

The Effect of Distance on Auditory Spatial Attention in the Peripersonal Space

著者	MONASTEROLO Florent
学位授与機関	Tohoku University
URL	http://hdl.handle.net/10097/00127080

The Effect of Distance on Auditory Spatial Attention in the Peripersonal Space

by

Florent MONASTEROLO

A thesis submitted to the Graduate School of Information Sciences of Tohoku University in partial fulfillment of the requirements for the degree of Master of Information Sciences

Evaluation Committee

Prof. Yôiti SUZUKI

Prof. Satoshi SHIOIRI

Prof. Yoshihiko HORIO

Prof. Shuichi SAKAMOTO

Tohoku University

February 2019

Preface

Auditory spatial attention is a core ability of human behavior. It allows specific examination of a desired sound while ignoring irrelevant sounds during various important tasks. It also allows unconscious monitoring of a listener's surroundings while doing different unrelated tasks. The study of these attentive capacities has inspired investigation of the so-called Cocktail Party Effect first described by Colin Cherry in 1953. This effect illustrates human auditory spatial attention, with the capacity of examining one conversation specifically against other competing conversations and noises during a cocktail party. Ever since, studies of auditory space perception, e.g. sound source separation and auditory spatial attention, have flourished, leading to a great leap in our understanding of auditory processing. Inspired by this knowledge, modern sound processing mechanisms have been developed for artificial intelligence, intelligent sensing in robots, and human hearing aids. Despite that progress, no machine has reached the levels of processing that the human brain can achieve. Therefore, many research efforts are still necessary to explain human auditory processes.

Among spatial auditory processes, the study of angular localization of sound source has been the most thoroughly conducted. Direction separation of competing sound sources has been shown to help greatly in elucidating our auditory environment, allowing the identification of the different sound sources surrounding us. We then have the ability to examine the desired sources specifically to achieve a particular task. To this day, however, few studies of the effects of sound source distance on auditory attention have been reported. Yet, the sound source distance strongly affects the properties of the sound reaching the listener. In fact, for sounds in the space within grasping reach, known as the peripersonal space, i.e. from within 1 m from the listener's head, the distance of sound sources affects direction perception. The results of these interactions on auditory attention remain quite unclear.

Therefore, for sound sources in peripersonal space, this thesis presents specific exam-

ination of a study of the effects of sound source distance on attention. In a first set of experiments, reaction times in a target speech sound search task are examined as a function of distance of sound sources. The effects of the relevant distance perceptual cues in auditory attentional tasks are analyzed. In a second experiment, the ability to examine a target presented at a particular distance specifically and ignore the competing sources is investigated. This is conducted by implicitly directing the attention of listeners to a focal distance and by comparison to conditions with no focal distance. These two experiments are discussed in separate chapters because they investigate different aspects of auditory attention.

Chapter 1 first introduces the background and motivation of this study. Studies of distance perception and of the potential importance of sound source distance on auditory attention are summarized. Finally, the study objective is presented.

Chapter 2 introduces the methods used to manipulate the sound source distance used for this study. Virtual sound sources are presented to listeners through headphones. Based on results of a numerical and subjective evaluation, the validity of this method is assessed.

Chapter 3 then presents a study of capture of attention using reaction time as a function of the sound source distance. This chapter specifically examines the stimulus-driven effects of source distance on auditory attention, investigating several absolute and relative source distances. According to the spatial information included in the stimuli, the effects of very near source distances on attention are evaluated.

On the other hand, Chapter 4 describes investigation of the effects of voluntary auditory distance attention. In this chapter, listeners' attention is attracted implicitly to the specific distance at which the target sound source is positioned. According to the distance at which the listener focuses, the difference in effects of auditory spatial attention is evaluated.

Finally, in chapter 5, all gathered results and interpretations are summarized to conclude the work done for this thesis.

Contents

Preface	i
1 Introduction	1
1.1 Auditory Scene Analysis (ASA)	2
1.1.1 Sound stream segregation	2
1.1.2 Auditory selective attention	4
1.2 Auditory distance perception	7
1.2.1 Cues for judging auditory distance	7
1.2.2 Near field distance perception cues	9
1.2.3 Other cues	12
1.2.4 Accuracy of distance perception	13
1.3 Contribution of sound source distance to auditory spatial attention	15
1.3.1 Sound intensity	15
1.3.2 Peripersonal space	15
1.3.3 Benefits from distance separation of sounds	16
1.4 Study objectives	17
2 Production of proximal sound sources using head-related transfer functions filtered through distance varying filters	19
2.1 Chapter objectives	20
2.2 Using head-related transfer functions to produce accurate direction of sound sources	21
2.2.1 Definition	21

2.2.2	Measurement	23
2.2.3	Limitations	24
2.3	Using distance varying filters (DVF) to produce near-field HRTFs	25
2.3.1	Motivation	25
2.3.2	Definition	25
2.4	Numerical evaluation of the model used	27
2.4.1	Comparing to a target calculated using the boundary element method (BEM)	27
2.4.2	Comparing to a measured HRTF in the near field	30
2.5	Perceptual evaluation	32
2.5.1	Experimental design	32
2.5.2	Results	35
2.5.3	Discussion	40
2.5.4	Evaluation conclusion	40
2.6	Chapter conclusions	41
3	Effects of sound source distance on spatial auditory attention to speech stimuli	43
3.1	Chapter objectives	44
3.2	Experiment design	45
3.2.1	Test participants	45
3.2.2	Apparatus and stimuli	45
3.2.3	Experimental procedure	47
3.3	Results	49
3.3.1	Analysis method	49
3.3.2	Average reaction time	49
3.3.3	Normalized reaction time as a function of perceived distance	54
3.4	Discussions	59
3.4.1	Effects of peripersonal space in virtual presentation of sounds	59
3.4.2	Sound stream segregation	60
3.5	Chapter conclusions	62

4	Top-down spatial auditory attention effects for distance of sound sources	63
4.1	Chapter objectives	64
4.2	Experiment design	65
4.2.1	Test participants	65
4.2.2	Apparatus and stimuli	65
4.2.3	Experimental procedure	68
4.3	Results	69
4.3.1	Analysis method	69
4.3.2	Average results	71
4.4	Discussion	76
4.4.1	Existence of the auditory spotlight for distance and interactions with peripersonal space	76
4.4.2	Relevance of these results	78
4.5	Chapter conclusions	80
5	Overall conclusion	81
A	DVF filtered HRTF localization accuracy - individual results	85
B	Bottom-up effects of distance - individual results	97
C	Top-down attention on distance - individual results	107
	Acknowledgments	113
	Bibliography	114
	List of works	121

Chapter 1

Introduction

1.1 Auditory Scene Analysis (ASA)

It is quite remarkable how efficiently human beings are capable to decompose their sound environment and evolve naturally in any sound space. When in a noisy environment, humans are exposed to a mixture of sounds of different levels, nature and positions. All these sounds add up at the entrance of the ear and are transmitted as this sum of sound waves within the ear canal, through the middle ear and eardrum and to the cochlea. How is it then, that we perceive so naturally our environment and process individual sound objects independently? This is the subject of auditory scene analysis (ASA), and of this study.

In the beginning, ASA was introduced with a different name but with a similar problematic. In the 1950s, Colin Cherry [1] studied the human capacity to focus on a particular sound source against other competing sounds. He illustrated this capacity through an example that he named the Cocktail Party effect, in which one is fully capable of focusing on a desired conversation against noise and competing conversations during an event such as a cocktail party. His designation has been widely spread and a great number of studies on this effect have been conducted following his study. Among them, studies on the frequency components, the signal duration, temporal structure, spatial positioning of a sound, the health of the listener and on many other factors have been examined. Albert Bregman included this effect in a bigger field and a bigger problematic that he named auditory scene analysis in his book published in 1990 [2]. This problematic is that of the processes involved in transforming the mixture of sounds overlapped at the entrance of our ears into meaningful sound objects. These processes begin with a step named sound stream segregation, or streaming.

1.1.1 Sound stream segregation

Sound stream segregation is the cognitive process of separating one's sound environment into individual consistent sound streams. This capacity depends greatly on the amount of sound objects presented simultaneously, on the nature of the individual sound objects and on the properties of each sound object relatively to its competing sound environment. A rule of thumb is that sounds which are coherent temporally, spectrally and/or

spatially are grouped together in one same stream [2, 3, 4]. For example, hearing a complex harmonic sound with a fundamental frequency f_0 and several harmonics $f_k = k * f_0$ results in a grouping of all components in one single auditory stream (Fig. 1.1 (a)). If one of the components becomes considerably non-harmonic, the listener then hears two different auditory streams : the non-harmonic sound as one stream, and the set of all harmonic sounds as another stream (b). If one of the components has a different temporal structure, it will also be separated into a different auditory stream (c). Finally, if the components are separated into different sound sources, increasing the directional separation between sources results in a separation of auditory streams, where each separated sound source becomes an individual stream (d).

Understanding the underlying mechanisms in sound stream segregation is essential to comprehend our capacities to analyse our acoustic environment. Indeed, in real-life situations, we are able to distinguish almost instantly all different sound sources in a room. Yet, the result of auditory stream segregation is a great amount of individual streams. This amount is strongly dependent on the complexity of the relationship of the sounds, and of the auditory environment. In order to comprehend our auditory environment regardless of this great amount of simultaneous auditory streams, a filtering, or selection of information is essential. This capacity in humans is known as auditory selective attention.

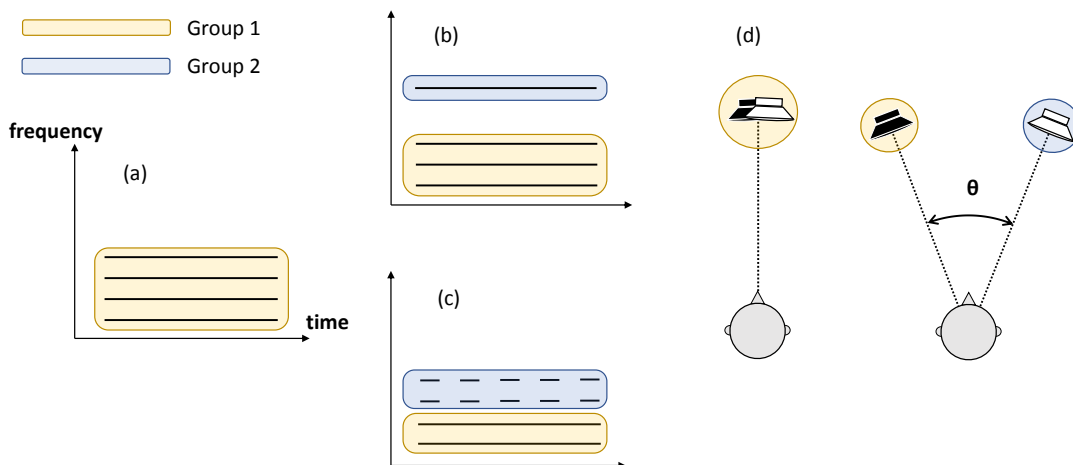


Figure 1.1: Schematic representing the principles of sound stream segregation with a harmonic sound. (a) the different components of a harmonic sound are grouped into one same stream. (b) a non-harmonic component within a complex sound is heard as a separate stream. (c) a component with a different temporal structure is heard in a separate stream. (d) spatial separation of sound sources leads to grouping into different streams.

1.1.2 Auditory selective attention

Auditory selective attention is our capacity to reduce the inconsiderable amount of auditory information we process at once to a reasonable amount. It can be compared to a bottleneck that selects only the most important sound information to go through the neck and be processed by higher cognitive levels. The selection of information is done either involuntarily or voluntarily using processes often designated as bottom-up and top-down attention.

Bottom-up attention

Bottom-up attention is the involuntary attention shift to certain sound streams. Hearing one's name while occupied with a task, for example, leads to automatic attention shift towards that sound source [5]. This capacity is essential for survival as it allows us to avoid dangerous situations or objects even without seeing them. Scharf [6] describes this type of attention as an "early warning system". The human brain constantly and unconsciously monitors our acoustic environment to scan for potential threats. This type of attention is believed to be stimulus-driven [7, 8]. That is, the process is believed to start from lower levels of the auditory system where the stimulus is analyzed, up to higher levels of cognition [8]. In this situation, the acoustic properties of the stimulus affect the mechanisms of auditory attention. For this reason, it is often qualified as bottom-up. The property that a sound has to capture or affect our attention is called salience, or saliency.

Top-down attention

As opposed to bottom-up attention, top-down attention is the conscious and voluntary focus of cognitive processes on a selected information stream. Attending to one conversation in a noisy environment, for example, leads to an increase of the isolation ability of the speech contents, revealing the increase of its intelligibility [1, 9]. Our capacities and the way we select information are task-dependent. That is, the listener selects the information relevant to achieve the given task and adapts his selection criteria and thresholds to this task. This process is believed to act as a feedback from higher level cognitive functions of the auditory system to bias the processing of incoming auditory stimuli. For this reason, it is often qualified as top-down.

Bottom-up attention and top-down attention are constantly acting together to better understand our auditory environment and react appropriately to sound events. Evidence for interactions between top-down and bottom-up processes in attention shifts [10, 11] and target search [8] are numerous. These interactions and their origins remain unclear for auditory attention.

Factors affecting auditory attention

Human auditory attention is affected by several factors. These factors can be related to the nature of the stimuli, the difference between competing stimuli, the position of sound sources, the listener's current health and mind-set, memory, and many other parameters...

The informational contents of the sounds are an important factor for salience of sounds. For example, Asemi *et al.* [12] showed that speech sounds are easier to detect than time-reversed speech sounds although the average spectral contents of both types of stimuli are identical, revealing the importance of informational contents. Reacting to one's name in a complex auditory environment [5] also goes to show this importance.

In addition to informational contents, the physical properties of sounds tend to affect auditory attention. Various studies have shown that the intensity of sound stimuli led to faster reactions and better processes [13, 14]. Additionally, the intensity ratio between a sound stimulus and its competing sounds directly affects its salience and intelligibility. The higher the ratio, namely the more intense the target sound is compared to other competing sounds, the more it "pops-out". In addition, spectral contents of the sounds and spectral differences between the target and competing sounds play an important role for stream segregation [2]. The bigger the spectral difference between target and competing sounds is, the easier the separation is. This leads to easier focus on one of the separated streams [15].

Finally, the position of the sound source is also believed to be an important factor. Sound source spatial separation leads to a better understanding of one's acoustic surroundings. It is therefore intuitive that this added understanding is essential in auditory selective attention. In addition, auditory processes are capable of identifying the nature and localization of sounds regardless of the the position of the sound source around the listener. Visual processes, on the other hand, cannot be relied on for sources outside of the field of view. Therefore, for sounds coming from the rear, attentional processes can only rely on auditory attention. Finally, human sound localization accuracy depends on the position of the sound source. The attentional selective capacities based on spatial position are therefore directly

dependent of the position of the sound source. Perhaps one of the most important finding is that when in a noisy environment, one can focus on a particular direction, leading to higher sensitivity to sounds coming from this direction, and ignoring the competing sound sources presented from other directions [16, 17, 18, 19]. This capacity was named auditory attention spotlight, inspired by its equivalent for visual attention, and by its properties.

In the past years, many studies have been led on the effect of azimuthal separation of sound sources and of azimuthal positioning of sound sources on attention[16, 18, 19, 20]. On the other hand, very little studies on distance have been led to this day. Yet, our perception of space is done both in direction and distance. In fact, the distance of a sound source from the listener's ears dramatically changes the sound reaching the listener. Beyond one meter from the head, change of distance is mainly perceived by the listener as a change in sound level, as explained later on in this chapter (Section 1.2.1). This change does not impact the perception of direction and is easily predictable both in physical and psychoacoustical dimensions [21]. However, for sounds within the space under one meter from the head, sound source distance is known to alter considerably the sound reaching the listener's ears [21, 22, 23, 24]. In this space, existing models and predictions for effects of source direction [25] may not be applied. To understand how proximal distances may impact sounds reaching the ears, let us review how judgement of distance is done by humans.

1.2 Auditory distance perception

1.2.1 Cues for judging auditory distance

Whereas direction perception is considered to be absolute, distance perception is believed to be relative. That is, when hearing a single sound presented from a single position in space, it is relatively easier to judge the direction of that sound than to give an accurate judgment of the distance of that sound. In fact, it is almost impossible to judge instantly the distance of a single sound source. The main reason may be that the judgement of distance is dominated by the judgement of sound intensity.

Sound intensity

Indeed, sound is a spherical wave propagating through the air. As the spherical wave propagates through space, its surface increases following a square law for radius. The power of the source, however is a finite value. In free field conditions, this results in an inverse square law of sound intensity for distance to the sound source :

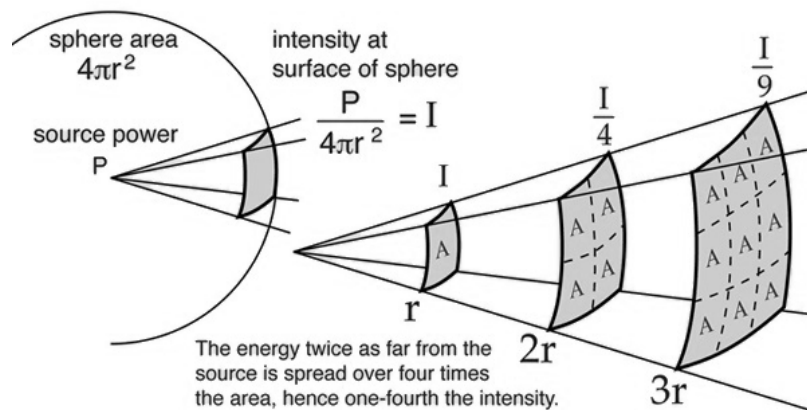


Figure 1.2: The inverse square law for sound intensity. Adapted from HyperPhysics, hosted by the Department of Physics and Astronomy, Georgia State University [26]

The further the sound source is, the lower the intensity reaching the listener's ears is. From the listener's point of view, this results in hearing sounds at a lower level. In free field, the relationship between the sound level at the listener's ears and the source distance from the center of the listener's head can be approximated to a reduction of 6 dB per doubling of distance.

For a long time, it has been believed that, in anechoic space, changing a source's intensity induces the same judgement of distance variation of the source position. For example, dividing the distance between listener and source and quadrupling the power of the source was considered to be perceptually equivalent. This illustrates how dominant the sound intensity cue is. However, we now know that this is not true for proximal sound sources, for sources within a reverberant environment, or for great distance changes.

Direct to reverberant intensity ratio

In anechoic space, the listener hears the direct sound wave travelling from the source. In reverberant space, reflections on objects or walls are added to the direct sound as a kind of secondary sound sources. The sound reaching the listener's ears is therefore a sum of the direct sound wave propagation and of all the secondary sound waves resulting from reflections. The reflected sounds consistently arrive with a delay to the ears, and with altered spectrum and phase characteristics.

Several studies show that listeners can use the reverberations in a room to improve judgement of sound source distance [27, 28]. The main cue used in distance localization using reverberation is believed to be the direct sound to reverberant sound energy ratio (DRR). It is defined, for a given location, by the ratio between the energy of the direct sound and the energy of the reverberant sounds simultaneously incident to the same location. DRR reaching the listener's ears depends of the sound source distance, the distance of the reflecting surfaces and the nature of these surfaces. Studies suggest that the DRR could be used as an absolute cue for distance perception in rooms [29, 30, 31].

High frequency attenuation

Air is not an ideal medium. Therefore, a soundwave's amplitude spectrum changes as it propagates through the air. This attenuation can be expressed by the following law in free field conditions :

$$I(k, r) = I(k, r_0)e^{-\alpha k(r-r_0)} \quad (1.1)$$

where α is an attenuation coefficient that depends on the temperature, pressure and humidity of the air and k is the wavenumber defined by $k = \frac{2\pi f}{c}$ where f is the temporal frequency and c is the speed of sound through the air. Practically, this results in high-frequency components being more attenuated than low-frequency components. As a consequence, far-away sounds are generally heard as darker, more muffled than the sound emitted at the source. The rate of frequency attenuation is rather small : 3 to 4 dB loss for every 100 m at 4 kHz for atmospheric conditions and average humidity [32]. This has therefore a limited effect for most everyday situations, where sounds are heard from much closer.

1.2.2 Near field distance perception cues

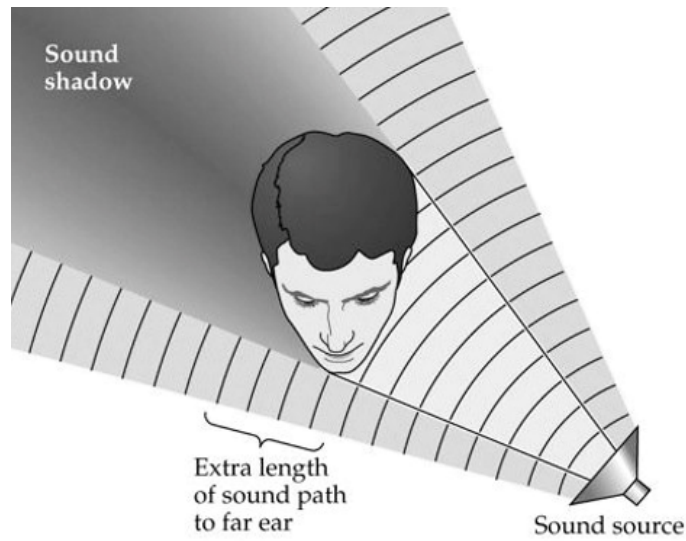
When sound sources are heard from distances within 1 m from the listener's head (near field), other cues for judgement of distance are dominantly used. The effect of the listener's head, shoulders and torso on the sound reaching both ears affects significantly sound within near field distances. In addition, a phenomenon designated as auditory parallax is used.

The head shadow effect

The head shadow effect is characterised by the effects of the head on the incoming sound wave. When the sound wave comes from the side, the presence of the head induces a difference in paths that the wave has to travel from the source to the ears. This leads to an interaural difference in level (ILD), time of arrival (ITD) and spectrum (Fig. 1.3). In anechoic space, the level at the closer ear (ipsilateral ear) is constantly higher than at the further ear (contralateral ear). Conversely, the sound wave reaches the contralateral ear with a time delay compared to the ipsilateral ear. ILD and ITD are functions of sound temporal frequency. For sound sources in the far field (beyond 1 m), the effect of the head on ITD and ILD is independent of source distance [21]. However, in the near field, the effect of the head significantly increases as a function of sound source proximity. The head shadow acts so that closer sounds lead to higher ILD, and to higher scattering which leads to alterations of the sound spectrum [21, 22, 23].

Auditory parallax

Auditory parallax is defined by the difference in angle between the paths from the source to both ears (Fig. 1.4). For distal sounds, this angle is small enough to be ignored. However, for proximal sounds, the difference is innegligibly increasing. Although the effect of this angle is rather unclear, Brungart [33] suggested that this angle led to a distortion of the spectrum at each ear. Kim *et al.* [34, 35] showed that listeners can achieve distance judgements comparable to judgements for real sound sources, by controlling the parallax for virtual sources.



© 2001 Sinauer Associates, Inc.

Figure 1.3: Schematic effect of the head shadow. The presence of the head in the sound field results in a difference in paths that the wave travels between the source and both ears. This difference in paths leads to level and time differences (ILD and ITD) and to a difference in the spectrum between the ipsilateral ear and the contralateral ear. Adapted from Sinauer Associates, Inc. 2001.

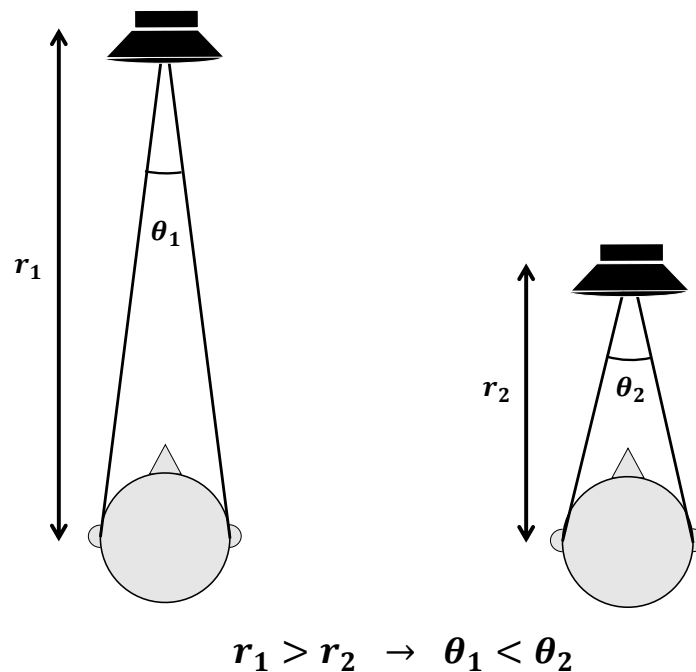


Figure 1.4: Schematic of the origin of auditory parallax. As sounds sources get closer, the angle between the paths to the left ear and to the right ear increases.

1.2.3 Other cues

Sound familiarity

Several studies show that familiarity of the target source helps greatly in distance localization [31, 36]. When a single unknown sound is presented in an unknown room, judgement of the source distance is almost impossible. However, knowledge of the room's acoustics or of the initial sound leads shows an increase in localization abilities.

A study on shouted, spoken and whispered sounds, showed that perceived distance of a speech sound depended on its vocalization and on the vocal effort [37]. Whispered sounds were consistently perceived as closer than phonated sounds, regardless of the presented sound level. Similarly, shouted sounds were consistently perceived further than phonated sounds. The study also claims that listeners are capable of using their acquired knowledge of speech characteristics to adapt their judgement of distance.

Crossmodal cues

In the processing of our environment, various sensory information is simultaneously integrated. Indeed, vision, audition and tactile sensations work together to create our perception of our proximal environment. Typically, seeing the sound source increases greatly our sensitivity to the audio contents produced by this source. For example, watching the lips of a talker affects positively speech intelligibility when the audio and visual contents match [38], and negatively when there is a mismatch [39]. In soundspace perception, spatially matching audio-visual stimuli affects very positively localization accuracy [40] whereas spatial mismatch leads to confusion and degraded localization accuracy [41]. The bias induced by other senses should therefore be considered carefully when presenting the listeners with sound sources.

1.2.4 Accuracy of distance perception

Judgement of distance of sources using sound only is poorer and more fluctuating than judgement of direction [42]. Humans tend to overestimate the distance of proximal sound sources, and to underestimate distance of distal sound sources. This is known as the auditory horizon phenomenon. The reasons for this phenomenon remain unclear to this day.

In 2005, Zahorik *et al.* published a review of 84 datasets on distance localization in different environments and test paradigms. All results in these individual evaluations were obtained by eliminating the effect of distance on sound level reaching the center of the listener's head, without eliminating the ILD information. This was done in order to extract only effects of distance perception cues. Results agree to fit estimated distance as a power function of presented distance : $r' = kr^a$, where r' is the distance estimated by the listener, r is the real source distance and k and a are fitting coefficients. Using these 84 datasets, the average value for k was 1.32 and the average value for a was 0.54. Note that the value of a is below 1, illustrating the overestimation for near distances and underestimation for far distances. This study included both results from virtual presentations of sound sources and for presentation of real sound sources. The result of the average psychophysical function is illustrated on Fig. 1.5.

A difference between distance perception for virtual sources and for real sources is observed. The overestimation of distance for virtual sound source distance is more important than for real source localization [43]. On the other hand, Kim *et al.* [35] reported that estimation of distance for both real sound sources and virtual sound sources are accurate for sources within 1 m from the listener's head (Fig. 1.6), with little overestimation. Overall, the results of accuracy from studies using different test environments differ greatly, indicating the importance of the test environment in the estimation of source distance.

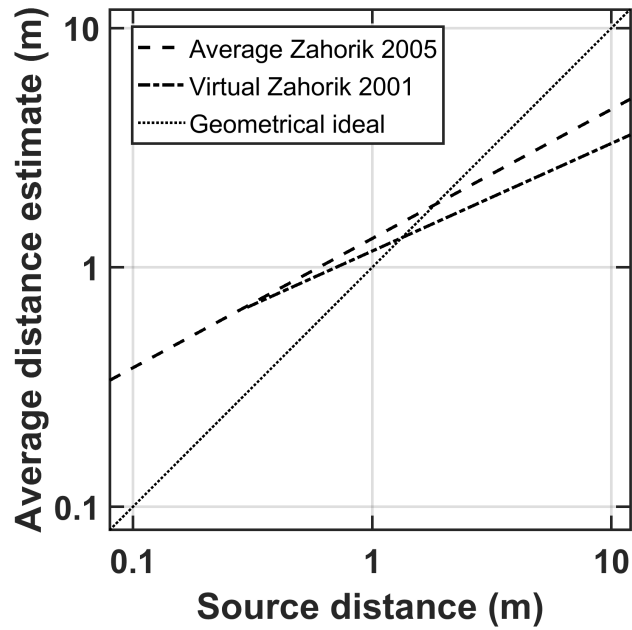


Figure 1.5: Schematic of the average perceived distance to presented distance scale. The dashed line represents the results gathered from 84 datasets by Zahorik *et al.* in 2005. The dash-point line represents the results for virtual sources presented by Zahorik and Wightman in 2001. The dotted line represents the geometrical ideal. Adapted from [42, 43].

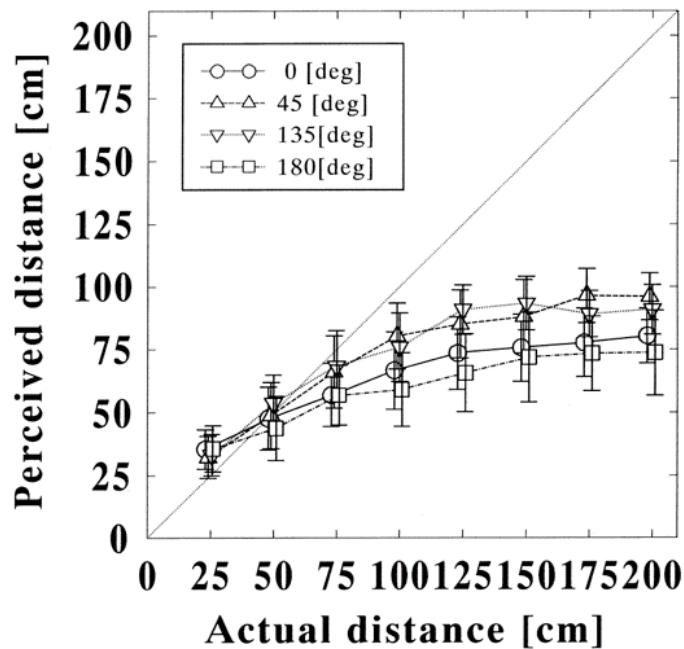


Figure 1.6: Results of distance estimation of real sound sources [35]. Circle marks represent results for sources directly in front of the listener, upper triangles for sources 45° to the left, lower triangles for sources 135° to the left (behind the listener) and squares for sources directly behind the listener.

1.3 Contribution of sound source distance to auditory spatial attention

To this day, very little studies have considered the effects of distance of sound sources in ASA. However, some main results on the effects of sound source distance on attention can be used.

1.3.1 Sound intensity

Sound intensity is a dominant cue for distance judgement and is a primary sound characteristic. It comes naturally that studies of the effect of sound intensity on attentional factors have flourished [13, 14, 44]. It is an intuitive statement that as the sound stimulus intensity increases, the greater its intelligibility, the higher the stimulus detection rate, and the faster the reaction time to this sound is. Chocholle was the first to find that the relationship between sound level (in decibels) and reaction time is a power function : $RT = k(SP)^n$ where RT is the reaction time, SP is the sound pressure level of the stimulus in decibels and k and n are fitting coefficients. The effect of sound intensity must therefore be carefully considered when studying auditory attention for different sound source distances.

1.3.2 Peripersonal space

Sounds presented from within one meter from the listener's head have a special role. First, the head shadow has its strongest effects within this space, and those effects are distance dependent (Section 1.2.2). We therefore have more information to accurately judge distance of sound sources when they are within this area. Second, one meter corresponds roughly to the upper limit of the reaching distance for the average adult. This reaching distance is named peripersonal space (PPS). Graziano *et al.* [45] studied the representations of auditory PPS on brain activity in monkey test subjects. They found that sounds within PPS lead to a higher brain activity. Their results show that some neurons responded to very near sounds (within 30 cm) with more activity than to further sounds, regardless of the

sound level reaching the ear. After these studies, several studies on neural representation of PPS in humans and monkeys were reported [46, 47]. Results from these studies suggest that the size of PPS is task dependent, and that neural processes for sounds within PPS differ from those outside of PPS. A possible explanation for this difference is that within reaching distance, the tactile sensation also becomes a factor. This means that processes become multimodal within PPS [45, 48].

1.3.3 Benefits from distance separation of sounds

In 2001, Barbara Shinn-Cunningham *et al.* [49] tested the effect of distance separation of a proximal speech sound source and a proximal noise sound source on speech intelligibility. They used a simple spherical head model to create spatialized virtual sound sources presented through headphones. The result is that distance separation of a speech sound from its masker noise sound leads to higher speech intelligibility and lower speech reception thresholds (SRT). This suggests that distance separation of sources benefits sound stream segregation.

Following this study, Brungart and Simpson [50] investigated the relationship between the nature of the sounds used in the experiment and distance unmasking. They used a similar experiment design as Shinn-Cunningham *et al.* using a generic dummy head model for virtual sound spatialization. Their study shows that the effect of spatial advantage due to the distance separation depends on the characteristics of target and masker sounds. When both sounds have very different characteristics, no effect is observed. When both sounds have similar characteristics, this effect is increasing. In the case they presented, spatial advantage from distance separation is most prominent when separating two speech sounds uttered by speakers of the same gender. They also suggest that the main contributors to this spatial advantage are interaural differences due to the head shadow that occur with source distance separation.

1.4 Study objectives

In light of these results, the question of the effects of sound source distance in attentional tasks remain unclear. If both sound space perception and neural processes are different for sounds within PPS, then it can be hypothesized that sound source distance greatly affects attentional capacities within this space. This study aims to contribute in verifying this hypothesis. Auditory attention is investigated as a function of peripersonal sound source distance in several conditions in this study. The questions that this study aims to answer are the following: (1) What are the effects of peripersonal source distance on auditory attention? (2) What distance cues in proximal space benefit auditory attention? (3) Is top-down spatial attention on source distance possible? Namely, are listeners capable of focusing on a particular distance and processing the information presented from this distance with higher selectivity? The novelty of this study is to evaluate spatial auditory attention as a function of distance of sound sources within PPS, using reaction time in an auditory search task as a measure of attention. This evaluation is done using accurate presentation of virtual sound sources through headphones.

Chapter 2 introduces the methods used to manipulate distance of the sound sources used in this study. Head-related transfer functions (HRTFs), manipulated using distance varying filters (DVF) are introduced. Both a numerical and a subjective evaluation of these calculated functions are presented. In comparison to previous studies, this evaluation justifies the use of HRTFs in distance-related psychoacoustic experiments.

Chapter 3 presents a study of simple reaction time to speech sounds as a function of source distance within peripersonal space. Conditions including and excluding source distance separation of competing sounds are considered. The configurations are chosen to separate the effects of individual distance cues, and particularly to eliminate the dominance of sound source intensity in distance localization. Results suggest a stronger salience of closest the sounds, especially when separated in distance from competing sounds.

Chapter 4 contributes to the study of top-down spatial attention related to the distance of the sound source. Here, the listener's top-down spatial attention is implicitly attracted to a specific source distance before sound presentation. Results suggest that focus on the specific source distance induces faster reaction time for sounds presented from this distance, and slower reaction time for sounds presented from different distances. This reveals effects

of top-down auditory attention to distance, for sources presented from within peripersonal space.

Finally, Chapter 5 considers all gathered results and interpretations to conclude on the work done in this thesis.

Chapter 2

**Production of proximal sound sources
using head-related transfer functions
filtered through distance varying filters**

2.1 Chapter objectives

There are various ways to present realistic sound space information to the listeners. Using real sound sources to present the sound space information has the advantage of being realistic and straightforward. However, this has some difficulties at the points of flexibility and reproducibility. Indeed, results using real sound sources depend on the precision and resolution of the infrastructure. This effect would be observed clearly for proximal sound source presentation. In addition, for proximal sources, the perceived width of the source depends of the size and distance of the loudspeaker. Finally, knowledge of the positions of the loudspeakers may bias auditory perception. Presenting virtual sound sources via headphones has an advantage at the points of flexibility and reproducibility. However, quality of generated sound space is not high, in general.

The objective of this chapter is to introduce the method used to present realistic virtual sound sources to the listeners who participated in the experiments presented in this thesis. The evaluation of the perceptual accuracy of sound information synthesized by the applied method has also been focused on. Binaural sounds were presented to the listeners through headphones. In order to simulate sound space virtually, a filtering method for the listener's individual head-related transfer function (HRTF) was used. The results of this filtering method were analyzed both numerically and perceptually in order to justify the use of the current method.

This chapter presents in section 2.2 how virtual sound sources are spatialized in direction using HRTF. In section 2.3, the filters applied to these HRTF to manipulate apparent distance of sources, named distance varying filters (DVF), are introduced. Section 2.4 presents a numerical evaluation of the result of filtering the HRTF. Section 2.5 presents a perceptual evaluation of these filtered HRTF. Finally, Section 2.6 presents the conclusions from this chapter.

2.2 Using head-related transfer functions to produce accurate direction of sound sources

2.2.1 Definition

A powerful tool commonly used to synthesize virtual sound sources binaurally is the head-related transfer function (HRTF). An HRTF for a specified ear for a certain sound source position is obtained using two transfer functions: (1) the transfer function of the sound propagation path from the sound source at a certain direction to the entrance of the subject's ear canal, and (2) the transfer function of the sound propagation path from the same source to the position corresponding to the center of the head with no subject present. The HRTF is calculated as the ratio of transfer function (1) to (2) [51]. They capture the spectral effects of the shadowing and scattering on the head, torso and pinna on the sound wave reaching the listener's outer ear. These effects strongly depend on the individual's physical features. In general, high intensities appear for sound sources on the same side of the ear (the ipsilateral side), and lower intensities for sound sources on the opposite side of the ear (the contralateral side). When spatial sounds are synthesized by using HRTFs, it is of great importance to use the individual's HRTF. Although, some generic HRTFs using a dummy head can also be used, the performances are poorer for space perception [52, 53]. An example of an HRTF for a dummy head is given in Fig. 2.2.

Suppose we know the left and right ear HRTFs for one position of a sound source obtained for one individual. Applying these two HRTFs to a monophonic sound results in a signal for each ear of the listener. When presenting these signals via headphones or earphones, the listener will hear a virtual sound source at a position matching that of the initial sound source. The process is illustrated in Fig. 2.1.

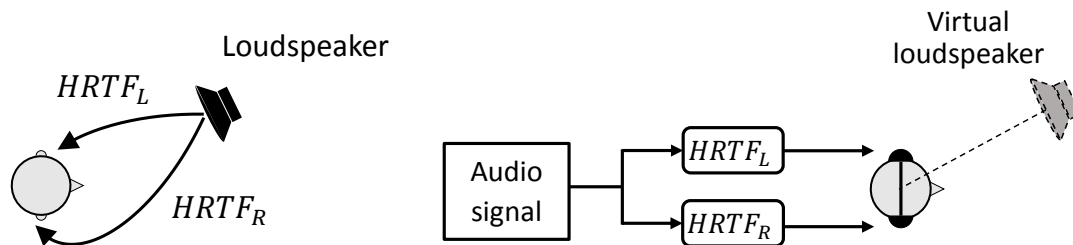


Figure 2.1: Schematic of the origins of the HRTF. If knowing an individual's HRTF for a position of a sound source, one can recreate virtually a sound source matching the position of the initial sound source.

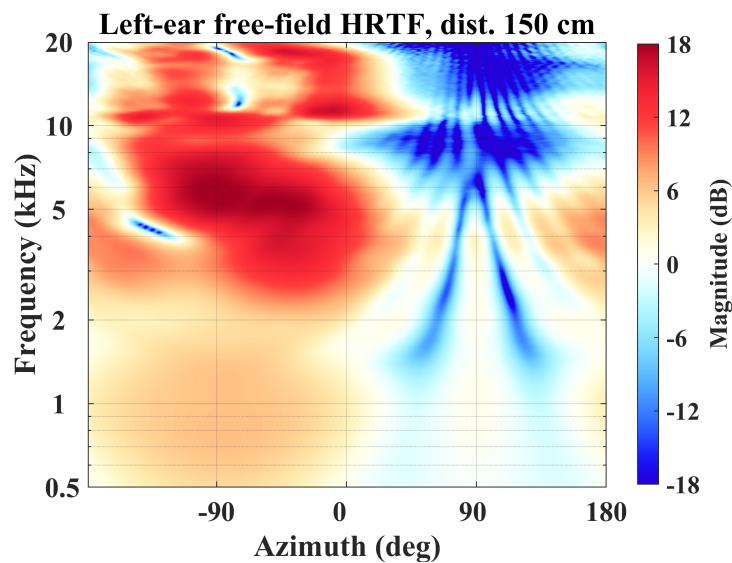


Figure 2.2: Example of a circular HRTF calculated for the left ear of a head model. The calculation was done at 1.5 m, with an angular resolution of 1° . Magnitudes indicate the difference in sound pressure level at the left ear compared to the sound pressure at the head's center in free-field conditions.

2.2.2 Measurement

Measurement of HRTFs is generally done in an anechoic room. The listener is sat at the center of a loudspeaker array, and these loudspeakers are placed at the desired measurement positions. Miniature microphones are set at the entrance of the listener's ear canal. To fix the microphones at the recording point, ear molds are inserted with the microphones. These microphones aim to record the sound reaching the listener's ear canal. An impulse or a train of swept-sine signal [54] is then presented from one loudspeaker at a time. An alternative is to use a miniature loudspeaker inserted in the listener's ear, and recording with a microphone array, based on the reciprocity method [55, 56, 57]. The presented sound is transmitted to the microphone position. The recorded sound using microphones is, then, converted to its frequency response. In this condition, reflection or scatterings on the listener's head is not included in the recorded sound. This sound is also converted to its frequency response. By normalizing the complex signals obtained with the head by the complex signal obtained in free-field condition, the left ear and right ear HRTF of the listener for one position of a sound source is generated. This procedure is repeated for every desired position of sound source.

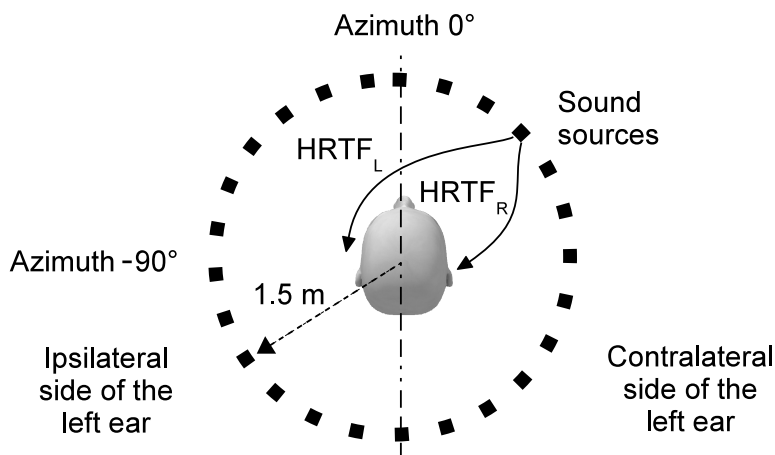


Figure 2.3: Top view of an example measurement system for the left and right ear HRTF of a head. Sound sources that lie on azimuths on the same side of an ear are said to be on the ipsilateral side, and the ones lying on azimuths on the opposite side of the ear, on the contralateral side.

2.2.3 Limitations

Measuring HRTFs is a complex, time-consuming and resource-consuming process. The infrastructure needed is a heavy and precisely calibrated one. Moreover, to measure a precise HRTF, listeners may need to stay motionless for several hours. The complexity of the needed infrastructure is greatly increased if one wants to measure HRTFs for several distances. Studies to accelerate this process have been conducted [55, 58] but a compromise between precision and speed of measurement must be done.

HRTF calculation also assumes that sound sources are point sources. Therefore for close distances, a smaller loudspeaker must be used to be able to meet this assumption. Yet, reducing the size of the loudspeakers impacts the frequency range that they can emit. It follows that the frequency range of the resulting HRTFs is limited. For this reason, most HRTF measurement systems only consider distal sounds with a fixed distance around 1.5 m. An additional reason for this is that beyond 1.5 m, HRTFs do not change with sound source distance [21]. In order to conduct psychoacoustic experiments in peripersonal space distances, the use of a conventional HRTF measurement system for a single far distance, followed by the application of distance varying filters is considered.

2.3 Using distance varying filters (DVF) to produce near-field HRTFs

2.3.1 Motivation

In order to synthesize accurate near field HRTFs, a filtering method using distance varying filters (DVF) [59, 60] is applied to the HRTF measured for distal sounds. This method aims to reduce the cost and increase the reproducibility of near-field HRTFs, with satisfactory precision. Other synthesis methods exist such as methods using broad anthropometric measurements [58, 61] or methods using a precise 3D model of the listener's head [62]. However, to obtain the listener's precise 3D model, heavy infrastructure is needed, such as an MRI measurement system or a multi-camera system. Using DVFs has for advantage that this complex 3D modelling is not needed.

2.3.2 Definition

The method used is the distance varying filter (DVF) developed by Salvador *et al.* [59, 60]. They are filters applied to an HRTF dataset measured for a circular array of distal sound sources in order to approximate HRTFs for closer distances (illustrated in Fig. 2.4). DVFs assume that the circular array used for measurement of the initial HRTF dataset is situated at 0° elevation angle. This means that they are only used on the horizontal plane which includes the listener's two ears.

DVFs are obtained as the result of solving the wave equation for sound in spherical coordinates. In this situation, invariance of sounds is assumed along elevation. The obtained filter is a function of frequency, and of sound source distance and azimuth.

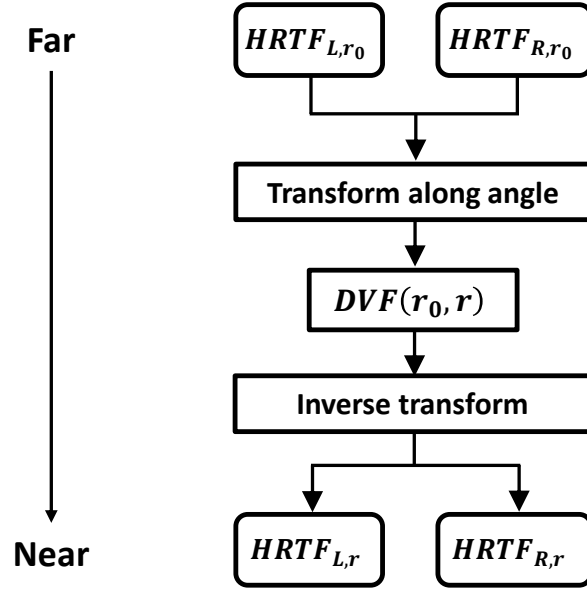


Figure 2.4: Schematic on how to apply DVFs to obtain an array of near field HRTFs. With a given dataset of HRTFs at a distance r_0 , applying the DVF from distance r_0 to r to these HRTFs results in a dataset of HRTFs at a distance r .

$$D_m(r, r_0, \omega) = \frac{r^{-\frac{1}{2}} H_\mu\left(\frac{\omega}{c} r\right)}{r_0^{-\frac{1}{2}} H_\mu\left(\frac{\omega}{c} r_0\right)} \quad (2.1)$$

Equation 2.1: Equation of the DVF D for angular mode m , desired distance r , distance of the initial HRTF data set r_0 , and angular frequency ω . H is the Hankel function of the second kind and fractional order μ . μ is defined by the value $\mu^2 = \frac{m^2}{\cos^2(\phi)} + \frac{1}{4}$ where ϕ is the elevation of the sound source. The angular mode depends on the angular resolution of the base HRTF dataset.

2.4 Numerical evaluation of the model used

In order to confirm the validity of using the considered sound source spatialization method, a numerical evaluation as compared to a reference method is presented in this section. Near-field measurements of HRTFs using the reciprocal method [55, 56], were also conducted with the Hirahara laboratory from Toyama Prefectural University. The DVF filtered individual HRTFs are then compared numerically to the near-field measurements of these individuals' HRTFs.

2.4.1 Comparing to a target calculated using the boundary element method (BEM)

In a previous study led by Salvador *et al.* [60], who developed the DVFs used in this thesis, DVF filtered HRTFs for an individual's head model were compared numerically to a dataset of reference HRTFs. These reference HRTFs were obtained by applying the Boundary Element Method (BEM) [62] for this individual's head 3D model. The 3D model was obtained using MRI imaging. The reference HRTFs were synthesized using the BEM for an angular resolution of 1° of the full circle and a distance resolution of 1 cm ranging from 10 cm to 150 cm from the center of the listener's head. DVF filters for the same angular and distance resolution was calculated. The DVF filters were applied to the BEM calculated HRTF at 150 cm in order to obtain the test data.

The numerical comparison was done based on two objective measures of overall accuracy. Overall accuracy along frequency is measured using the spectral distortion (SD) in decibels. This corresponds to a logarithmic spectral distance, shown to be suitable for predicting audible differences between measured and synthesized HRTFs [63]. Spectral distortion is defined as :

$$\text{SD}(\theta) = \left[\frac{1}{f_2 - f_1} \int_{f_1}^{f_2} \left[20 \log_{10} \left| \frac{\hat{\mathcal{H}}(\theta, f)}{\mathcal{H}(\theta, f)} \right| \right]^2 df \right]^{\frac{1}{2}} \quad (2.2)$$

where f_2 and f_1 define the frequency range over which the the SD is calculated, θ is the source azimuth, $\hat{\mathcal{H}}$ is the HRTF obtained by the DVFs and \mathcal{H} is the reference HRTF

obtained by the BEM for the same distance.

Overall accuracy along angles is measured using circular correlations (CC). They correspond to a measure of the similarity of directional patterns in the two compared HRTFs [64]. Normalized CC is defined as :

$$CC(f) = \frac{\int_{-\pi}^{\pi} \hat{\mathcal{H}}(\theta, f) \overline{\hat{\mathcal{H}}(\theta, f)} d\theta}{\int_{-\pi}^{\pi} |\hat{\mathcal{H}}(\theta, f)|^2 d\theta \times \int_{-\pi}^{\pi} |\hat{\mathcal{H}}(\theta, f)|^2 d\theta}. \quad (2.3)$$

The results of the numerical comparison for the HRTF for one ear of one head model are displayed in Figs. 2.6 and 2.7. Results suggest that the highest errors occur at the closest distances. These errors are most prominent at the contralateral ear, and around frequencies 10 kHz and 16 kHz. These errors are believed to be due to the naturally low energies of HRTFs at 10 kHz and 16 kHz, leading to higher tendency for error in this area. DVFs tend to slightly underestimate the decrease in level at the contralateral ear that normally occurs due to the increased effect of the head shadow as sources are brought closer. However, little errors occur at the ipsilateral ear, for which the energy is highest. Errors are acceptable until approximately 25 cm. The impact of these errors on the perception of space is to be determined in Section 2.5.

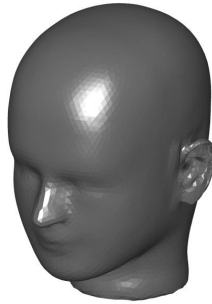


Figure 2.5: Three dimensional representation of the head model used for numerical comparison.

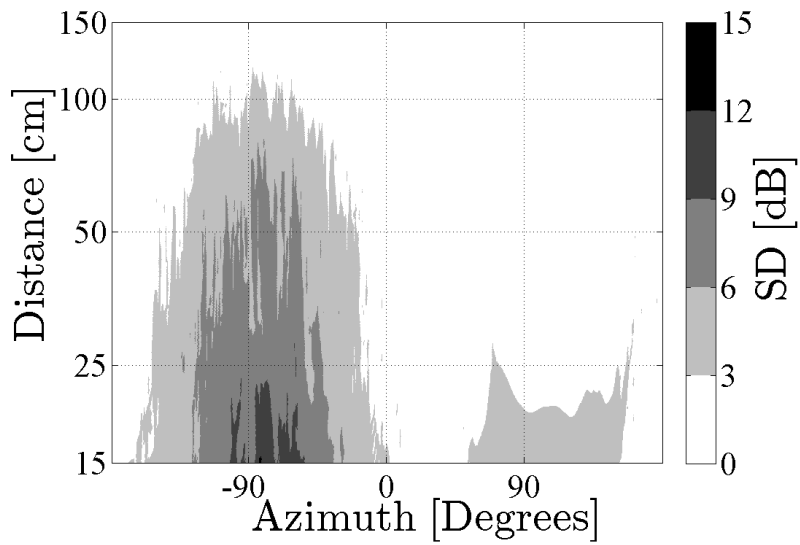


Figure 2.6: Results for the spectral distortion between the DVF filtered HRTF for one individual's right ear and the BEM calculated HRTF for that same ear. The lower the SD is, the closer the frequency spectrum of both HRTFs is. The highest errors occurred for the contralateral ear and for the closest distances.

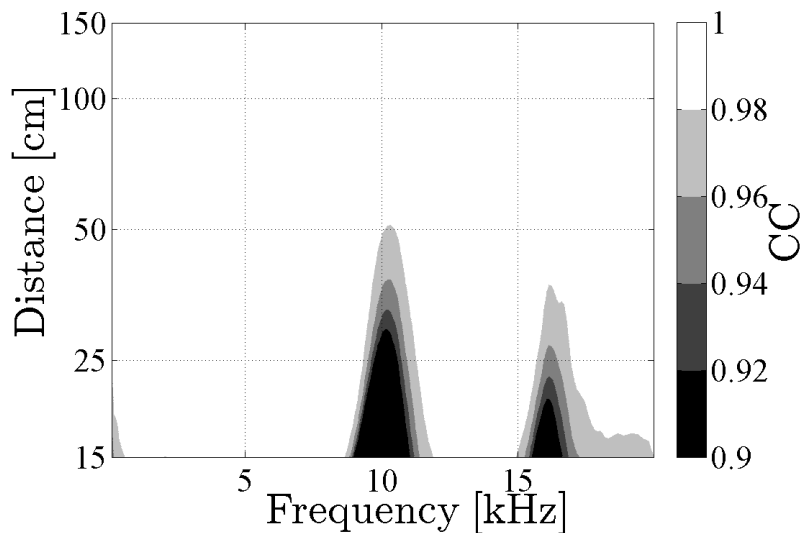


Figure 2.7: Results for the circular correlation between the DVF filtered HRTF for one individual's right ear and the BEM calculated HRTF for that same ear. The closer the CC is to the value one, the closer the angular patterns of both HRTFs are. The highest errors occurred around 10 kHz and 16 kHz and for the closest distances.

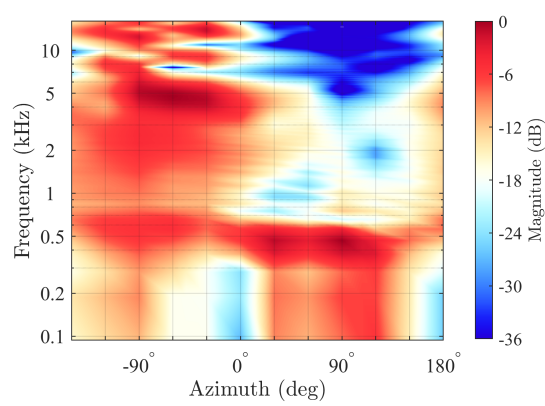
2.4.2 Comparing to a measured HRTF in the near field

This section compares the DVF filtered individual HRTFs with a dataset of measured HRTFs. Near-field HRTFs of three listeners' head were measured using the reciprocity method [55, 56, 57]. This measurement was done at the Hirahara laboratory from Toyama Prefectural University. Measurement was done in a sound proof room. The listener sat on a chair on one side of the room. The microphones and loudspeakers were similar to the ones used in [57] and [56]. These microphones are placed on an array of 11 distances from 15 cm from the head to 115 cm at a regular interval of 10 cm. HRTFs for 12 different azimuths from 0° to 330° with a 30° resolution were measured. Examples for the resulting HRTF and the calculated DVF filtered HRTF are illustrated in Fig. 2.8 for one head model and several distances.

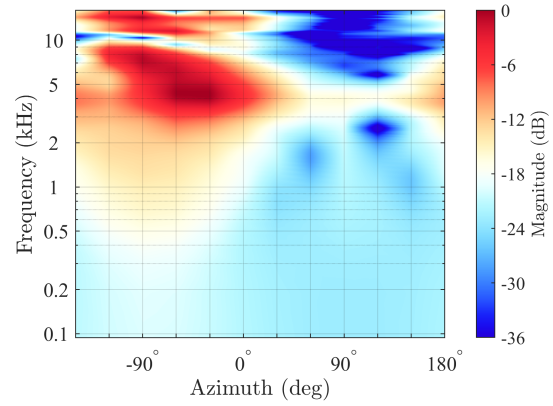
Because of the small size of the loudspeakers, and of the proximity between the listener's ear canal and the loudspeaker, the sound pressure level of the impulse at the measurement was very low. This results in a low signal to noise ratio, especially at low frequencies, leading to inaccuracies of the measured HRTFs for under approximately 1 kHz [56, 57]. This can be observed on Figs. 2.8a and 2.8c. Whereas little variation of the HRTF along angle is normally observed at low frequencies for dummy heads [21], there are high variations observed in this measurement.

However, beyond 1 kHz, the measured HRTF and the calculated DVF filtered HRTF are similar. A maximum of magnitude appears for the ipsilateral side between 3 kHz and 7 kHz. The angle of maximum of magnitude is closer to 90° with closer sources. Moreover, the angular width of the peak decreases with sound source distance [21, 33]. A second maximum at the ipsilateral side appears beyond 10 kHz for both HRTFs. The main difference between measured and calculated HRTFs is in the lower frequencies at the ipsilateral ear. Magnitude of the calculated HRTF at the frequency range between 1 kHz and 3 kHz is around 6 dB lower than that of the measured HRTF. This could highlight an underestimation of the impact of the head shadow on the interaural level differences for the closest sounds when DVFs are used to generate near-field sounds.

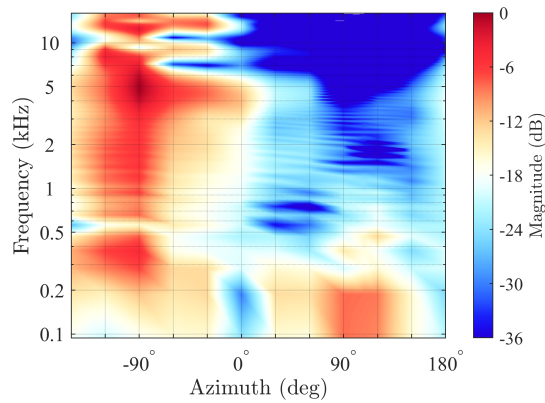
The impact of the numerical differences between measured HRTF and DVF filtered HRTF on distance perception is unclear. A subjective experiment is therefore conducted to test the accuracy of distance localization using DVF filtered HRTFs.



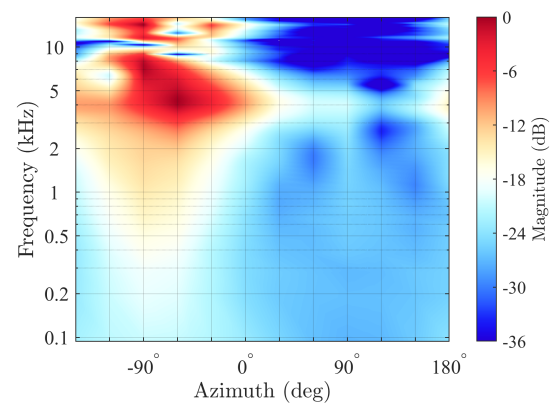
(a) Left ear HRTF measured for a distance of 45 cm using the reciprocity method.



(b) Left ear HRTF calculated for a distance of 45 cm using DVF filters.



(c) Left ear HRTF measured for a distance of 15 cm using the reciprocity method.



(d) Left ear HRTF calculated for a distance of 15 cm using DVF filters.

Figure 2.8: Comparison of measured and generated left ear HRTFs: (a) and (b) HRTFs for 45 cm, (c) and (d) HRTFs for 15 cm, (a) and (c) HRTFs measured using the reciprocity method, (b) and (d) HRTFs calculated using DVFs applied to the BEM calculated HRTF for 1.5 m. The HRTFs are normalized by their maximum value of amplitude and presented on a log scale for magnitude. The frequency range is limited to 16 kHz, above which the DVFs can not be relied on.

2.5 Perceptual evaluation

No perceptual evaluation of DVF filtered individual HRTFs has been investigated in previous studies. The objective of this section is thus to provide a perceptual evaluation to determine whether this spatialization method is available in psychoacoustic studies. This evaluation is done through the measurement of listeners' accuracy in localization of virtual sound sources. The accuracy test is performed for both azimuth and distance, separately.

2.5.1 Experimental design

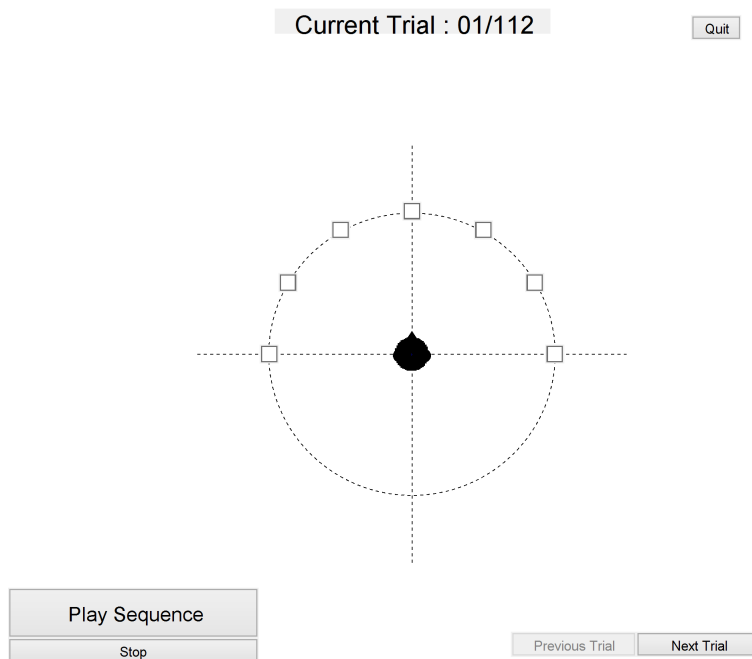
In order to evaluate the accuracy of angular and distance estimation of sound sources, two separate experiments were examined consecutively. The first experiment was the evaluation of angular accuracy. It consisted in an absolute estimation of the direction of the presented virtual sound. The second experiment was the evaluation of distance accuracy. It consisted in a relative estimation of the target sound source distance by comparing to the distance of a reference source. In the experiments, ten young students (9 male, 1 female, ages 21-24, average age 23) with normal hearing participated. All were from the Graduate School of Information Sciences of Tohoku University. Before starting the experiment, listeners were trained to use the response interface by evaluating the position of 20 sound sources picked at random.

In both experiments, listeners sat in a sound proof room in front of a computer with a Matlab UI window displayed (illustrated in Fig. 2.9). All answer inputs and the progression of the experiment used this interface. Sound stimuli were presented through a RME BabyfacePro [65] sound card and headphone amplifier, with a 48 kHz sampling frequency, connected to Sennheiser head phones (HDA-200). The sound stimuli were a train of 5 consecutive rectangular windowed 150 ms white noise bursts, with an inter-stimulus interval (ISI) of 30 ms. These stimuli provide the listener with a broadband spectrum and a temporal structure, making them robust for sound localization. The headphone transfer function was compensated by applying the inverse transfer function calculated using a BK4153 artificial ear and repeated swept-sine signals [54]. Sound pressure level was set to 65 dBA at each ear when the virtual sound source was at 1 m and straight in front of the listener.

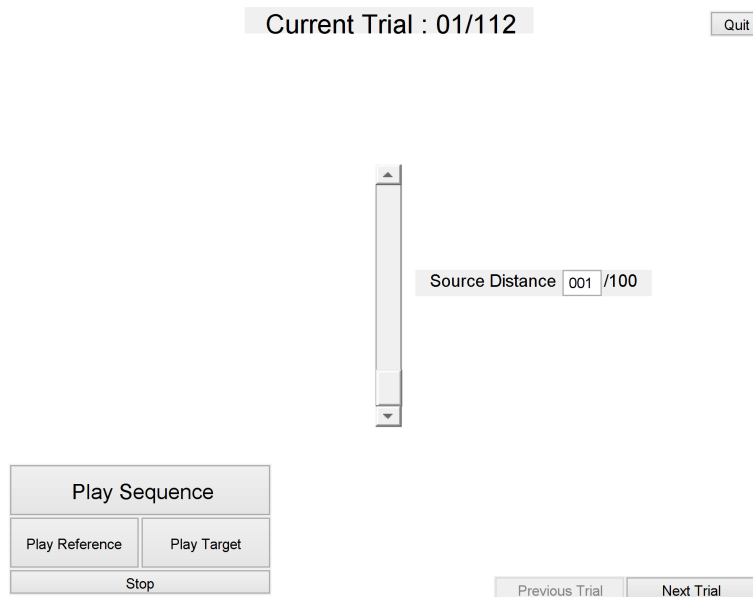
All stimuli in this experiment were spatialized using the listeners' individual DVF filtered HRTFs. These HRTFs were constructed by DVF filtering the individual HRTFs measured at 1.5 m with a 5° angular resolution. The considered directions and distances for the experiments were seven azimuths in front of the listener ranging from -90° (directly in front of the listener's left ear) to $+90^\circ$ with a 30° resolution, and four distances chosen on a logarithmic scale : 0.13 m, 0.25 m, 0.50 m and 1.00 m. This means that there were 27 possible source positions (7 directions \times 4 distances). The sound for each source position was heard four times.

In the angular accuracy evaluation, a single virtual sound was presented to the listener. The listener was free to listen to it again as many times as desired before making a judgement of the direction of the source. The listener answered the direction where he/she believed the source came from all possible positions, before moving on to the next trial. In the distance accuracy evaluation, first a reference virtual source located at 1 m was presented. Then, the target virtual source was presented, located at one of the four possible distances. Direction of the target source was always the same as that of the reference source. The listener was free to listen to both sounds as many times as he desired, before making a judgement of the distance of the target source. Both a slider and a text input were available for the listener to answer at which distance he believed the source distance was. The slider was ranged from 1 cm to 100 cm with 1 cm resolution. If the listener believed the sound source came from further away than 100 cm, he could answer via the text input.

For the distance evaluation experiment, the effect of the intensity cue for distance was considered. In one condition, this cue was included, while in another condition, this cue was excluded. This was done in order to study judgement of distance independently from the dominant judgement of sound intensity. To do this, the source power was changed accordingly to distance so as to compensate for the inverse square law presented in section 1.2.1.



(a) View of the interface used during the angular localization accuracy experiment. The listener selected from one of the available azimuths which direction he believed the sound source was presented from.



(b) View of the interface used during the distance localization accuracy experiment. The listener heard a reference and target sound and answered using a slider or text input the distance he believed the target sound source was presented from.

Figure 2.9: Matlab Interfaces used during the localization experiments: (a) the interface for direction evaluation, (b) the interface for distance evaluation.

2.5.2 Results

The results obtained for angular and distance accuracy were gathered and averaged for all listeners. The results are presented as a perceived dimension to presented dimension scale in Fig. 2.10 and 2.12. The Pearson's correlation coefficient between perceived dimension and presented dimension is also calculated and illustrated in Fig. 2.11 and 2.13. This correlation coefficient estimates a global accuracy value on a scale of 0 to 1. A correlation coefficient of 1 corresponds to a perfect matching of perceived and presented dimension.

Azimuth estimation

A consistent overestimation of the laterality of sources presented from diagonal azimuths ($\pm 30^\circ$, $\pm 60^\circ$) is observed in Fig. 2.10. This overestimation is clearly observed for closer sounds, up to 45° estimation for sounds at 0.13 m. This can be explained by the increased effect of the head shadow. As sources are closer to the listener's head, the interaural level differences become a more important cue which results in the impression of a more lateral position of the presented sound source. This suggests that the listener is not used to judge the direction of sound sources from within peripersonal space.

A two way analysis of variance (ANOVA) is conducted for the parameters of Distance (four distances) and Direction (seven azimuths). Results show significance for direction ($F(6, 54) = 255$, $p < .001$) and interaction (Distance \times Azimuth, $F(18, 162) = 1.86$, $p < .05$), but not for distance ($F(3, 27) = 0.44$, $p = .73$). The results for interactions of distance on azimuth localization confirm the effects of distance on direction perception.

The results imply that almost all listeners can judge angular of the sound source accurately. Overall error between perceived and presented scales is small (absolute: 16° , RMS: 18°). This error may be lower if increasing the resolution of possible angles proposed for the response. The correlation coefficient between presented azimuth and perceived azimuth is also very high (Fig. 2.11), with an average of 0.917. This indicates good localization accuracy.

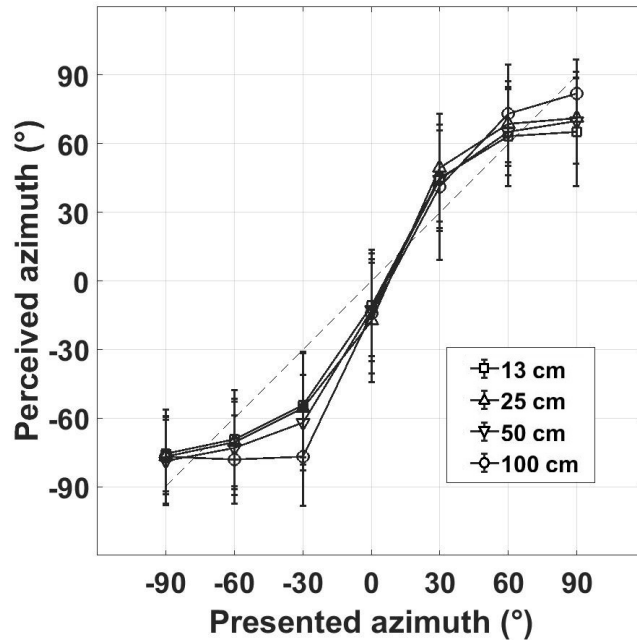


Figure 2.10: Average results for the direction accuracy experiment. Each point corresponds to the average estimated direction among all test subjects for this configuration. Vertical bars correspond to one standard error. Results for the different considered distances are plotted on the same graphic.

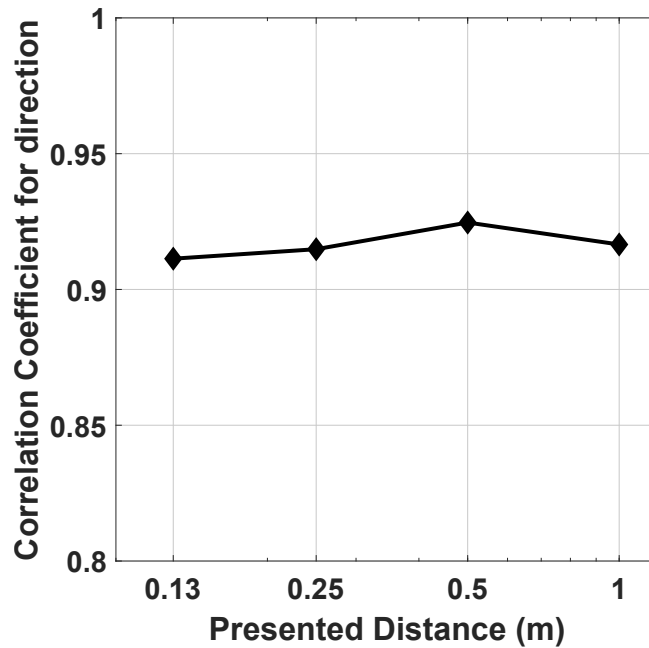


Figure 2.11: Average correlation coefficients for the direction accuracy experiment. For each distance, the pearson's correlation coefficient between perceived azimuth and presented azimuth is calculated.

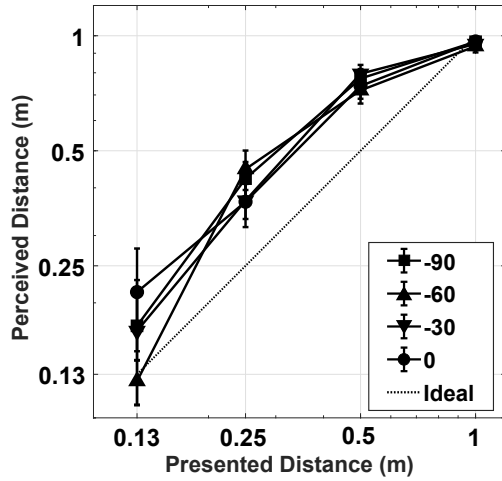
Distance estimation

A consistent overestimation of the distance of sources is observed, regardless of whether the intensity cue was included or not (Fig. 2.12). The overestimation was greater for conditions excluding intensity. This overestimation is consistent with the human average distance estimation scale established by Zahorik *et al.* [42] presented in section 1.2.4.

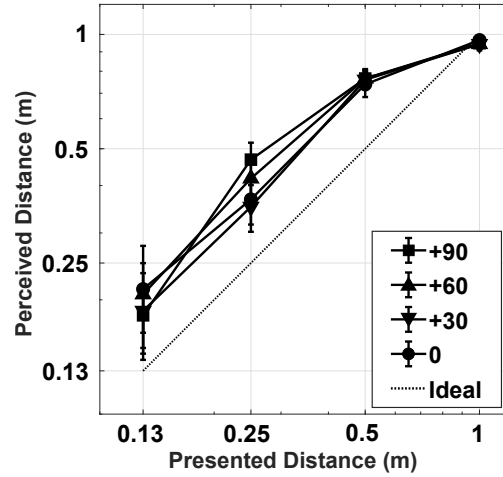
A three way analysis of variance is conducted for the parameters of Distance (four distances), Direction (seven azimuths) and Condition (with/without intensity cue). Results show significance for Distance ($F(3,27) = 223, p < .001$) and Condition ($F(1,9) = 37.5, p < .001$), but not for Direction ($F(6,54) = 0.2, p = .97$). Interactions were significant between Direction and Distance ($F(18,162) = 2.15, p < .01$), Condition and Distance ($F(3,27) = 11.6, p < .001$), and between all three factors ($F(18,162) = 2.18, p < .01$) but not between Direction and Condition ($F(6,54) = 1.19, p = .32$).

These results indicate the effects of azimuth on distance localization in both conditions. Distance of sources coming from the interaural axis ($\pm 90^\circ$) are estimated more accurately than sources coming from the median plane (0°). The interaural differences are therefore more consistent in localizing distance than the auditory parallax and spectral modifications. Despite the consistent overestimation of distance, results for correlation, presented in Fig. 2.13, are suitable. The average correlation value with intensity is 0.85, and without intensity, 0.64. The mean absolute error (with intensity : 0.13 m, without intensity : 0.31 m) and root mean square (with intensity : 0.16 m, without intensity : 0.35 m) are also low.

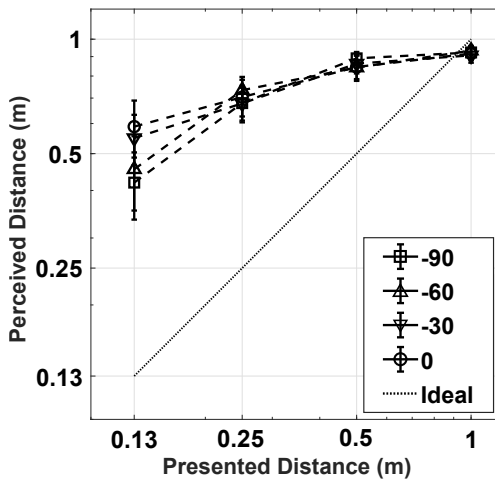
Regarding distance perception, individual differences were important in conditions excluding the intensity cue. Several listeners struggled to perceive distance for sound sources presented from the median plane. This can be explained by the little information provided by auditory parallax to the listeners for sound localization when sources were on the median plane, leading to great uncertainties in judgement of distance. Some listeners reported a shift in the sound image rather than a change in perceived source distance.



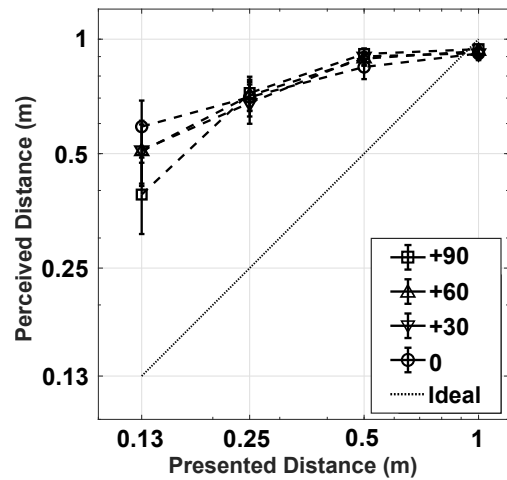
(a) Conditions including the intensity cue for sources on the left side.



(b) Conditions including the intensity cue for sources on the right side.



(c) Conditions excluding the intensity cue for sources on the left side.



(d) Conditions excluding the intensity cue for sources on the right side.

Figure 2.12: Average results for the distance accuracy experiment: (a) and (b) conditions with the intensity cue, (c) and (d) conditions without the intensity cue, (a) and (c) results for sources on the left side, (b) and (d) results for sources on the right side. Each point corresponds to the average estimated distance among all listeners for this configuration. Vertical bars correspond to one standard error.

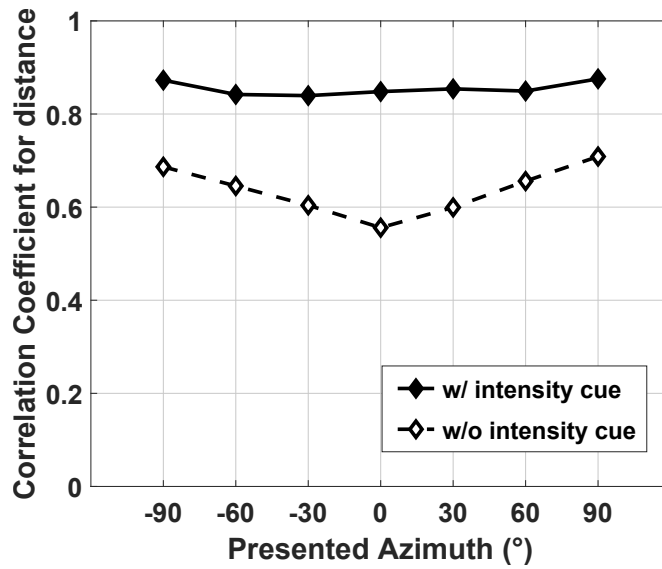


Figure 2.13: Average correlation coefficients for the distance accuracy experiment. For each direction and condition, the pearson's correlation coefficient between perceived distance and presented distance is calculated. Filled and open symbols respectively correspond to results with and without the intensity cue.

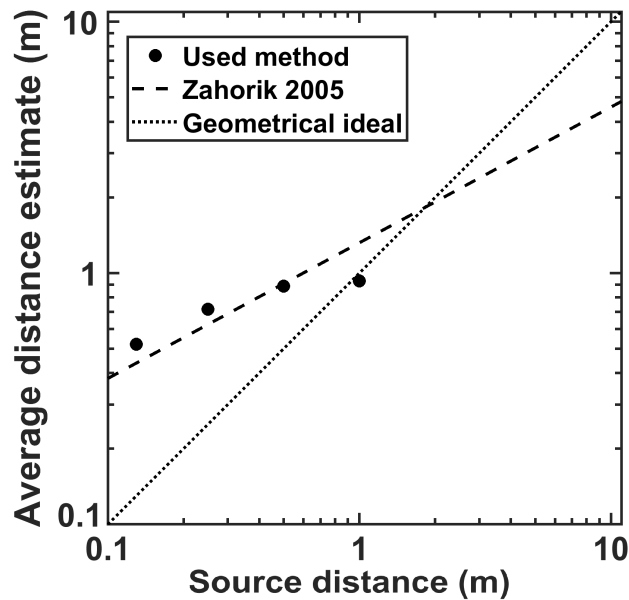


Figure 2.14: Schematic of the average perceived distance to presented distance scale. The full line represents the results gathered from 84 datasets by Zahorik *et al.* in 2005. The filled circles represent the distance perception scale averaged over azimuths obtained in this experiment.

2.5.3 Discussion

The high correlations between perceived and presented dimension are promising for the use of this method. Indeed, if comparing to previous studies using virtual sound sources excluding the intensity cue [66, 67] such as presented in the table below (Table 2.1), this method leads to overall better correlations. The correlation coefficients for distance estimation without intensity for this study range from 0.57 to 0.78, whereas correlations for past studies range from 0.05 to 0.6. In addition, if comparing to the average human distance estimation accuracy [42] illustrated in Fig. 2.14, the results for this method are acceptable.

However, if comparing to the distance estimation accuracy for real sources [35, 66], this method remains slightly inaccurate.

	Virtual sources			Real sources
Study	Brungart & Simpson 2001 [66]	Qu <i>et al.</i> 2009 [67]	This method	Brungart <i>et al.</i> 1999 [22]
Correlation coefficient	0.05~0.6	0.2~0.6	0.57~0.78	0.4~0.85

Table 2.1: Overview of results gathered from previous localization studies when excluding the intensity cue.

2.5.4 Evaluation conclusion

To conclude this perceptual evaluation of DVF filtered individual HRTFs, the distance and direction estimation accuracy is comparable to that of the average human distance estimation accuracy. It also gives better results than other virtual source spatialization methods that were previously used in psychoacoustic experiments, indicating that these HRTFs can be used in psychoacoustic experiments involving sound spatialization.

2.6 Chapter conclusions

Measuring individual HRTFs for several distances is a time-consuming, complex, and still perfectible task. It can be especially complex and inaccurate for near field sound sources. Therefore, the use of DVFs on HRTFs to approximate near-field virtual source positions was evaluated. Numerical results suggested that DVF filtered HRTFs are less accurate for the closest distances than for further distances, and that the errors appear at the contralateral ear for high frequencies of sound. However, the result of a perceptual experiment using the considered method showed distance localization accuracy comparable to the average human distance estimation accuracy. The considered method was therefore used for virtual positioning of sound sources in the next chapters. Furthermore, the obtained localization scale for each listener was extracted for use in the following chapters.

Chapter 3

Effects of sound source distance on spatial auditory attention to speech stimuli

3.1 Chapter objectives

This chapter aims to investigate the effects of sound source distance on stimulus-driven auditory spatial attention. Some researchers reported that sounds presented from within the peripersonal space lead to a higher activity in the brain [45, 46, 47]. This increase is induced by particular auditory and multimodal processes activated when the sounds are presented within peripersonal space [45, 48]. I hypothesize that this particular process of peripersonal sounds could affect auditory attention. When listeners search for a target sound, the closest sources would attract more attention.

Firstly, the effect of sound source distance on auditory attention is investigated. Reaction times (RT) in a target search task are used as a measure of auditory attention. The listener is instructed to look for a particular target word. A distracting background speech sound is presented simultaneously. By varying the sound source distance of both target and background, RT would be changed. This change is considered to reflect the amount of attention. If the hypothesis is appropriate, listeners would respond to the target sound faster when the sound is more salient.

Next, the effect of the distance separation between target and competing sound sources on RT is investigated. Shinn-Cunningham et al. [49] and Brungart and Simpson [50] showed in their separate studies that distance separation of speech sound sources benefits intelligibility. The result is a reduced masking effect of competing sounds on a target sound. By investigating RT in this target search task, the benefits of the separation on auditory attention are analyzed.

These two conditions are selected to investigate the effects of individual distance perception cues separately. Section 3.2 presents the design of the experiment conducted in this chapter. Section 3.3 presents the results obtained from this experiment and whether the hypothesis that distance of sound sources affects auditory spatial attention is supported or not. In Section 3.4 these results are considered and are discussed by relating to previous studies of auditory spatial attention. Finally, in section 3.5, conclusions on this chapter are presented.

3.2 Experiment design

The experiment consists of a target (T) sound search task in a competing background (B) sound. The listeners are instructed to respond to the target sound as fast as possible when they noticed its presentation. The reaction time is defined as the time delay between the presentation of the target stimulus and the input response of the listeners. Moreover, by changing the positions of the target and background sound sources, the relationship between observed RT and target sound distance is investigated.

3.2.1 Test participants

Nine young students (8 male, 1 female, ages 21-24, average age 23) with normal hearing acuity participated in this experiment. All were from the Graduate School of Information Sciences, Tohoku University. All listeners had also participated in the localization accuracy experiment presented in section 2.5.

3.2.2 Apparatus and stimuli

The stimuli were presented with the same experimental apparatus and in the same environment as in the sound source localization accuracy experiment in section 2.5. The listener was provided with a gamepad plugged to the experiment computer. The listener was instructed to respond to the target sound by pressing the buttons of this gamepad. The mechanical and electrical delay of the gamepad were assumed to be negligible compared to the human reaction time.

Stimuli

The target sound was a single 4 mora word (A-DO-RI-BU, アドリブ) chosen from the Japanese word corpus FW07 [68]. It was uttered by a male speaker and lasted 610 milliseconds. It was fixed for all trials of the experiment. The background sound was a superposition of meaningless speech sounds. To create this sound, six streams of meaning-

less speech were created by connecting words sequentially from the FW03 Japanese word corpus [69] uttered by the same speaker as the target sound. All words in this stream were different from the target sound. These six streams of sound were then added with random delay. The result is a meaningless speech sound resembling that of several speakers talking simultaneously. The length of this background sound was 7 s. Once the background sound finished presenting, the current trial was terminated. Between each trial, a short break of 1 s was done for computing the next stimuli.

The background sound was presented to the listener. After a random time delay ranging from 2 to 5 s, the target sound was presented. The listener was therefore not capable of predicting when the target sound would be presented. The target to background sound level ratio (TBR) of the speech sounds at the center of the head was set to be 0 dB. Virtual sound sources were obtained by convolving the monophonic sounds with the listener's 512 point DVF filtered HRTFs for the desired position.

Conditions and spatial configurations

Three azimuths and four distances were considered in this experiment. The three possible azimuths were -90° , 0° and $+90^\circ$. They were chosen so as to separate distance perception cues. Indeed auditory parallax could not be used for distance perception on the interaural axis. On the other hand, interaural differences could not be used for distance perception on the median plane. The distances followed the same logarithmic distance scale as used in the sound source localization accuracy test. This resulted in 12 possible positions of the target and background sources (4 distances \times 3 azimuths).

Two conditions were investigated. In "Same-Distance" condition, the target sound source and the background sound source were presented from the same position. In each trial, this position was one of the twelve possible positions. In "Distance-Separation" conditions, the background sound source was always presented at a distance of 1 m. The target sound source was presented from one of the four possible distances. The target sound was always presented from the same direction as the background sound. These conditions and spatial configurations are illustrated in Fig. 3.1.













In the experiment, the effect of intensity cue was also investigated. For sounds excluding the intensity cue, no difference in source intensity between the target and the background could be perceived. In contrast, for sounds including the intensity cue, the distance of the sound source affected the perceived intensity. With this choice of spatial configurations and

conditions, the individual bottom-up contribution of distance perception cues to auditory attention could be investigated.













3.2.3 Experimental procedure

The listeners were first instructed with the task. Then, the target word to search for was instructed to them and presented to them via headphones. The listener was then seated in front of a computer screen indicating the current trial number. Instruction was given to the listener to keep his/her head straight during the session. No visual feedback on the positions of the sound sources or the correct response was given. During one trial, if the gamepad input occurred before the target sound presentation or after a time delay of more than 2 s, this trial was repeated once more later in the session. Before the beginning of the experiment, the listener was trained for the experiment task on twenty trials picked at random from the possible conditions and source positions. During the experiment, all conditions and sound source positions were randomly mixed.

The sound sources presented from each position were heard ten times in each experimental condition, and each sound intensity condition. “Same-Distance” and “Distance-Separation” conditions were identical when both the target sound source and the background sound source were presented from 1 m, regardless of whether the intensity cue was provided or not. Therefore, this spatial configuration was not repeated for each condition. This resulted in 30 trials in which both target and background sound were presented at 1 m (3 azimuths \times 10 repetitions). Every other target and background sound source position resulted in 180 trials in each condition (3 azimuths \times 3 distances \times 2 intensity conditions \times 10 repetitions). This resulted in 390 trials per experiment in total. In order to preserve the listener’s attentive capabilities, the experiment was divided into five sessions of equal length (78 trials, 10 minutes), with a short rest between each session.

	-90°	0°	$+90^\circ$
1 m	\bar{B} 	\bar{B} 	 \bar{B}
0.5 m	\bar{B} 	\bar{B} 	 \bar{B}
0.25 m	\bar{B} 	\bar{B} 	 \bar{B}
0.13 m	\bar{B} 	\bar{B} 	 \bar{B}

(a) Schematic of the configurations considered in the Same-Distance conditions. The target (T) and the background (B) are presented from the same position. This position is chosen from one of all possibilities for distance and azimuth.

	-90°	0°	$+90^\circ$
1 m	\bar{B} 	\bar{B} 	 \bar{B}
0.5 m	B T 	B T 	 T B
0.25 m	B T 	B T 	 T B
0.13 m	B T 	B T 	 T B

(b) Schematic of the configurations considered in the Distance-Separation conditions. The target (T) and the background (B) are presented from the same direction. The distance of the background is fixed at 1 m. The distance of the target is one of all distance possibilities.

Figure 3.1: Configurations considered for both Same-Distance conditions (a) and Distance-Separation conditions (b). The azimuth of the sources is one of three azimuths : -90° , 0° , $+90^\circ$. The distances of the sources is picked from one of four : 1 m, 0.5 m, 0.25 m and 0.13 m.

3.3 Results

3.3.1 Analysis method

The average RT for each listener was calculated in all conditions and target sound source positions. The mean RT over all listeners for each condition was then obtained for sources on the interaural axis and on the median plane separately in Fig. 3.2 and 3.3.

3.3.2 Average reaction time

A three-way analysis of variance (ANOVA) with factors Condition (four: With/Without intensity \times Same-Distance/Distance-Separation), Azimuth (three), and Distance (four) was conducted. Results show a significant effect for Distance ($F(3, 24) = 44.2$, $p < .001$) and Condition ($F(3, 24) = 78$, $p < .001$) but not for Azimuth ($F(2, 16) = 0.84$, $p = .45$). Interactions were significant for Condition \times Distance ($F(9, 72) = 19.7$, $p < .001$), but not for Azimuth \times Distance ($F(6, 48) = 1.93$, $p = .095$), Azimuth \times Condition ($F(6, 48) = 0.29$, $p = .94$) or three-way interactions ($F(18, 144) = 0.59$, $p = .90$).

These results suggest the importance of distance of target sound source in attentive target detection tasks. The processes involved in this detection task depend on source distance. The overall tendency is that the closer sounds induce the fastest responses. This confirms the hypothesis that closer sounds capture auditory attention.

Interactions between Condition and Distance show effects of Condition for 0.13 m ($F(3, 96) = 81$, $p < .001$), 0.25 m ($F(3, 96) = 60$, $p < .001$), 0.5 m ($F(3, 96) = 29$, $p < .001$) but not for 1 m ($F(3, 96) = 0$, $p = 1$). All conditions were identical for 1 m, explaining the absence of effect for this distance. In addition, effects of Distance are significant for Distance-Separation with intensity ($F(3, 96) = 98$, $p < .001$) and without intensity ($F(3, 96) = 8.8$, $p < .001$), and for Same-Distance with intensity ($F(3, 96) = 13$, $p < .001$) but not without intensity ($F(3, 96) = 1.7$, $p = .16$). This indicates that the task is responded to differently in Same-Distance condition and in Distance-Separation conditions. These two separate conditions are therefore analysed separately.

Same-Distance

Results for sounds on the interaural axis and on the median plane are analyzed separately. On the interaural axis, a three-way ANOVA with factors Intensity (With/Without intensity), Azimuth ($\pm 90^\circ$) and Distance (four) is conducted. On the median plane, a two-way ANOVA with factors Intensity (With/Without intensity) and Distance (four) is conducted.

Results on the interaural axis suggest a consistent effect of sound source distance on RT ($F(3, 24) = 15, p < .001$). The closer the sounds were presented to the listeners, the faster the RT was. This was true regardless of providing the intensity cue. Indeed, interactions between Intensity and Distance were not significant ($F(1, 8) = 1.87, p = .16$). The RT reduction was at best of 58 ms in conditions including intensity and 25 ms in conditions excluding intensity. This suggests that the absolute distance of sound sources have a great effect on auditory processes. Differences between condition with intensity and without intensity ($F(1, 8) = 9.26, p < .05$) could also be observed. No difference between -90° and $+90^\circ$ was observed ($F(1, 8) = 0.01, p = .91$). Processes seem to be identical for sounds coming from the left or the right.

Results on the median plane show no significant effect of the Distance ($F(3, 24) = 0.45, p = .72$) or Intensity factors ($F(1, 8) = 3.39, p = .102$).

Comparison between the three directions leads to the conclusion that interaural differences are used consistently in this attentive target detection task when target and background are presented from the same position. In contrast, auditory parallax alone could not be used consistently. This difference in results for the median plane and interaural axis can be explained by the difficulty of distance judgement in the median plane, as illustrated in Chapter 2. Therefore, the differences between source distances on the median plane could not lead to differences in attention.

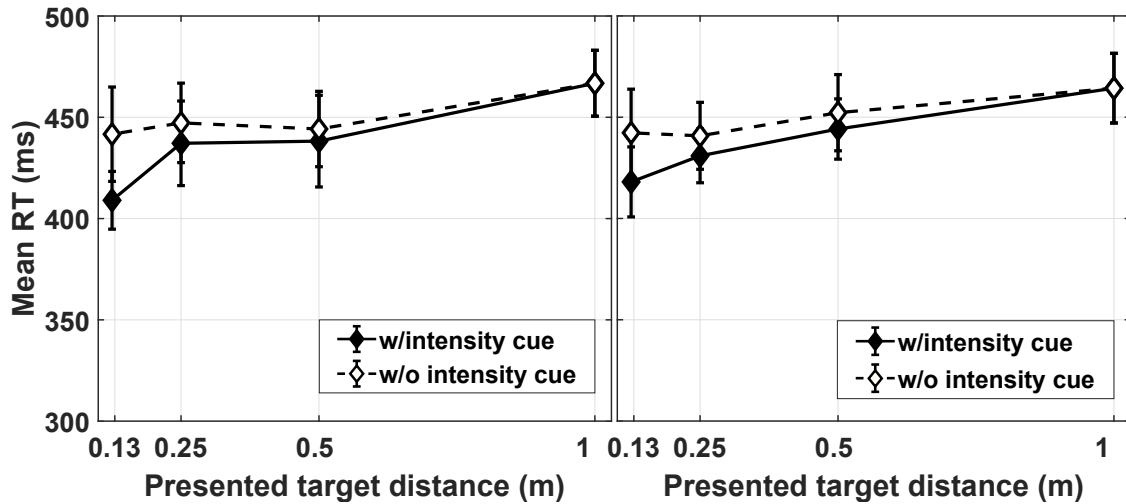
Distance-Separation

Similarly to Same-Distance conditions, in Distance-Separation conditions, results for sounds on the interaural axis and on the median plane are analyzed separately.

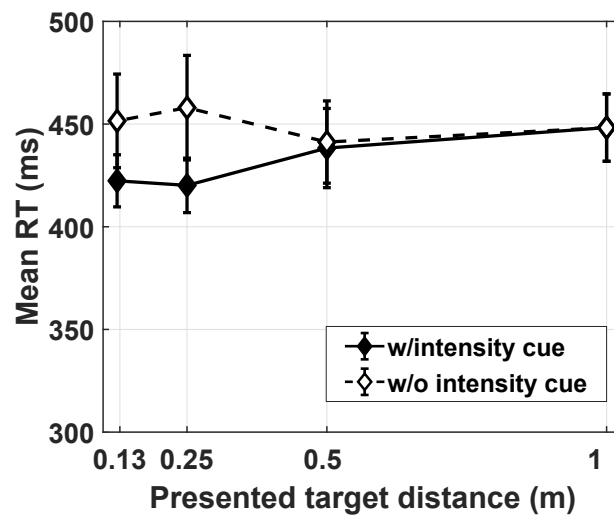
Results on the interaural axis suggest a consistent effect of target-background distance separation on RT ($F(3,24) = 54.7, p < .001$). The RT reduction was at best of 126 ms in conditions including intensity and 50 ms for conditions excluding intensity. The closer the target sound was presented to the listeners, the faster the RT was, regardless of the intensity cue. Both Intensity ($F(1,8) = 155, p < .001$) and Distance \times Intensity interactions ($F(1,8) = 39, p < .001$) were significant. The simple main effect of Distance was significant both with intensity ($F(3,48) = 90.8, p < .001$) and without intensity ($F(3,48) = 10.8, p < .001$). Results of a multiple comparison (Ryan's method) show that all differences in distance were significantly different with intensity ($p < .01$). Without intensity, only the difference between 0.25 m and 0.5 m was non significant ($p = .49$). No difference between -90° and $+90^\circ$ was observed ($F(1,8) = 0.65, p = .44$). Processes here again seem to be identical for sounds coming from the left or the right.

Results on the median plane show significant effects of Distance ($F(3,24) = 12.5, p < .001$), Intensity ($F(1,8) = 127, p < .001$) and Distance \times Intensity interactions ($F(3,24) = 19.5, p < .001$). Distance had a significant effect with intensity ($F(3,48) = 27, p < .001$) but not without intensity ($F(3,48) = 1.02, p = .39$). Results of a multiple comparison in conditions with intensity show that only the difference between 0.13 m and 0.25 m is non-significant ($p = .62$). This difference with the interaural axis leads to the conclusion that when competing sound sources are separated in distance, only the source intensity and the interaural differences results in differences in attention.

A larger contribution of target source distance was observed in these Distance-Separation conditions than in Same-Distance conditions ($F(1,8) = 86, p < .001$). This can be explained by the lower effects of masking by the background sound when sources are separated in distance. These tendencies are consistent with results reported by Shinn-Cunningham [49] and Brungart & Simpson [50].

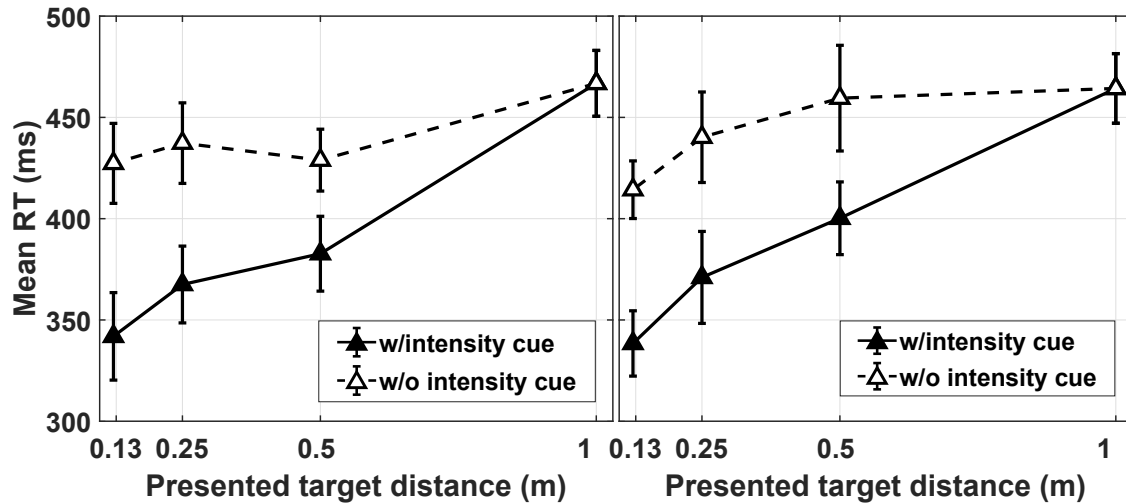


(a) Results on the interaural axis. On the left figure, -90° . On the right figure, $+90^\circ$.

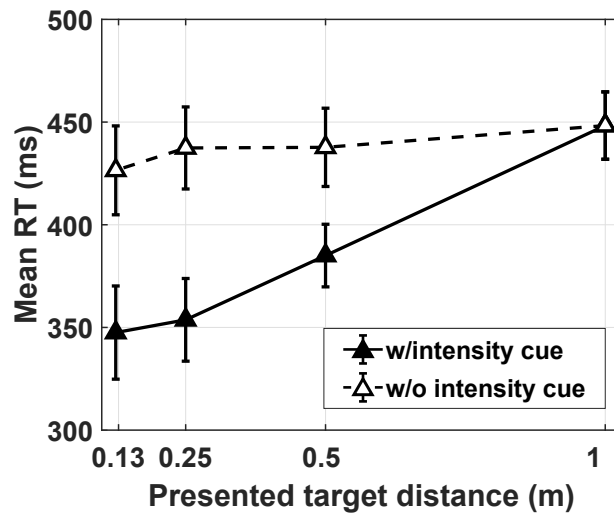


(b) Results on the median plane (0°).

Figure 3.2: Reaction time averaged for all listeners as a function of presented distance of the target sound source in Same-Distance conditions. The results for the interaural axis (a) and the median plane (b) are shown separately. Error bars indicate standard error. The plots with black markers and full lines correspond to sounds including the intensity cue, while plots with white markers and dashed lines correspond to sounds excluding the intensity cue.



(a) Results on the interaural axis. On the left figure, -90° . On the right figure, $+90^\circ$.



(b) Results on the median plane (0°).

Figure 3.3: Reaction time averaged for all listeners as a function of presented distance of the target sound source in Distance-Separation conditions. The results for the interaural axis (a) and the median plane (b) are shown separately. Error bars indicate standard error. The plots with black markers and full lines correspond to sounds including the intensity cue, while plots with white markers and dashed lines correspond to sounds excluding the intensity cue.

3.3.3 Normalized reaction time as a function of perceived distance

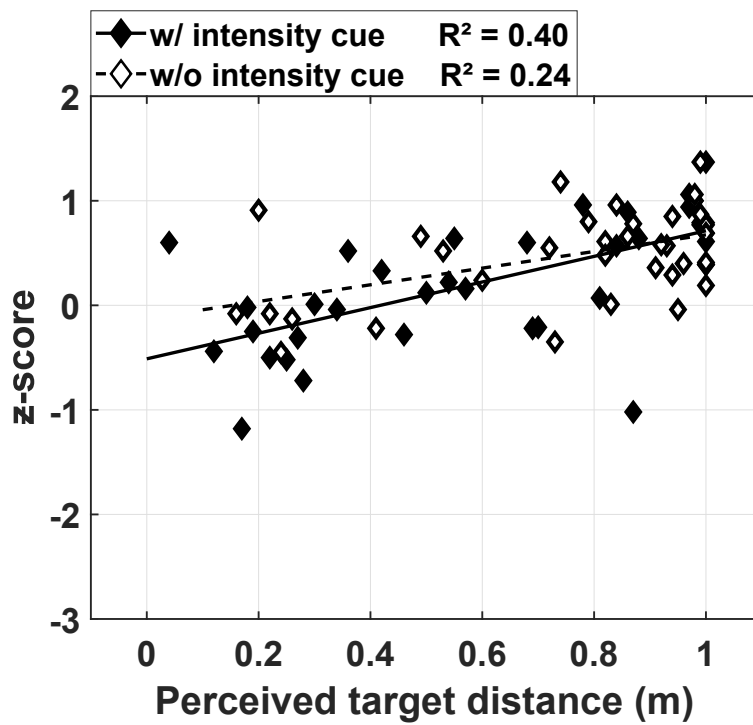
In the previous section, the mean RT for all listeners was analyzed. However, when analyzing the listeners' individual average RT, reaction speed strongly depends on individual listeners. Some listeners have a faster average reaction time than others. Therefore, in order to compare individual listeners' RT results, each listener's average RT was normalized. In this section we normalize individual RT scores using a z-score transformation, using Equation (3.1) :

$$z\text{-score} = \frac{RT - \mu}{\sigma} \quad (3.1)$$

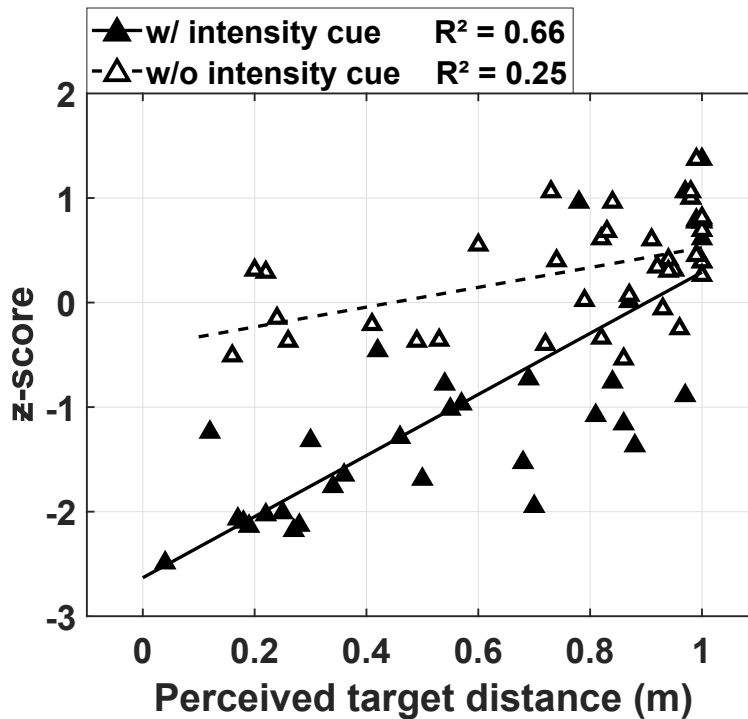
where μ is the listener's average RT over all conditions and σ is the standard deviation of the listener's RT over all conditions. The result of this transformation follows the same monotony as the individual RT : when the RT is faster, the z-score is smaller. Using this transformation, individuals' RT behavior to source distance can be compared equally.

Then, the same individual's perceived distance scale from section 2.5 is extracted this individual's z-score to perceived distance is plotted. This plot for sounds on the interaural axis is represented in Fig. 3.4. This results in a scatter plot of the average behavior of RT as a function of perceived distance for an average listener. After this conversion, I aim to analyze the effects of perceived source distance, rather than presented source distance.

At the beginning of the analysis, a linear regression of the resulting scatter plot was calculated and presented on the figures. R^2 values are low (0.24 and 0.25) in conditions excluding intensity but are acceptable when including intensity (0.40 and 0.66). This discrepancy highlights the dominant effect of sound source intensity in distance-related experiments. Overall, the values are scattered around the linear model. However, in Distance-Separation conditions, it can be observed that the initial small separation in distance leads to larger reduction in RT than further separation. The relationship between the average z-score and average perceived distance is plotted in all conditions in Fig. 3.5. This figure highlights that the initial small separation between target and background results in large RT reduction for the Distance-Separation conditions.

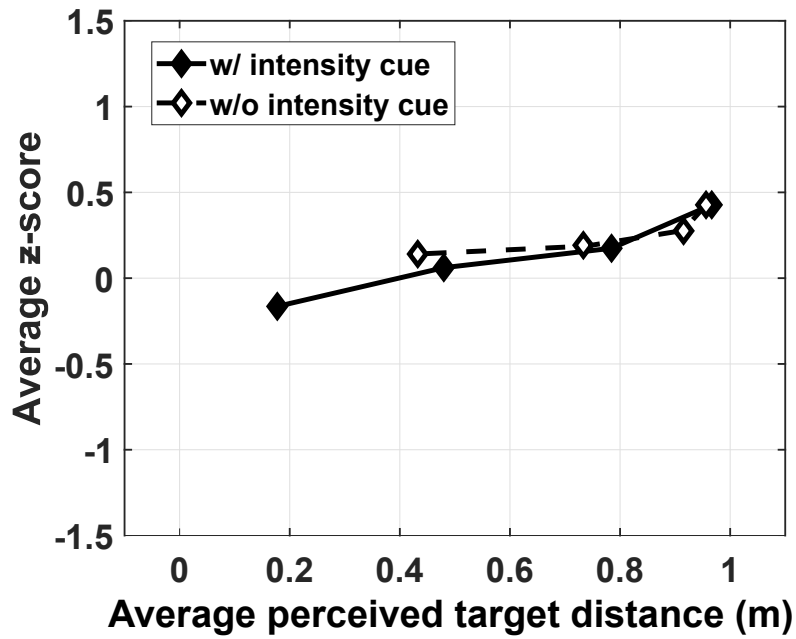


(a) Same-Distance

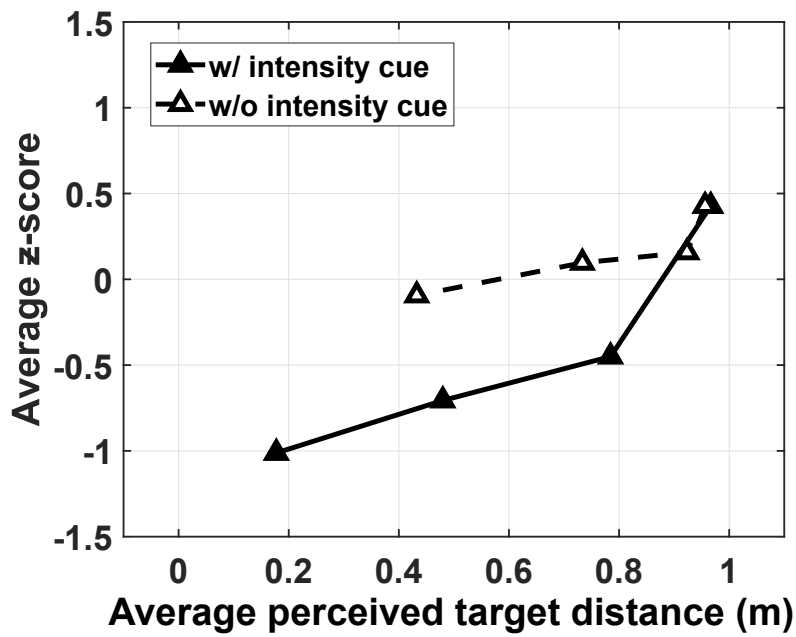


(b) Distance-Separation

Figure 3.4: Individual z-score to perceived distance. Same-Distance conditions (a) and Distance-Separation conditions (b) are represented. Sounds including (black marks, full line) and excluding intensity (white marks, dashed lines) are also analyzed. In these figures, a linear fitting is applied to the data.



(a) Same-Distance



(b) Distance-Separation

Figure 3.5: Relationship between average z-score and average perceived distance. Same-Distance conditions (a) and Distance-Separation conditions (b) are represented. Sounds including (black marks, full line) and excluding intensity (white marks, dashed lines) are also analyzed. In Distance-Separation conditions, a larger reduction of z-score can be observed in the initial separation of target and background than in further separation.

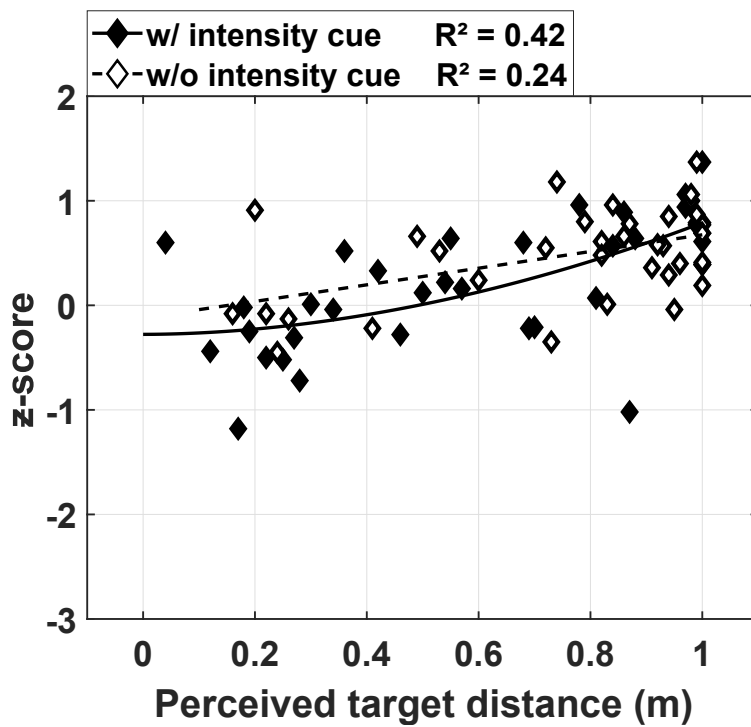
Taking this consideration into account, a power function model is proposed in Fig. 3.6. The z-score is fitted by following equation :

$$z\text{-score} = a \times (r')^b + c \quad (3.2)$$

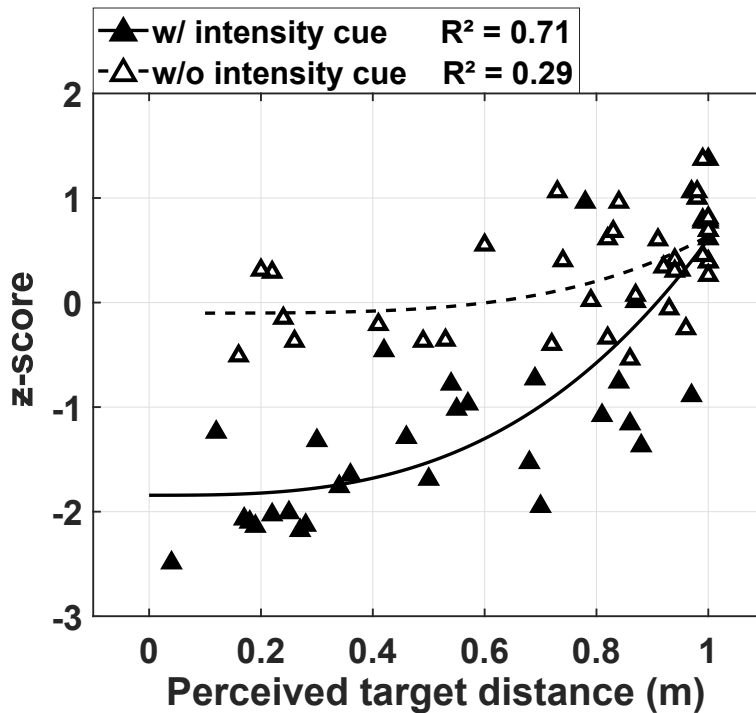
where r' is the perceived distance and a , b and c are fitting parameters. As a result, the model predicts a rapid reduction of RT for the initial distance separation of target and background sources. The R^2 values for each fitting model is presented in Table 3.1. The small increase in Distance-Separation condition, as well as similar observations reported by Brungart & Simpson [50], could justify the use of the power model.

	Same-Distance		Distance-Separation	
	w/ intensity	w/o intensity	w/ intensity	w/o intensity
Linear model R^2	0.40	0.24	0.66	0.25
Power model R^2	0.42	0.24	0.71	0.29

Table 3.1: R^2 values for both linear and power fitting models of z-score to perceived distance. A small increase of R^2 value is observed when fitting the data with a power model.



(a) Same-Distance



(b) Distance Separation

Figure 3.6: Individual z-score to perceived distance. Same-Distance conditions (a) and Distance-Separation conditions (b) are represented. Conditions including (black marks, full line) and excluding intensity (white marks, dashed lines) are analyzed. Here, a power fitting is applied to the data, leading to better R^2 values.

3.4 Discussions

3.4.1 Effects of peripersonal space in virtual presentation of sounds

During the experiment, the listener attended to a target word. While attention is on the nature of the word, the observed tendency of the reduction of RT with distance indicates that the presentation of the sound in peripersonal space enhances auditory process. In Same-Distance condition, no benefits of sound source separation in distance were included. However, under this condition, the closest sounds resulted in faster RT when presented on the interaural axis, regardless of the source intensity. This suggests that the absolute distance of sound sources impacts how much the perceived sounds capture attention. This can also be explained by the nature of auditory attention and peripersonal space. As explained by Scharf in 1998 [6], the auditory system is “an excellent early warning system”, and auditory attention selects which information is especially important for the organism. Because the position of the sound sources would be considered as dangerously close, the organism gives high priority to these closest sounds.

An explanation of the mechanisms underlying this priority of near sources could lie in the multimodal processes involved in the peripersonal space. As sound stimuli were presented from the very closest distances, multimodal neurons which are related to the tactile stimulation also activate for these auditory stimuli [45]. This finding was also true regardless of the intensity of the auditory stimuli. This multimodal activation would result in the integration of very near sound sources with knowledge of tactile contact. This integration results in the attribution of a high priority to these auditory stimuli.

The effect of distance was especially consistent for sources on the interaural axis. Two possibilities to explain this phenomenon are considered. The first is that distance perception is more precise on the interaural axis as could be seen in Section 2.5 and previous studies [22, 66], leading to more consistent effects of virtual source distance on attention. A second possible explanation could be that sources on the interaural axis are not in the visible field, leaving only the auditory system to process sounds coming from these directions. The absence of information for sources outside of the visible field would make them more urgent, resulting in higher priority given to these sources.

Whereas previous studies of auditory peripersonal space were done with real sound sources, the results in this experiment using virtual sound sources suggest also the existence of peripersonal space processes for virtual sound sources. This means that the high priority given to the very near sound sources is extracted from the sound's acoustic properties.

3.4.2 Sound stream segregation

Effects of distance on RT in Distance-Separation conditions were consistently larger than in Same-Distance conditions, even when excluding the intensity cue. In this experiment, eliminating the intensity cue also eliminated variation in target to background sound level ratio (TBR) at the center of the listener's head. The TBR gives a measure of difference in level between competing sounds, similar to the signal to noise ratio for signal processing. It is calculated using Equation (3.3) below. Although distance separation did not change this TBR at the center of the head when intensity change is eliminated, sound source separation resulted in faster processes. This supports the hypothesis proposed by Shinn-Cunningham [49] and Brungart & Simpson [50] that separation of competing sound sources in distance results in unmasking benefits, regardless of the TBR at the center of the head. Additionally, Brungart & Simpson argued that the TBR at the ipsilateral ear can not explain these benefits neither. Indeed, during a speech intelligibility test for peripersonal distances similar to those in this experiment, they controlled the TBR at the ipsilateral ear to be constant while conserving natural interaural level differences. The results for control of TBR at the center of the head and for control at the ipsilateral ear induced similar spatial unmasking benefits.

$$\text{TBR} = 20 \log_{10} \left(\frac{\text{RMS}(\text{target})}{\text{RMS}(\text{background})} \right) \quad (3.3)$$

where *RMS* is the root mean square value calculated on the length of the target word. The TBR is estimated in decibels.

Brungart & Simpson also found that the initial separation between 1 m and 0.5 m had more impact on unmasking benefits than further separation of sound sources. This result is consistent with what was found in Section 3.3.3, in which the initial separation of sound sources had more impact on RT than further distance separation.

These two main results lead to the speculation that separation of sound sources contributes to the mental sound stream segregation process, regardless of TBR. When both sounds are presented from the same position, listener would perceive sounds in one mixed auditory stream. The initial separation in distance results in the perception of two independent streams. This segregation induces faster and easier process of both individual stream. This phenomenon is similar to the contribution of angular separation of sound sources on sound stream segregation as presented in Section 1.1.1. Further distance separation then results only in the easier segregation of sound streams.

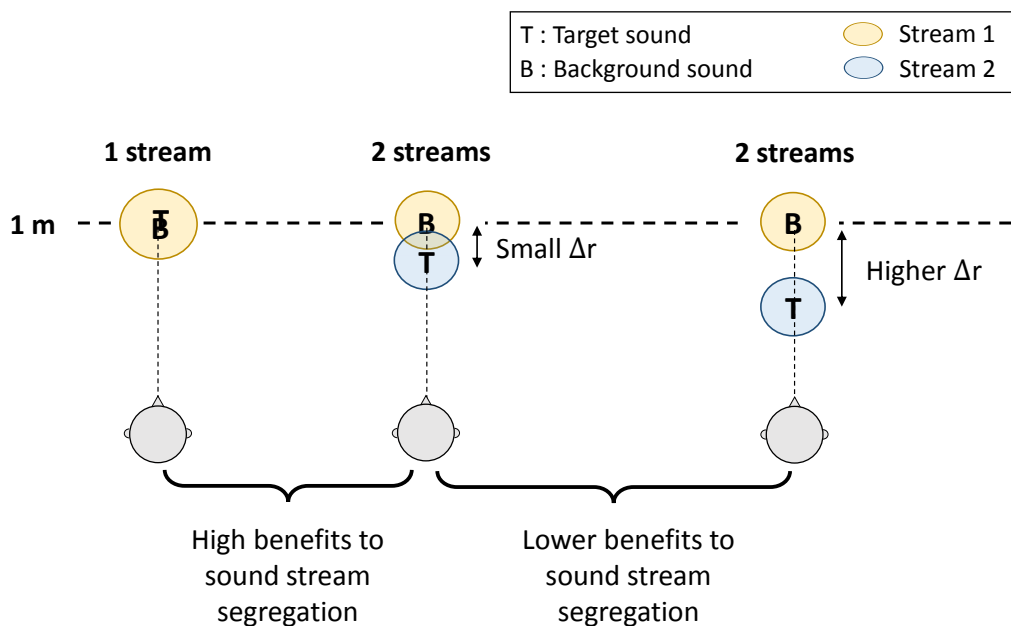


Figure 3.7: Schematic representation of benefits of distance separation on sound stream segregation. The slight separation in distance between target and background sounds leads to a difference in sound image localization. This results in a separation of perceived sound streams from one to two streams, which benefits greatly attention. Further separation of sounds in distance only makes the separation of streams clearer, benefitting less than the initial separation into two streams.

3.5 Chapter conclusions

An experiment on auditory attention was conducted for a target detection task masked by a background sound. In this experiment, no prior knowledge of the position of the target word was given. Whereas the listeners paid attention to the target word, they were not focused on the spatial localization of the sound source. Therefore, by changing the sound sources' position, the stimulus-driven auditory spatial attention could be evaluated. Results showed large bottom-up effects of distance of sound sources and distance separation of sound sources. Two main results were obtained. First, when sounds were presented from closer distances in peripersonal space, they captured attention more consistently. Furthermore, this was true regardless of the intensity cue when sounds were presented on the interaural axis. This suggests effects of peripersonal space on auditory attention. Second, distance separation of sources benefitted considerably sound stream segregation. Not all distance cues could be used consistently by the listener. Only the intensity cue and the interaural differences cue resulted in variation of reaction time.

The experiment in this chapter focused on stimulus-driven auditory attention. Therefore this leaves us with the question of task-dependent, voluntary attention. When listeners focus on a particular source distance, can the listeners process sound sources presented from this distance with high priority ? This issue is investigated in the next chapter.

Chapter 4

Top-down spatial auditory attention effects for distance of sound sources

4.1 Chapter objectives

This chapter aims to investigate the existence and capabilities of top-down spatial auditory attention to sound source distance. In other words, the research question in this chapter is whether listeners can focus on a specific sound source distance, or not. If this top-down spatial auditory attention exists, then paying attention to a specific distance would enhance auditory processes for sound sources presented from this distance. Such ability for direction of sound sources was reported by previous studies [16, 17, 18, 19, 20]. In this chapter, the effect of top-down spatial auditory attention is investigated using a similar experiment as in Chapter 3. Focus on a distance is drawn implicitly by using a probe-signal method, in which probability of the target speech sound presentation is controlled. Two distances are set to focus : far (1 m) and near (0.13 m). If the listener is capable of focusing on one of these distances then the reaction time (RT) should be faster for sounds presented from this distance.

Section 4.2 of this chapter presents the design of the experiment conducted in this chapter. Section 4.3 presents the results from this experiment and whether the hypothesis that top-down auditory spatial attention on source distance is available is supported or not. In section 4.4, these results are discussed and put in perspective with results from the field of auditory spatial attention. Finally, in section 4.5, this chapter is concluded.

4.2 Experiment design

A target sound detection task similar to Chapter 3 was conducted. However, the experimental conditions were different. The previous chapter aimed to study the effects of sound source distance on stimulus-driven auditory attention by analysing which distance perception cues are used dominantly. In this experiment, the potential top-down auditory attention for distance is focused on. Therefore, two types of conditions were considered: conditions where focus on distance is attempted and conditions where no focus on distance is attempted.

4.2.1 Test participants

In the experiment, seven young students (all male, ages 23-24, average age : 23.7) with normal hearing acuity participated. All were from the Graduate School of Information Sciences, Tohoku University. All listeners except for one (Subject 11) had also participated in the localization accuracy test presented in section 2.5 and in the experiment in Chapter 2.

4.2.2 Apparatus and stimuli

The stimuli were presented through the same experimental apparatus and in the same environment as in the sound source localization accuracy experiment in section 2.5 and the experiment in Chapter 3. The head of the listener was fixed using a chin rest.

Stimuli

The scheme of the experiment in this chapter was the same as in the previous experiment. The target sound was a four mora word chosen from the Japanese word corpus FW07 [68], (A-DO-RI-BU, アドリブ). The background sound was made of six layers of meaningless speech constructed from words extracted from the FW03 [69] Japanese word corpus. However, in order to maximize the effects of top-down auditory attention, the background sound lasted longer than in the previous chapter (more than 3 min). The

target sound was presented several times during the length of the background sound. The hypothesis is that within this time, the listener should be able to fine-tune auditory spatial attention on the target distance.

In addition, so as to further study the listener's selective capabilities, another type of stimulus was also presented. To control the difficulty of the task, distracter sounds were also presented between each target sound. Distracters were four mora words from the FW07 Japanese word corpus. These were all different from the target word and the same distracter word was not presented twice during one trial. They were all uttered by the same speaker as the target and background sounds. The length of the distracter sounds ranged between 650 ms and 1000 ms. Using these distracter sounds, the selective capabilities of auditory attention for distance can also be investigated using signal detection theory.

First, the background sound was presented. After a 500 ms time delay from the beginning of the background sound, the first distracter or target sound was also presented. Between each presentation of the target sound, either one, two, three or no distracter sounds were presented. Furthermore, the inter-stimulus interval between two consecutive sound was ranged between 750 ms and 1250 ms. The listener could therefore not predict at which time the next target sound would be presented. This was done in order to eliminate potential temporal or rhythmic cues affecting attention. The trial ended once the target sound was heard a fixed number of times from each distance.

Conditions and spatial configurations

Spatial configurations were similar to the ones in the experiment in Chapter 3. Four egocentric distances in peripersonal space : 1 m, 0.5 m, 0.25 m and 0.13 m were considered. No significant difference between results for sources presented from the left and right directions was observed in the experiment in Chapter 3. Therefore, the experiment in this chapter considered only the left (-90°) and front (0°) azimuths.

The background sound was always presented from 1 m. The distance of each distracter sound was randomized from one of the four distances. The target and distracter sounds were always presented from the same azimuth as the background sound. The intensity cue for distance was always eliminated.

Three conditions were investigated separately. The first condition was the No-Focus condition, in which no a priori knowledge of the distance of the target sound source was

given to the listener. Here, the auditory attention of the listener was not directed to a particular distance. In this condition, the target sound was presented with equal probability from each source distance. Therefore, the target sound was presented with 25% probability from each of four distances. The second and third conditions were Focus conditions, in which the listener’s auditory attention was implicitly directed to a particular distance using the probe-signal method [20]. This method can direct the listener’s attention to a specific position by controlling the probabilities of presentation of the stimuli from this position. The target sound was presented from one particular distance with 80% probability, and from each of the other three distances with $20 \div 3 = 6.66\%$ probability. The hypothesis was that the listeners would gradually expect the next target sound to be presented from the high probability distance, and therefore focus on the distance implicitly. The two distances of focus considered were set at 1 m and 0.13 m. These two distances were chosen to clearly observe the effects of peripersonal space. The probabilities of presentation of the target sound at one distance are summarized in Table 4.1.

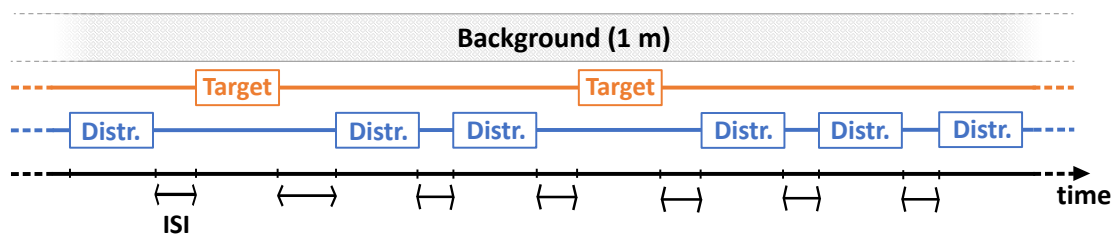


Figure 4.1: Schematic of the time course of presentation of the different stimuli. The background sound lasted more than 3 mins. The target sound is presented several times, separated by the presentation of one, two, three or no distracter sounds. Between each sound, the inter-stimulus interval (ISI) ranged from 500 ms to 1000 ms. The trial stops once the fixed amount of target sounds was presented.

4.2.3 Experimental procedure

The listeners were instructed to press a gamepad button as soon as they heard the target word. The target word was instructed and heard before the beginning of the experiment. For each azimuth and condition, the target sound was presented from each distance twelve times. Therefore, the target sound was heard 48 times (12 presentations \times 4 distances) per azimuth in No-Focus conditions. As mentioned previously, the probe-signal method was applied in Focus condition. This means that the number of target sound presentation is increased. In these conditions, for each azimuth the target was heard 144 times from the focus distance and 12 times from each of the remaining three distances. In both Focus 1 m and Focus 0.13 m conditions, the target was therefore presented 180 times (144 presentations + 12 presentations \times 3 distances) for each azimuth.

The direction of the target sound source was set to -90° and 0° . These directions were used in separate sessions, during which all sound stimuli were presented from the same azimuth. In order to preserve the listener's attentive capabilities as best as possible, Focus 1 m and Focus 0.13 m were each separated into three consecutive independent trials of equal length (60 presentations of the target, approximately 3 mins 30 s). The full experiment for one azimuth consisted of 7 separate sessions of approximately 3 mins 30 s (one session in No-Focus condition, three sessions per Focus condition). The order of each conditions were counterbalanced between all listeners. Each trial was followed with a short break. The session always started with a training session conducted in No-Focus condition, for a 1 min long trial.

Condition	Distance			
	1 m	0.5 m	0.25 m	0.13 m
No-Focus	25%	25%	25%	25%
Focus : 1 m	80%	6.66%	6.66%	6.66%
Focus : 0.13 m	6.66%	6.66%	6.66%	80%

Table 4.1: The probabilities of presentation of the target sound at one distance, in each experiment condition. In Focus conditions, the target sound is presented from a particular distance with high probability. This results in implicitly orienting the listener's attention to this distance.

4.3 Results

4.3.1 Analysis method

The hit rate and false alarm rates were investigated in this experiment. A response via gamepad input was considered as a hit if it occurred within the time interval between the beginning of a target sound and the end of the following ISI. If no input was detected within this interval, it was considered as a miss. Likewise, an input was considered as a false alarm (FA) if it occurred within the time interval between the beginning of a distracter sound and the end of the following ISI. If no input was detected within this interval, it was considered as a correct reject. These rules are summarized in Fig. 4.2.

The average RT for all individuals was calculated only for inputs considered as hits. Furthermore, in Focus conditions, the target sound was presented from the focus distance 144 times, while it was only presented 12 times for each other distance. In order to compare average values calculated on the same amount of data, the average RT for the focus distance was therefore calculated using the last 12 inputs considered as hits. It is believed that the listener's top-down attention on the focus distance was at its maximum potential for these last inputs.

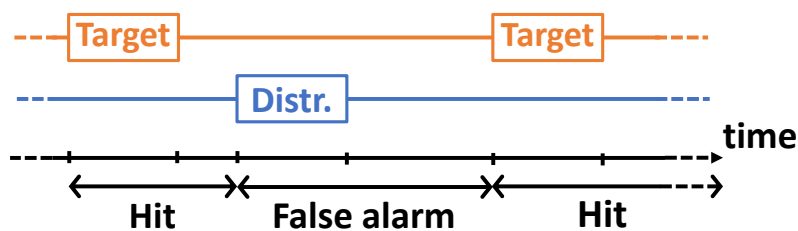


Figure 4.2: Schematic explaining the definitions of hit and false alarm in this experiment. If a gamepad input is detected within the time interval of a target sound followed by its ISI, it is considered as a hit. If no input is detected within this interval it is considered as a miss. If a gamepad input is detected within the time interval of a distracter sound followed by its ISI, it is considered as a hit. If no input is detected within this interval it is considered as a miss.

Except for one listener (Listener 2), a decrease of RT with number of presentations of the target sound at the focus distance could be observed. This tendency was observed for both Focus conditions on the interaural axis, and for Focus on 0.13 m on the median plane (Fig. 4.3). The average decrease in RT for all other listeners ranged between 5 ms and 33 ms. There were little learning effects for focus on 1 m on the median plane.

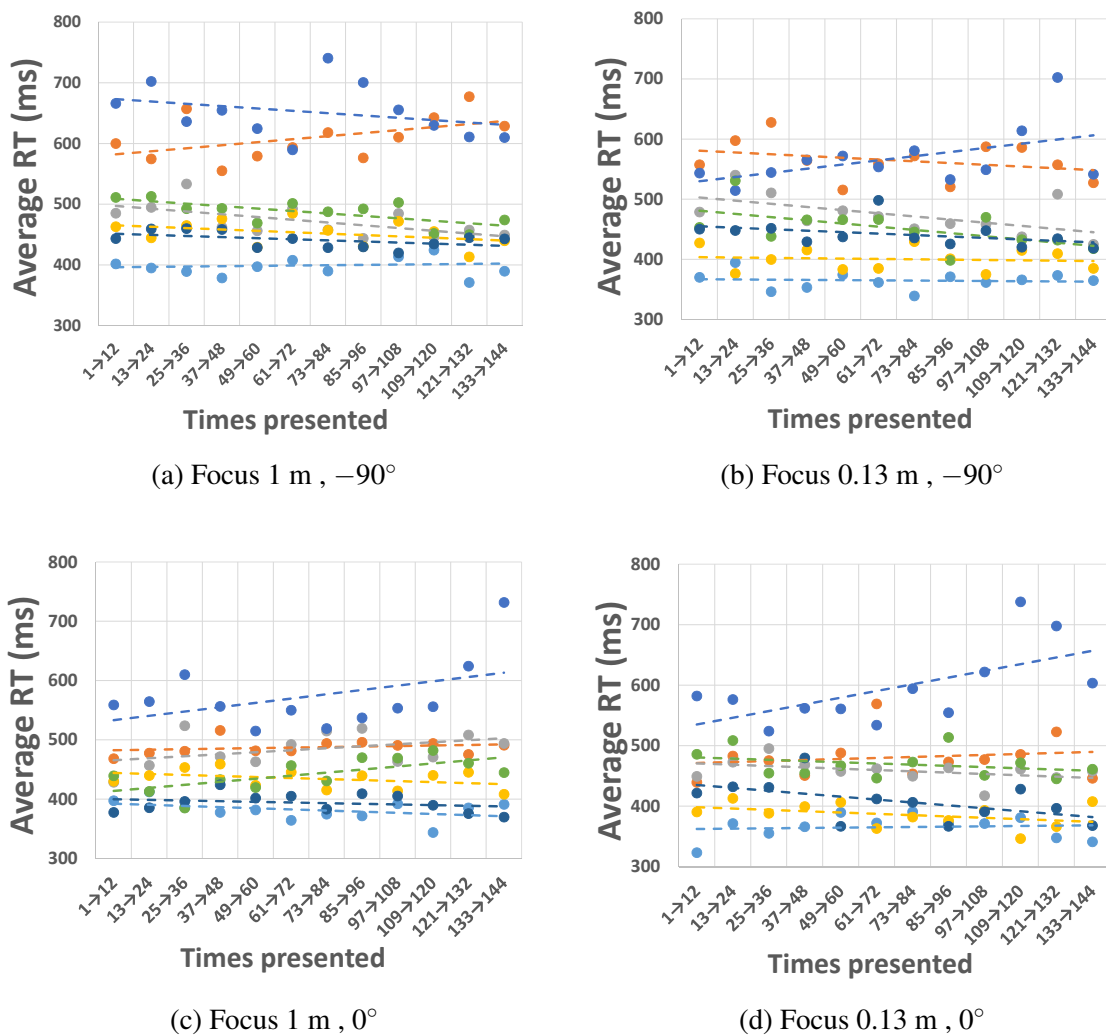


Figure 4.3: Average RT of each listener as a function of the number of target presentations from the focus distance. (a) (b) interaural axis, (c) (d) median plane, (a) (c) Focus 1 m, (b) (d) Focus 0.13 m. Each color corresponds to an individual listener. One listener (dark blue, Listener 2) presented an increase of RT towards the end of the test session, suggesting effects of fatigue due to the length of the sessions.

4.3.2 Average results

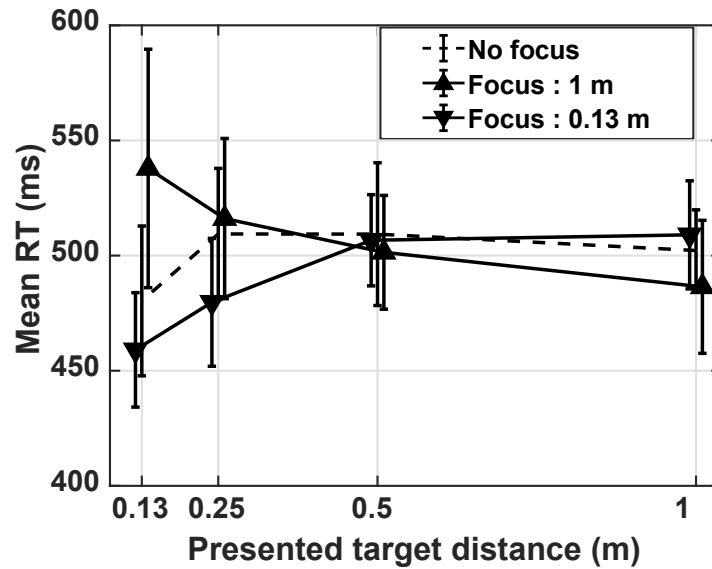
Mean reaction time

The mean RT calculated over all listeners are compared in Fig. 4.4 for No-Focus, Focus 1 m and Focus 0.13 m. The difference of RT between No-Focus condition and each of other two Focus conditions is represented in Fig. 4.5. By comparing the obtained RTs in No-Focus and Focus conditions, the effect of top-down attention can be analyzed. When the target sound is presented from the focus distance, a decrease of RT can be observed. In contrast, the RT obtained when the targets are presented from other distances is longer than that when the targets are presented from the focus distance. In the Focus 0.13 m condition, the obtained decrease of RT for targets presented at the focus distance was as much as 23 ms for targets presented at the focus distance, while the increase of RT for targets presented from the farthest distance was 43 ms. For the Focus 1 m condition, the decrease was as much as 13 ms and the increase of 57 ms. When the listener focused on 1 m, the RT became a decreasing function of target distance.

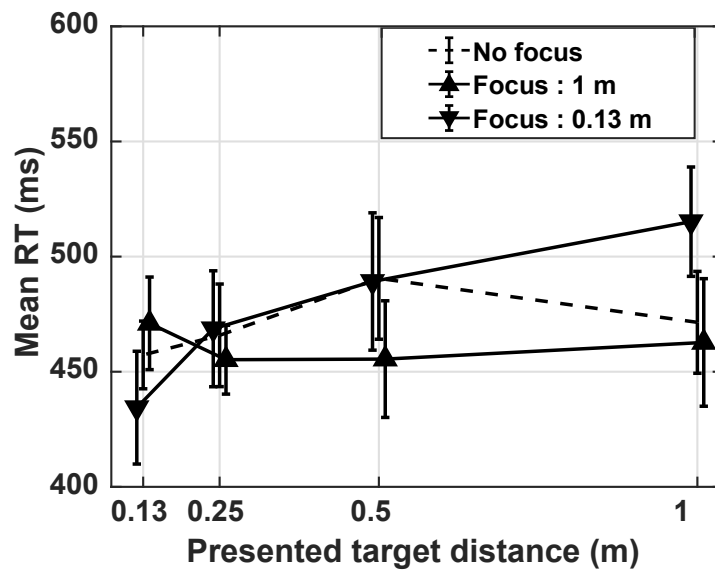
A three-way analysis of variance ANOVA was conducted on mean RT for the parameters of Distance (four target distances), Azimuth (-90° and 0°) and Condition (No-Focus, Focus 1 m, Focus 0.13 m). Results suggest an effect of Azimuth ($F(1,6) = 6.15, p < .05$) and Distance ($F(3,18) = 4.02, p < .05$) but not of Condition ($F(2,12) = 0.10, p = .90$). Interactions were significant for Condition \times Distance ($F(6,36) = 5.0, p < .001$), but not for Azimuth \times Condition ($F(2,12) = 1.32, p = .31$), Azimuth \times Distance ($F(3,18) = 0.46, p = .71$), or three-way interactions ($F(6,36) = 0.44, p = .85$).

The simple main effect of Distance was statistically significant in Condition at the two Focus distances : at 1 m ($F(2,48) = 3.9, p < .05$) and at 0.13 m ($F(2,48) = 8.75, p < .001$). A multiple comparison (Ryan's method) reveals that for these two focus distances, the difference between Focus 0.13 m and Focus 1 m is statistically significant ($p < .05$). This contributes to show that implicitly orienting the listener's attention to a particular distance considerably changes the listener's spatial selective processes, especially at the focus distance. The effect of Distance is largely significant for the Focus 0.13 m condition ($F(3,54) = 10.0, p < .001$), almost not significant in Focus 1 m condition ($F(3,54) = 2.2, p < .1$), and not significant in No-Focus condition ($F(3,54) = 2.06, p = .12$). A multiple comparison (Ryan's method) reveals that in Focus 0.13 m condition, the differences between 1 m and 0.13 m, between 1 m and 0.25 m and between 0.5 m and 0.13 m are significant ($p < .01$). The

non-significance of effects of Distance in No Focus condition is inconsistent with the results obtained in Chapter 3. Although the order in which Focus conditions were investigated was mixed, interactions between Focus and No Focus conditions may explain this inconsistency. The shift of attention in a Focus trial may have affected the results for the following No-Focus trial. Another possible explanation is the small number of listeners. Further study is needed to explain this inconsistency.

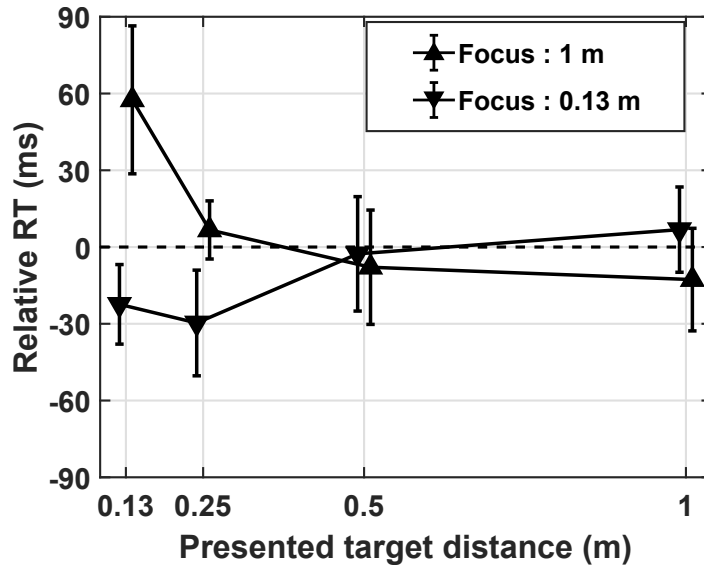


(a) Results on the interaural axis (-90°).

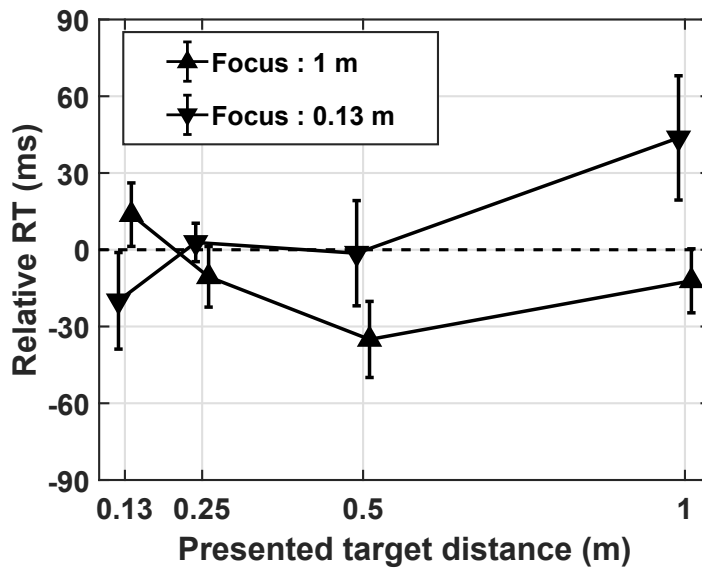


(b) Results on the median plane (0°).

Figure 4.4: Reaction time averaged for all listeners as a function of presented distance of the target sound source. The results for the interaural axis (a) and the median plane (b) are shown separately. Error bars indicate standard error. Dashed lines show the results when no focus on distance is attempted. Filled triangles show results for focus on 1 m (upper triangle) and 0.13 m (lower triangle).



(a) Results on the interaural axis (-90°).



(b) Results on the median plane (0°).

Figure 4.5: Relative reaction time advantage of focusing as a function of presented distance of the target sound source. The calculation was made by subtracting the results for No-Focus condition from results for Focus conditions. The results for the interaural axis (a) and median plane (b) are shown separately. Error bars indicate standard error. Upper triangles show results for focus on 1 m and lower triangles for focus on 0.13 m.

The effect of Azimuth was significant, but interactions between Azimuth and other factors were not. The RT averaged on all distances is faster on the median plane than on the interaural axis, regardless of focus conditions. This was not observed in the results from Chapter 3. A previous study of auditory attention to direction reported faster reaction time to sounds coming from the median plane than for sounds coming from the sides [17] when that direction is attended to. Further study is needed to understand this difference.

Hit rate and false alarm rate

The hit rate was consistently close to 100% and the false alarm rate consistently close to 0%. The maximum false alarm (FA) rate was obtained for the closest distance : 0.13 m in No-Focus condition and Focus 0.13 m condition. However, the FA rate was constant in Focus 1 m. These effects were not significant. Indeed, the maximum difference in FA rate between conditions was 3%, which was not considered as statistically significant. Moreover, the results of a three-way ANOVA for FA rate show only the significant effect of Condition ($F(2, 12) = 5.3, p < .05$), and not of Distance ($F(3, 18) = 2.08, p = .14$) or Condition \times Distance interactions ($F(6, 36) = 1.25, p = .30$). The difference between No-Focus and Focus 1 m was significant ($p < .01$). However, the difference between No-Focus and Focus 0.13 m was small ($p = .06$) and there was no significant difference between Focus conditions ($p = .28$). Results of ANOVA for hit rate show no significant effects.

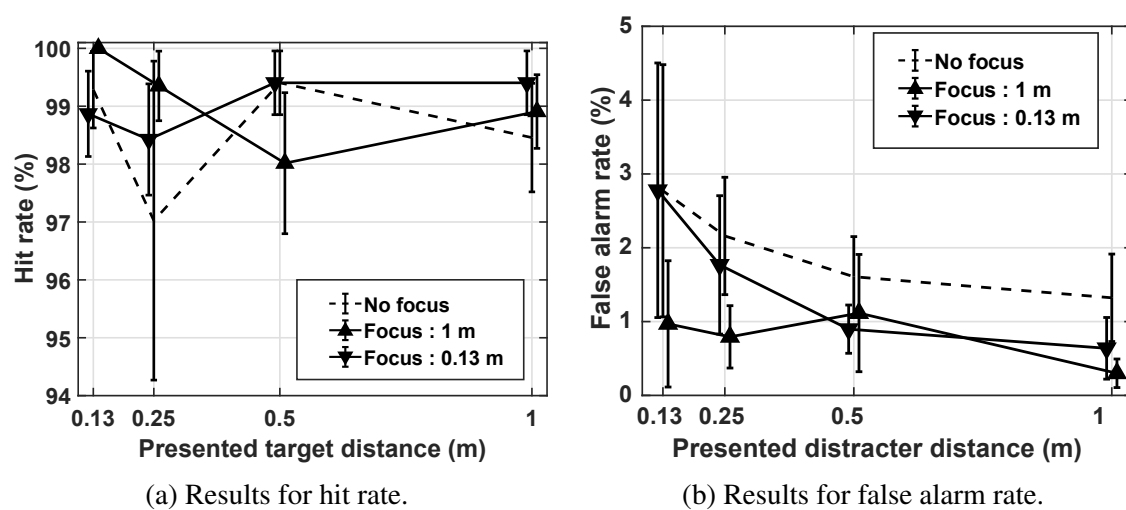


Figure 4.6: Hit rate and false alarm rate averaged over all listeners and azimuths. Upper triangles show results for focus on 1 m, lower triangles for focus on 0.13 m and dashed lines for No-Focus.

4.4 Discussion

4.4.1 Existence of the auditory spotlight for distance and interactions with peripersonal space

Results showed that attention on a particular distance affected the RT to the target speech sound. When the attention of the listener was implicitly directed to a particular distance, the response to sounds coming from this distance became faster. In contrast, the response to sounds coming from a different distance became slower. A similar tendency could be obtained by the study of auditory spatial attention on direction by using RT [17] and correct response rate [16, 18, 19]. For direction, auditory spatial attention is compared to a “spotlight” which shines on the desired direction. Auditory process of sound sources from within the spotlight would be enhanced, while sources presented from a larger angular separation from the center of the spotlight would be processed with decreased capabilities. Therefore, the effect of the spotlight for direction is a function of angular distance, defined as the absolute value of the angle between direction of focus and direction of presented target sound source. For direction, a recent study (unpublished data, Teraoka *et al.*) suggests that the spotlight of attention is of similar shape and effect for several different directions of focus.

For distance, the spotlight does not seem to be identical for all focus distances. Indeed, the shape of the spotlight for 1 m and 0.13 m was different, as represented on Fig. 4.7. On this figure, the effect of attention on RT is represented as a function of distance from focus point. If taking the distance at which the benefits of focusing on a particular source distance is null (relative RT is 0 in Fig. 4.7), the spotlight seems to be broader when focusing on far distances than on near distances. The spotlight in the Focus 1 m condition results in faster RT until 0.75 m separation from the point of focus. However, in Focus 0.13 m, it results in faster RT only for sources until 0.37 m separation from the point of focus. There would be two reasons to explain this phenomenon. One reason could be perceptual resolution of the space which is simulated using HRTFs. It was more difficult to distinguish the position of sound source between 0.5 m and 1.0 m than that between 0.5 m and 0.25 m. This effect could be observed in section 2.5. Because distances of far sound

sources are hard to differentiate, focusing on 1 m results in a broad spotlight focused on all far distances. Another reason could be the effect of peripersonal space. This tendency could be slightly observed in section 3.4.1. When a sound is presented from the near field in peripersonal space, auditory processes would be enhanced. If this speculation is true, an enhancement of the resolution of the top-down auditory spatial attention could occur for sounds in peripersonal space. On the other hand, when a sound is presented outside of the peripersonal space, no enhancement of auditory process would occur. As a result, the resolution of auditory attention in the area is wide and the listener responds to all sounds outside of the peripersonal space.

The difference in false alarm rate also suggests different properties of auditory attention between close sources in peripersonal space and farther sources. No-Focus resulted in a higher false alarm rate for the closest distances. This means that the listeners responded to distracter sounds presented from 0.13 m although they were instructed to respond only to target sounds. This contributes to the theory that near sounds in peripersonal space are naturally alarming. Focusing on 0.13 m did not change this behavior. However, focusing on 1 m led to an overall smaller false alarm rate for all distances, including 0.13 m. This suggests a more composed selection of sounds when the focus is on 1 m.

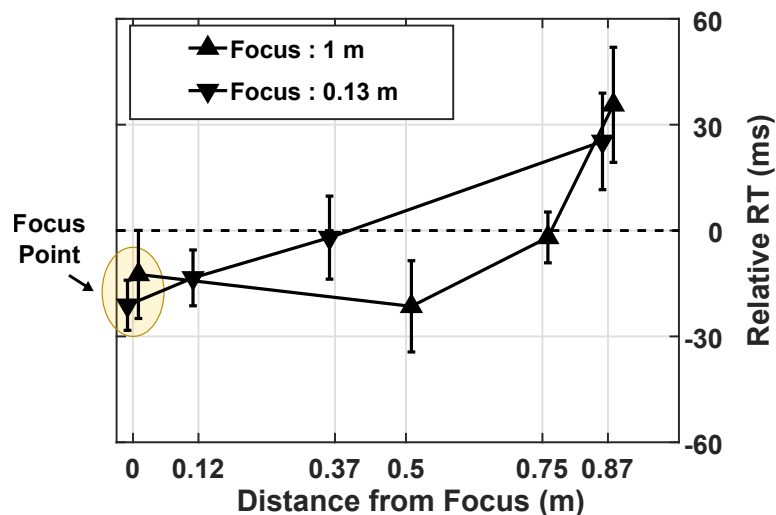


Figure 4.7: Relative contribution of focusing on distance as a function of distance to focus distance. This figure is obtained by averaging results for the left and front azimuths. Upper triangles show results for focus on 1 m, lower triangles for focus on 0.13 m. As the target source is presented farther away from the focus distance, the RT increases. The size and shape of the attention spotlight is different as a function of focus distance.

4.4.2 Relevance of these results

As mentioned in the previous chapter, the relationship between the distance of sound source and auditory attention is unclear. No known study of the auditory attention spotlight as a function of distance of sound source had been conducted. This section aims to discuss the results obtained in this chapter and to present how these results could be further completed in future studies.

The existence of the auditory spotlight for distance, as well as a change of the shape of this spotlight along distance can be speculated from the results in this study. The change in shape of the spotlight according to focus distance is speculated to be due to enhanced auditory processes for sounds from within peripersonal space. This enhancement of auditory processes could also affect the auditory attention spotlight for direction. The effects of peripersonal distance of sound sources on auditory attention for direction is an interesting subject for further study.

In this study, the distances considered were either always closer than the far focus distance or always farther than the close focus distance. The shape of the auditory spotlight was not investigated symmetrically around a focus distance. Moreover, in neither focus condition did the effect of attention seem to reach a saturation although this was reported for direction in previous studies [16, 17]. In these studies, beyond a certain angular distance, the effect of greater angular separation did not lead to faster processes [17] or better correct response rate [16]. Studying the effect for further distances than 1 m could complete the study of both the 0.13 m and the 1 m spotlight's shapes. This study is bound to be complex in anechoic conditions, as distances beyond 1 m in anechoic conditions are difficult to differentiate [21]. The presence of reverberation has been reported to benefit the localization accuracy of sound sources presented from far distances [29, 30, 31]. Therefore, using a reverberant environment could extend the range of possible source distances beyond 1 m to further study auditory spatial attention for distance.

It is also important to understand the conditions in which these results were obtained. The sound sources were virtual, presented through headphones in an anechoic environment excluding the intensity cue. If supposing that the capacity and shape of the attention spotlight depends on the localization accuracy, then these results are bound to change if using real sound sources, if including the intensity cue, or if conducting the experiment in a reverberant environment.

Finally, the average RT in this experiment was fast, and the correct response rate was very high. This shows that the task was done with ease by the listeners. Furthermore, little learning effects were observed. Several past studies have shown that the effects of auditory spatial attention become more important when the task is complex [20, 50, 70, 71, 72]. If making the task in this experiment more complex, the effects of auditory attention for distance may potentially be more obvious.

4.5 Chapter conclusions

An experiment of auditory spatial attention was conducted by implicitly orienting the listener's auditory attention to specific distances. The listeners performed a target detection task in which the distance of presentation of the source was varied. Evaluation of the listeners' reaction time revealed that the sounds presented from the focus distance were consistently responded to faster than from other non focused distances. Moreover, reaction time increased according to the distance from the focus point. This reveals the availability of top-down attention on distance. In addition, the distance at which the listeners focus affects their spatial selective capabilities. Focus on far distances resulted in broad spatial selectivity, whereas focus on very near distances resulted in a narrow spatial selectivity. This suggests a difference in processes for very near sound sources, as compared to far sound sources. This difference in processes is believed to be due to peripersonal space effects for the very near sound sources.

Chapter 5

Overall conclusion

Auditory spatial attention for distance of sound sources was investigated in this thesis. The novelty was to investigate the effects of distance of sound sources within peripersonal space. It is believed that human auditory processes of sound within peripersonal space are different from sources in extrapersonal space. The sound sources were virtual, rendered in an anechoic environment. The generated sound was presented via headphones. The listeners were instructed to respond as fast as possible once they detected a target word under a distracting speech sound environment. The distances of the target and distracter sound sources were varied to investigate the relationship between auditory spatial attention and sound source distance, as well as its relationship with distance of competing sound sources.

An introduction to the method used to generate accurate virtual sound sources in an anechoic environment was presented in Chapter 2. Head-related transfer functions (HRTFs) were filtered through distance varying filters (DVF) in order to obtain transfer functions for distances within peripersonal space. The numerical evaluation of these functions, as well as a psychoacoustic localization accuracy experiment were conducted. The results showed that virtual sound sources synthesized using these functions were localized with accurately. The accuracy was comparable to the average human accuracy for judgment of sound source distance. Based on these results, it was concluded that this DVF method is appropriate to generate near-field sound source.

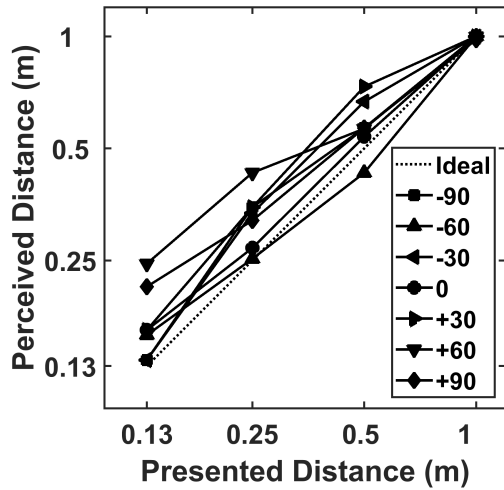
Using the spatialized virtual sound sources, auditory attention as a function of sound source distance was investigated in Chapter 3. Effects of distance separation of competing sound sources were also investigated. Spatial configurations and experiment conditions were chosen to investigate the individual contribution of each distance perception cue. Results showed that the closest sound sources capture auditory attention more consistently than further sounds. This suggests effects of auditory peripersonal space on auditory processes. Indeed, the acoustic properties of very near sounds resulted in enhanced auditory processes. In addition, distance separation between competing sound sources was used effectively to decrease reaction time even when the intensity cue for distance of sound was not provided. The initial small separation between sound sources resulted in a higher contribution to reaction time than further separation. This suggested that distance separation benefits sound stream segregation. This seemed to have effect regardless of source intensity.

While auditory attention with no a priori knowledge of the position of the target sound source was the focus of Chapter 3, the contribution of knowing the position of the source prior to stimulus presentation was investigated in Chapter 4. The listeners' attention was implicitly directed to specific focus distances using the probe-signal method. A similar target word detection task as Chapter 3 was applied. Results suggested the effect of top-down attentional capabilities for distance. The sounds presented from the focus distance were consistently responded to faster than when presented from other distances. Furthermore, the size of the selective area of spatial attention was a function of focus distance. When focusing on a very near distance, the area of selection was narrower than when focusing on a far distance.

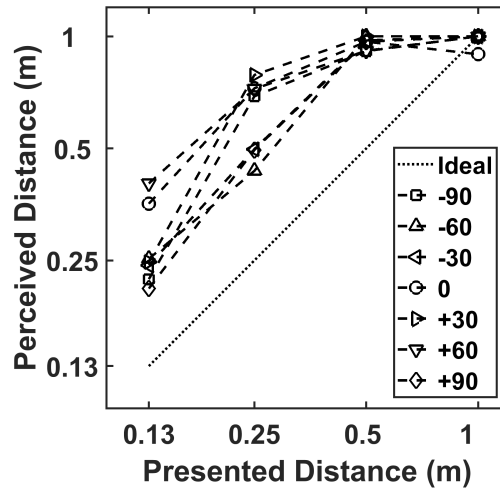
Although these findings suggest a spatial selective ability with auditory distance, the study presented in this thesis opens for many questions on the mechanisms of human auditory attention as a function of sound source distance. Mechanisms underlying the enhanced auditory processes of the closest sounds can be linked to the multisensory mechanisms that occur within peripersonal space, especially within the space very near to the listener's head. As for the study of the top-down auditory attention spotlight for distance, its shape and size are still unclear. Similar investigation for a larger range of distances would contribute to establish a model of the spotlight for distance. Based on its interactions with the attention spotlight for direction, a two-dimensional model of the auditory attention spotlight would be constructed. Finally, it would be interesting to investigate the mechanisms of auditory spatial attention in the whole peripersonal space, including rear space.

Appendix A

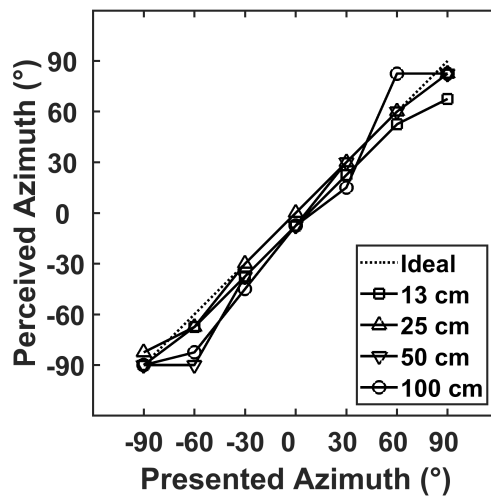
DVF filtered HRTF localization accuracy - individual results



(a) Including the intensity cue.

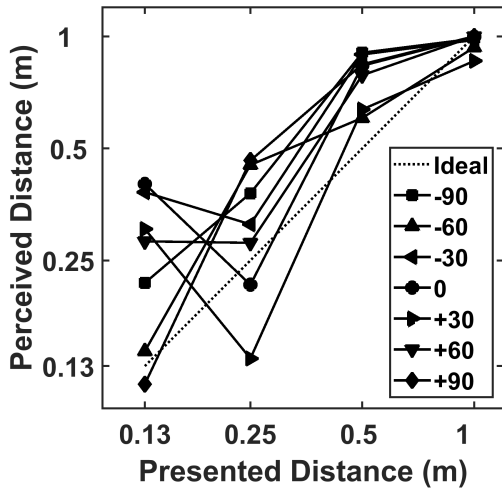


(b) Excluding the intensity cue.

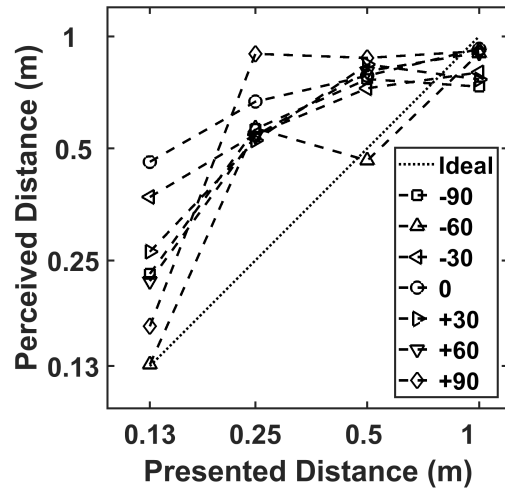


(c) Azimuth localization.

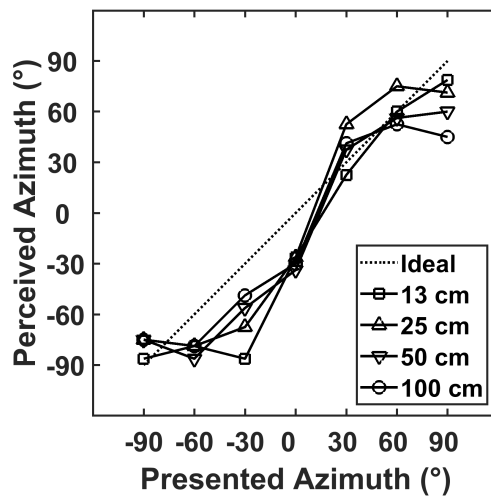
Figure A.1: Localization accuracy results for Listener 1. (a) distance localization results including intensity cue, (b) distance localization results excluding intensity cue, (c) azimuth localization results.



(a) Including the intensity cue.

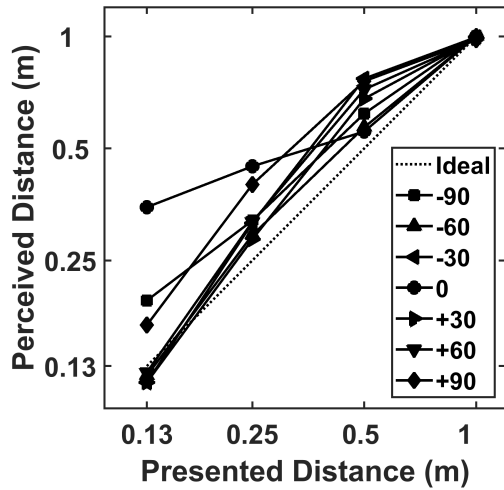


(b) Excluding the intensity cue.

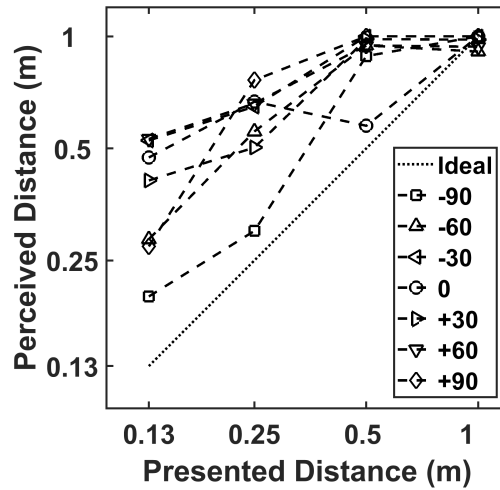


(c) Azimuth localization.

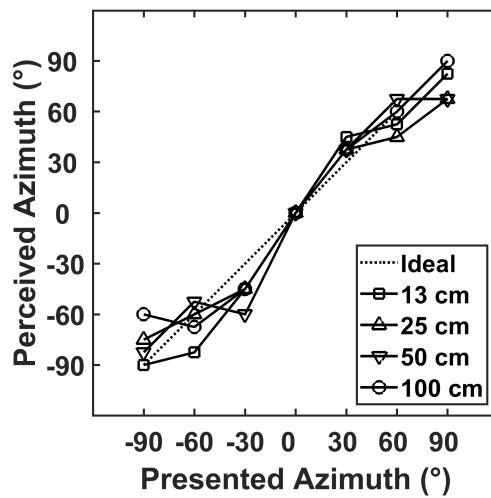
Figure A.2: Localization accuracy results for Listener 2. (a) distance localization results including intensity cue, (b) distance localization results excluding intensity cue, (c) azimuth localization results.



(a) Including the intensity cue.

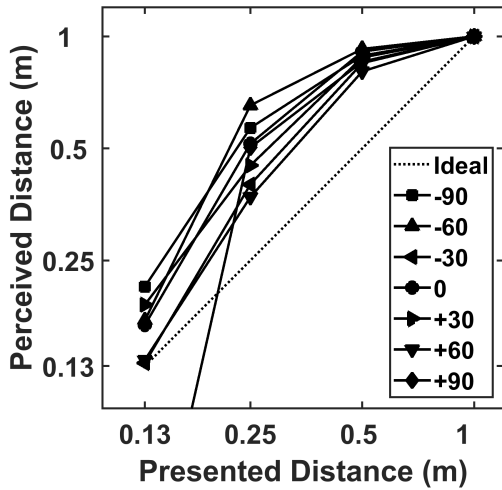


(b) Excluding the intensity cue.

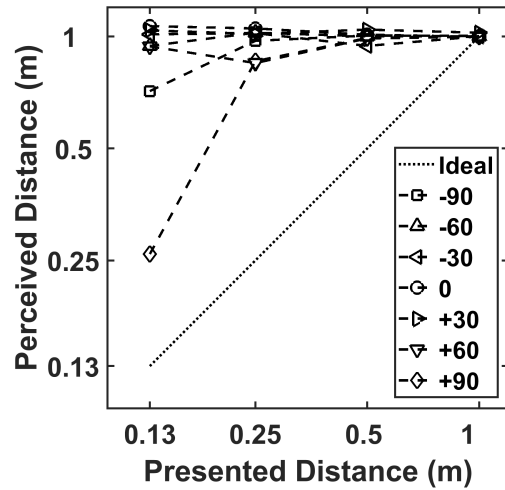


(c) Azimuth localization.

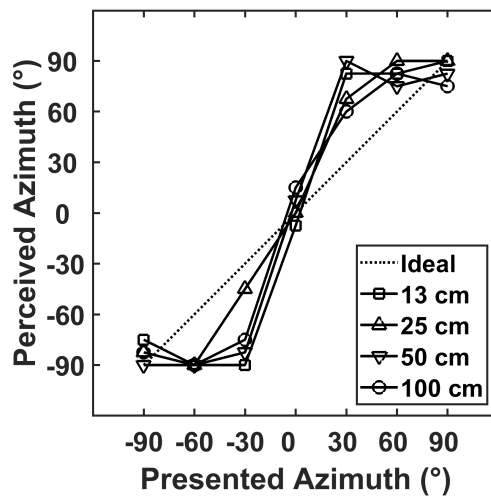
Figure A.3: Localization accuracy results for Listener 3. (a) distance localization results including intensity cue, (b) distance localization results excluding intensity cue, (c) azimuth localization results.



(a) Including the intensity cue.

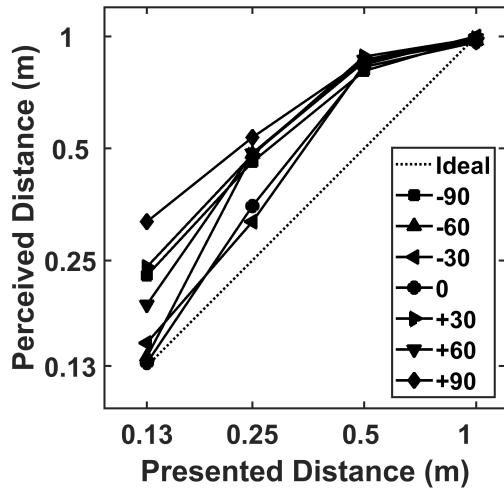


(b) Excluding the intensity cue.

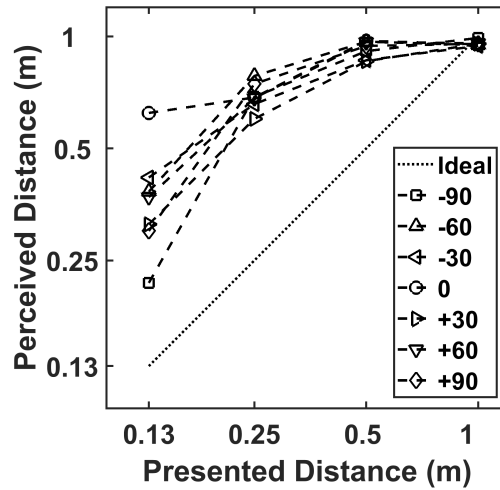


(c) Azimuth localization.

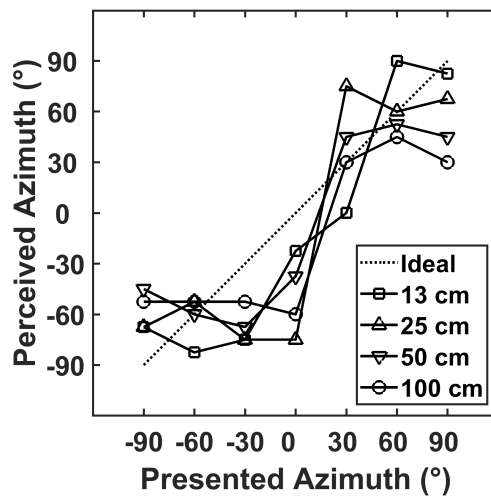
Figure A.4: Localization accuracy results for Listener 4. (a) distance localization results including intensity cue, (b) distance localization results excluding intensity cue, (c) azimuth localization results.



(a) Including the intensity cue.

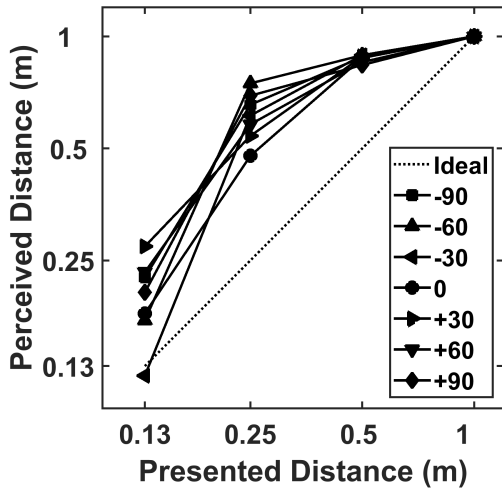


(b) Excluding the intensity cue.

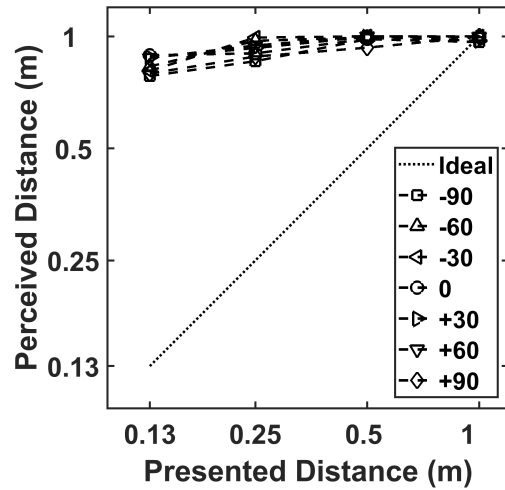


(c) Azimuth localization.

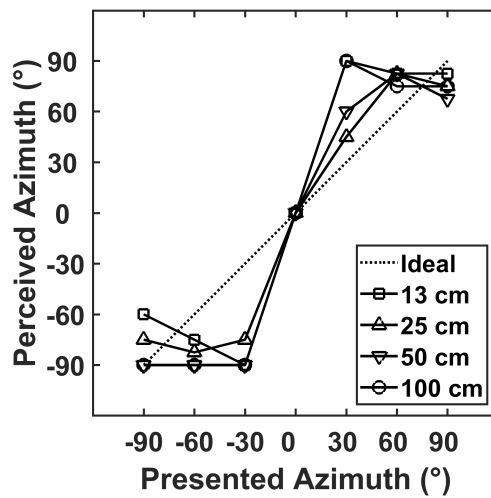
Figure A.5: Localization accuracy results for Listener 5. (a) distance localization results including intensity cue, (b) distance localization results excluding intensity cue, (c) azimuth localization results.



(a) Including the intensity cue.

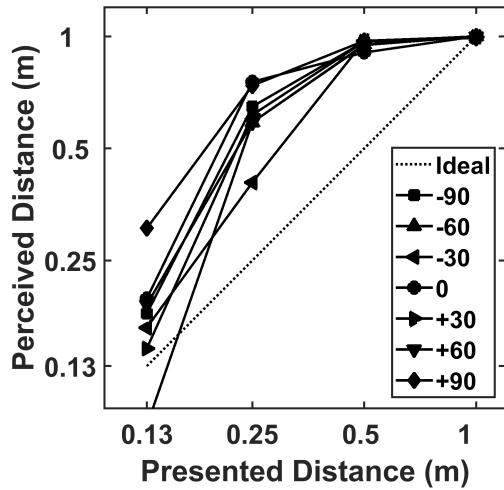


(b) Excluding the intensity cue.

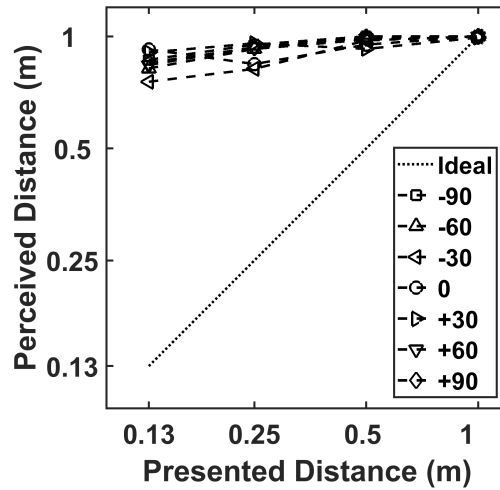


(c) Azimuth localization.

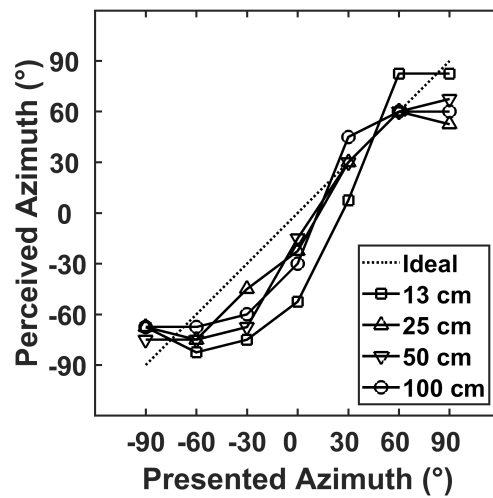
Figure A.6: Localization accuracy results for Listener 6. (a) distance localization results including intensity cue, (b) distance localization results excluding intensity cue, (c) azimuth localization results.



(a) Including the intensity cue.

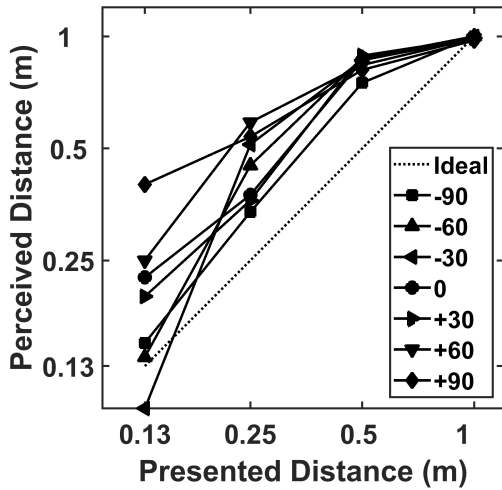


(b) Excluding the intensity cue.

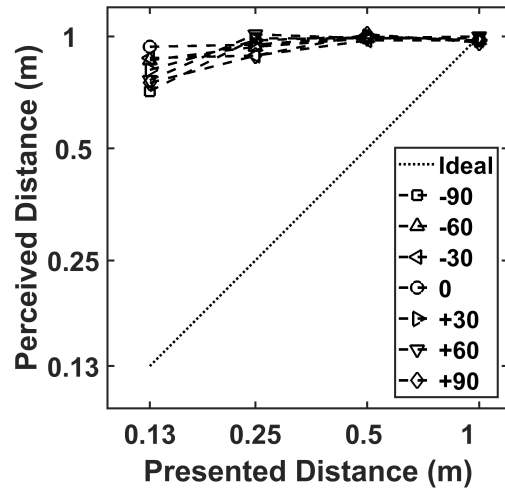


(c) Azimuth localization.

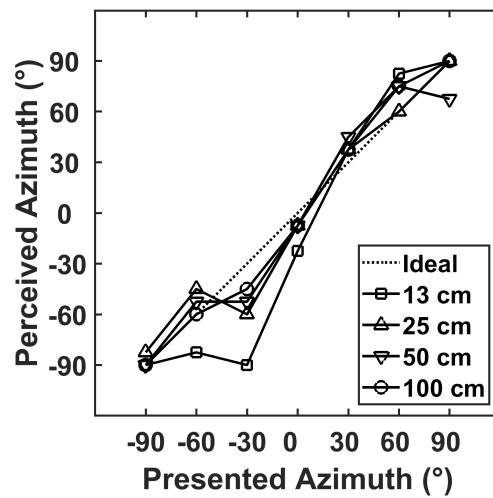
Figure A.7: Localization accuracy results for Listener 7. (a) distance localization results including intensity cue, (b) distance localization results excluding intensity cue, (c) azimuth localization results.



(a) Including the intensity cue.

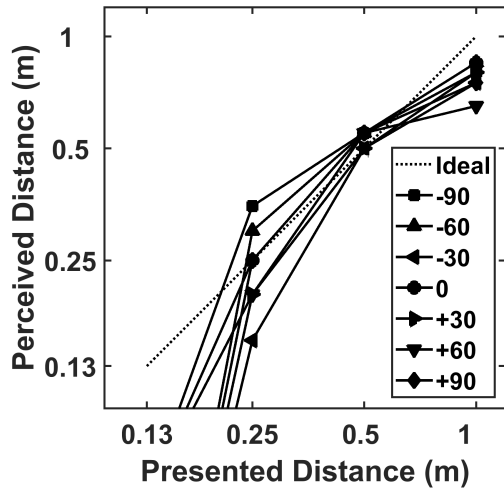


(b) Excluding the intensity cue.

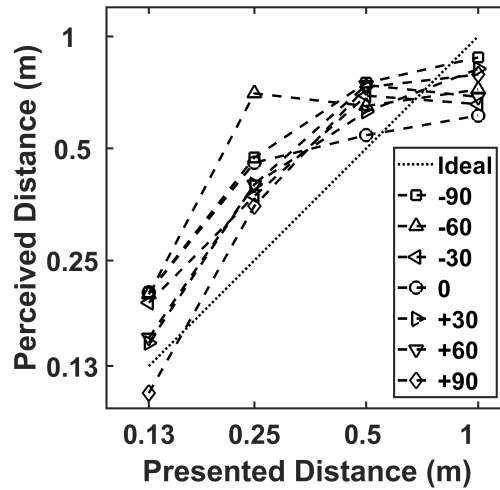


(c) Azimuth localization.

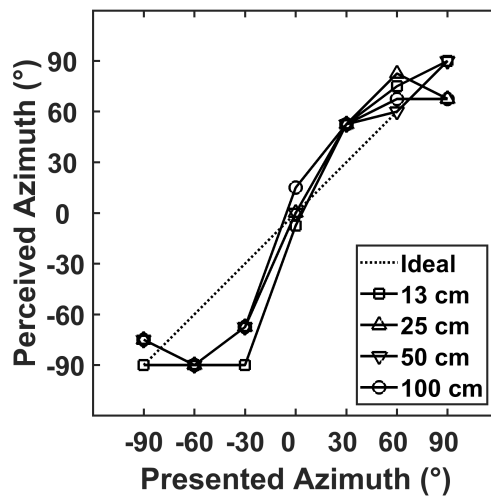
Figure A.8: Localization accuracy results for Listener 8. (a) distance localization results including intensity cue, (b) distance localization results excluding intensity cue, (c) azimuth localization results.



(a) Including the intensity cue.

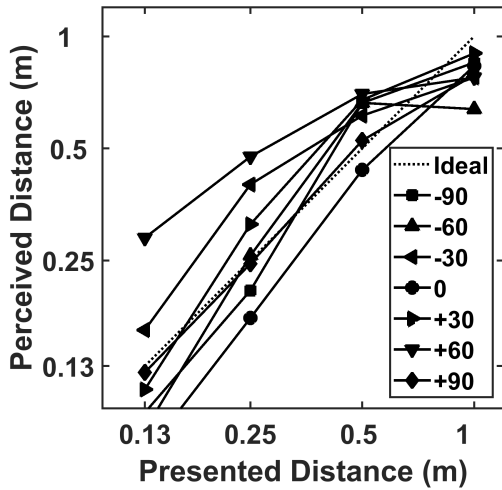


(b) Excluding the intensity cue.

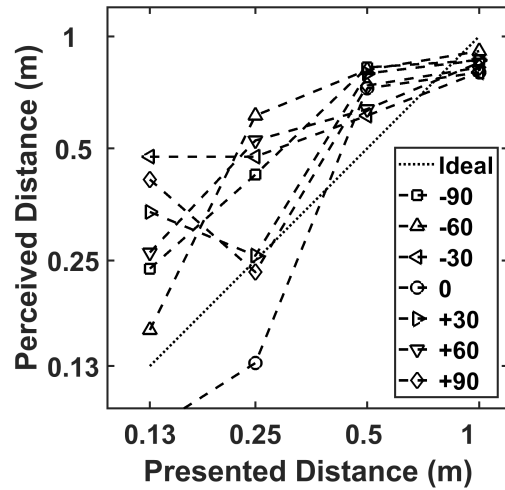


(c) Azimuth localization.

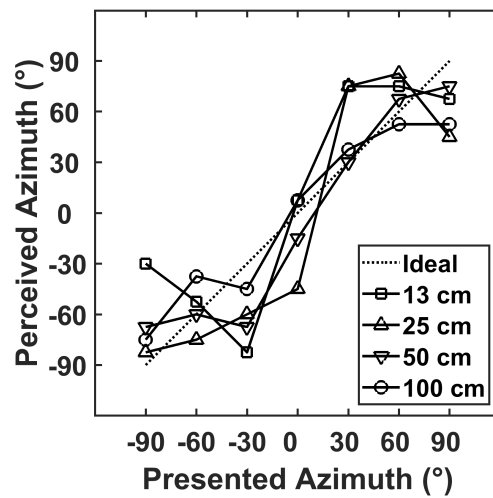
Figure A.9: Localization accuracy results for Listener 9. (a) distance localization results including intensity cue, (b) distance localization results excluding intensity cue, (c) azimuth localization results.



(a) Including the intensity cue.



(b) Excluding the intensity cue.

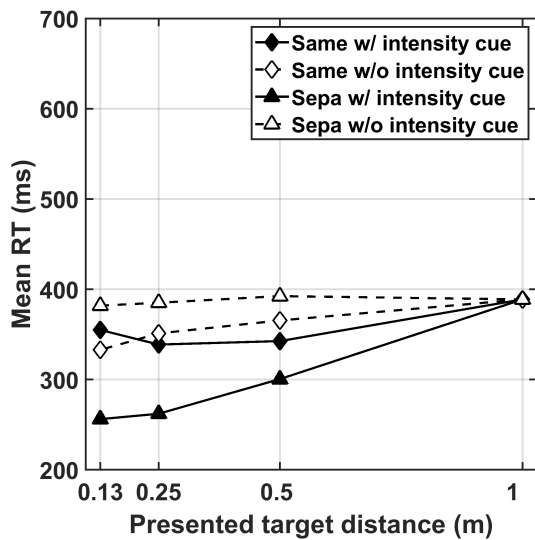


(c) Azimuth localization.

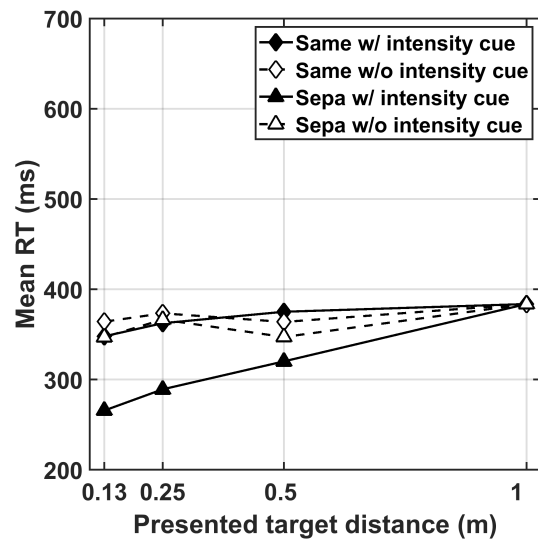
Figure A.10: Localization accuracy results for Listener 10. (a) distance localization results including intensity cue, (b) distance localization results excluding intensity cue, (c) azimuth localization results.

Appendix B

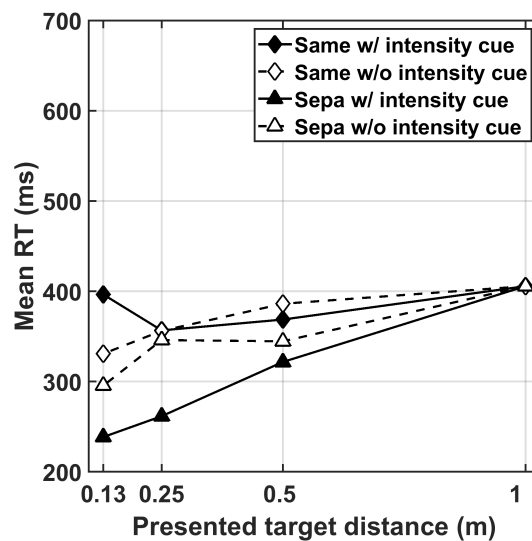
Bottom-up effects of distance - individual results



(a) Sources presented from -90° .

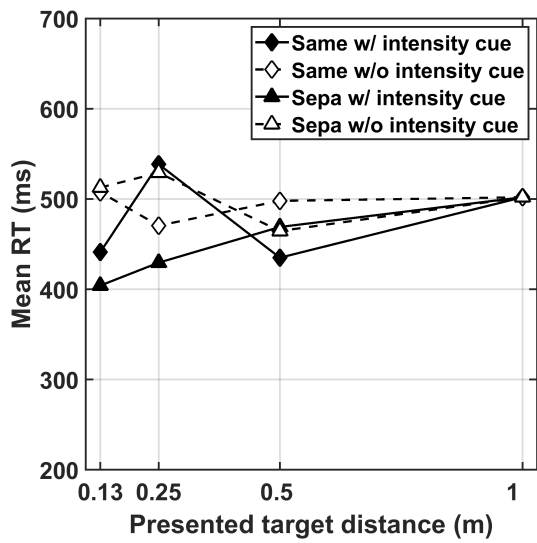


(b) Sources presented from $+90^\circ$.

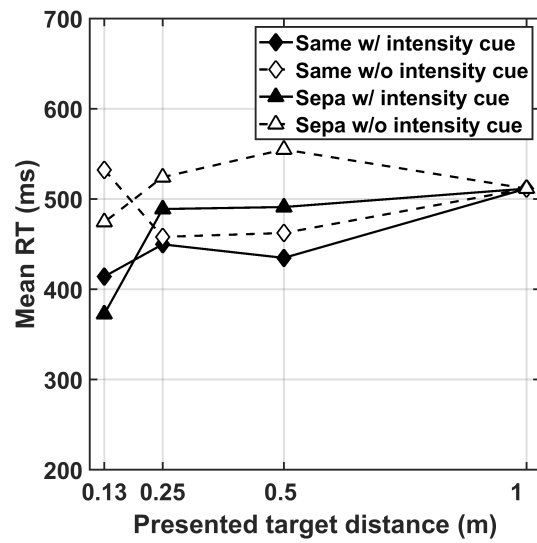


(c) Sources presented from 0° .

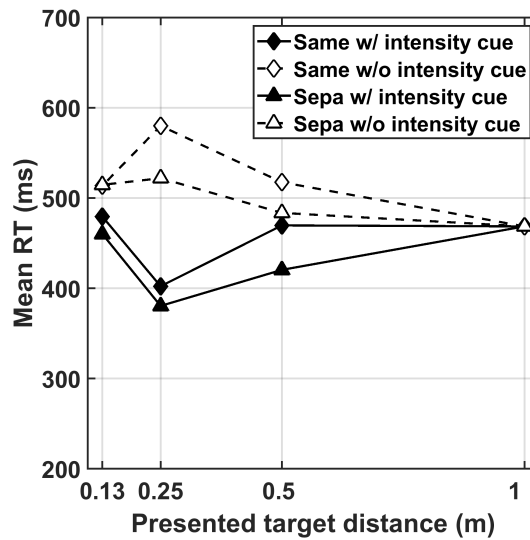
Figure B.1: Reaction time experiment results for Listener 1. (a) mean RT for sounds presented from -90° , (b) mean RT for sounds presented from $+90^\circ$, (c) mean RT for sounds presented from 0° . Full black lines are results when including the intensity cue, dashed white lines are results when excluding the intensity cue. Diamonds are results for Same Distance condition, Triangles for Distance Separation condition.



(a) Sources presented from -90° .

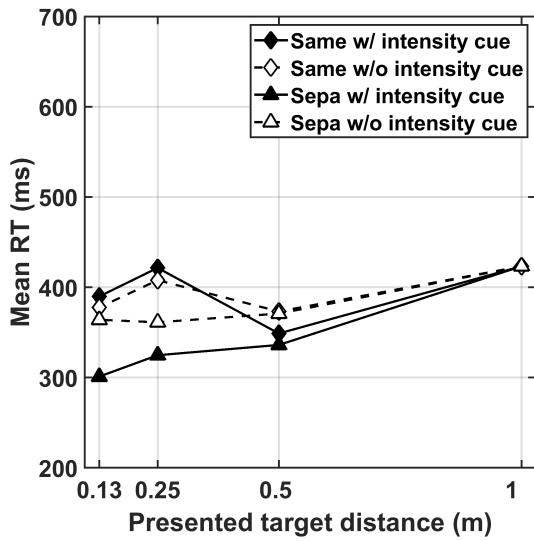


(b) Sources presented from $+90^\circ$.

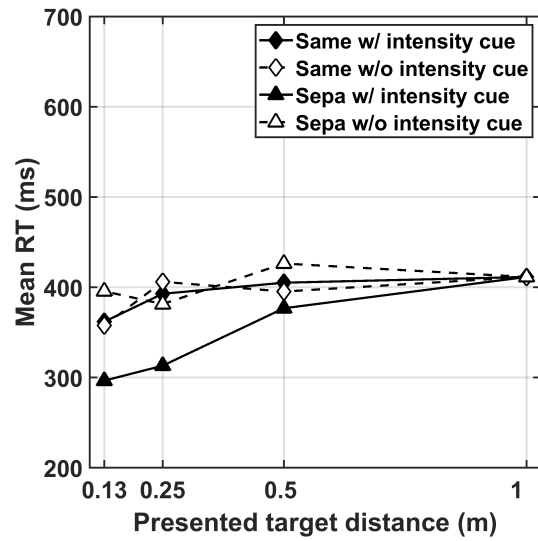


(c) Sources presented from 0° .

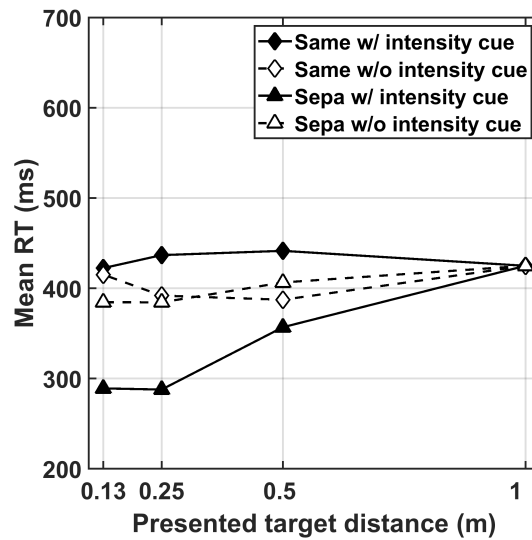
Figure B.2: Reaction time experiment results for Listener 2. (a) mean RT for sounds presented from -90° , (b) mean RT for sounds presented from $+90^\circ$, (c) mean RT for sounds presented from 0° . Full black lines are results when including the intensity cue, dashed white lines are results when excluding the intensity cue. Diamonds are results for Same Distance condition, Triangles for Distance Separation condition.



(a) Sources presented from -90° .

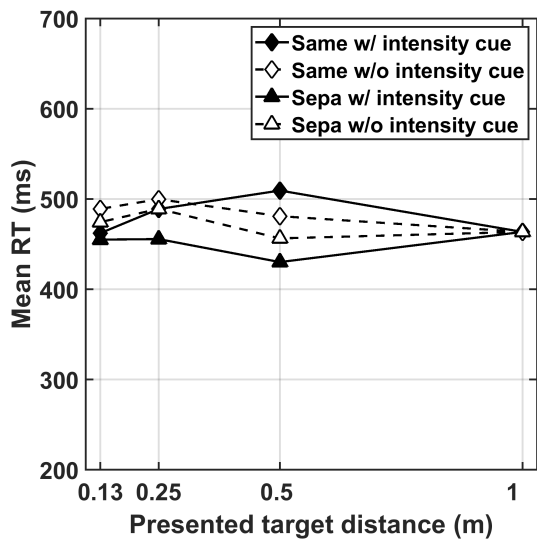


(b) Sources presented from $+90^\circ$.

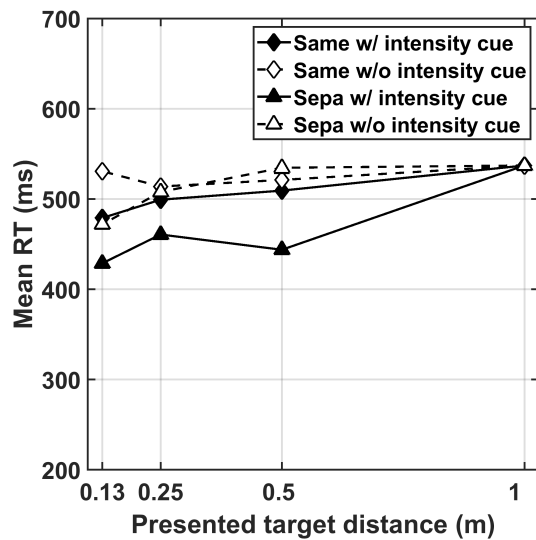


(c) Sources presented from 0° .

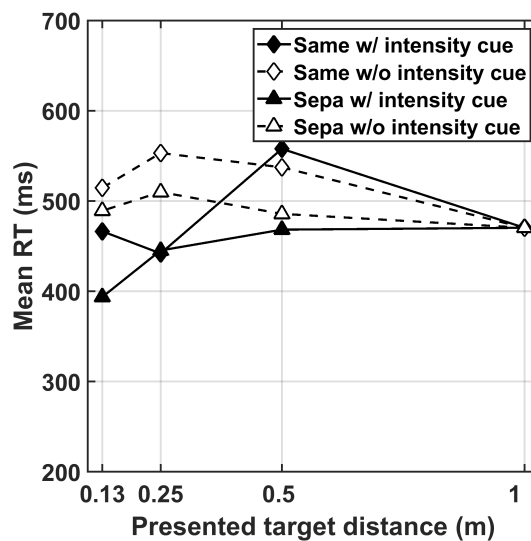
Figure B.3: Reaction time experiment results for Listener 3. (a) mean RT for sounds presented from -90° , (b) mean RT for sounds presented from $+90^\circ$, (c) mean RT for sounds presented from 0° . Full black lines are results when including the intensity cue, dashed white lines are results when excluding the intensity cue. Diamonds are results for Same Distance condition, Triangles for Distance Separation condition.



(a) Sources presented from -90° .

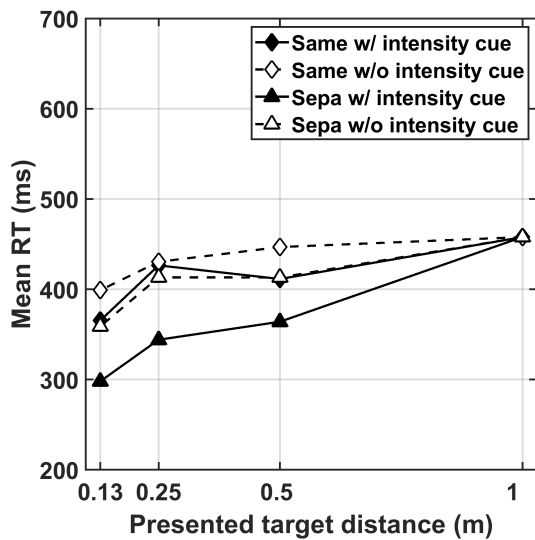


(b) Sources presented from $+90^\circ$.

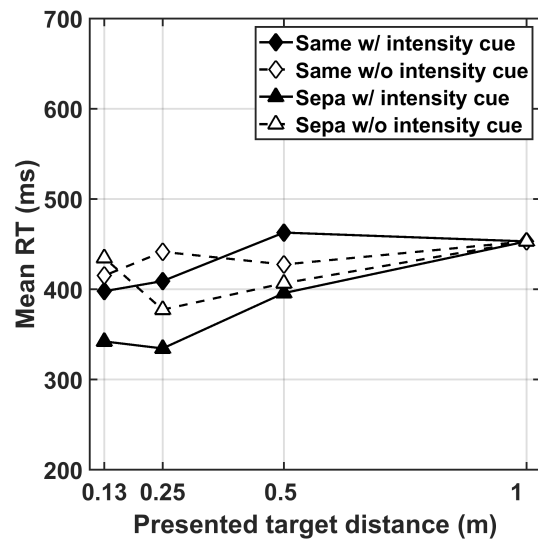


(c) Sources presented from 0° .

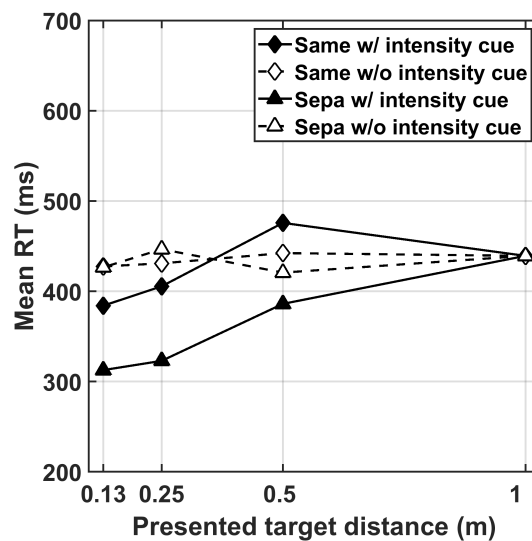
Figure B.4: Reaction time experiment results for Listener 4. (a) mean RT for sounds presented from -90° , (b) mean RT for sounds presented from $+90^\circ$, (c) mean RT for sounds presented from 0° . Full black lines are results when including the intensity cue, dashed white lines are results when excluding the intensity cue. Diamonds are results for Same Distance condition, Triangles for Distance Separation condition.



(a) Sources presented from -90° .

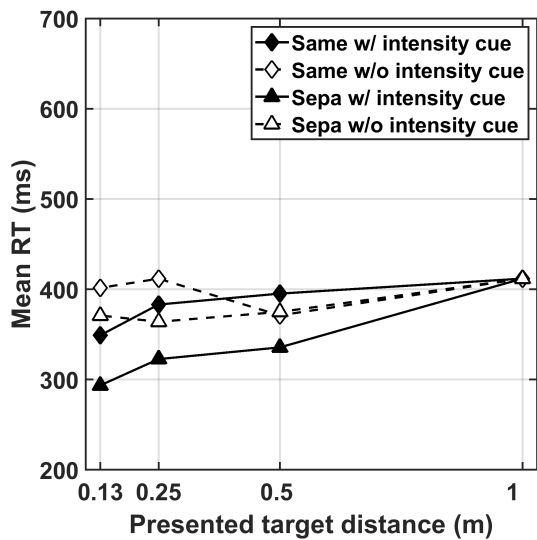


(b) Sources presented from $+90^\circ$.

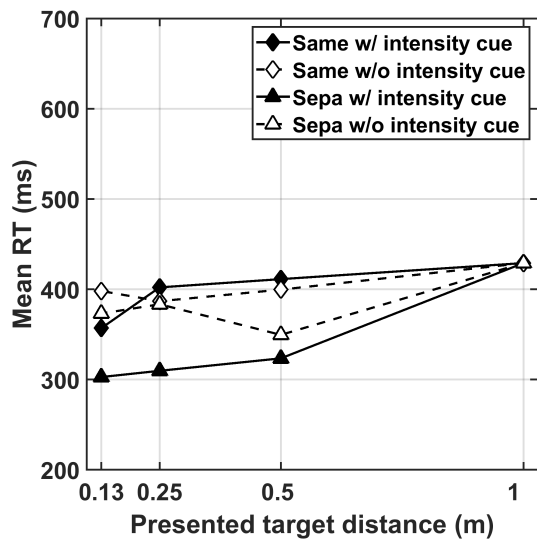


(c) Sources presented from 0° .

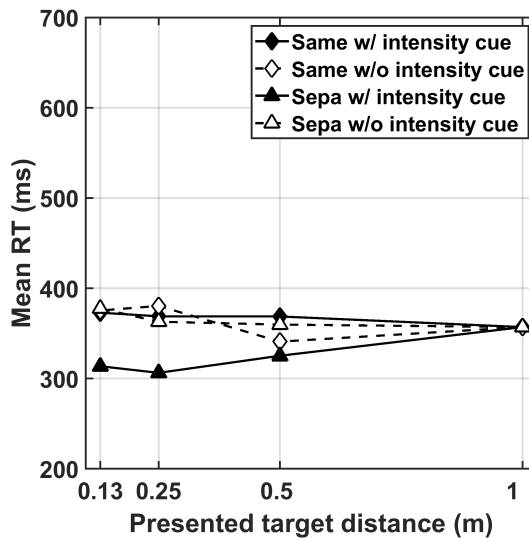
Figure B.5: Reaction time experiment results for Listener 5. (a) mean RT for sounds presented from -90° , (b) mean RT for sounds presented from $+90^\circ$, (c) mean RT for sounds presented from 0° . Full black lines are results when including the intensity cue, dashed white lines are results when excluding the intensity cue. Diamonds are results for Same Distance condition, Triangles for Distance Separation condition.



(a) Sources presented from -90° .

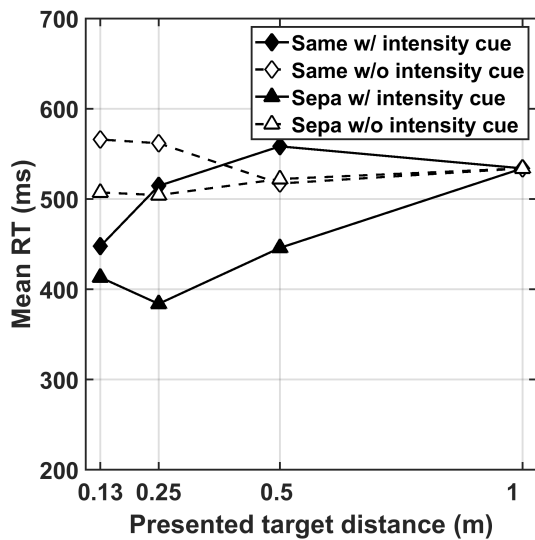


(b) Sources presented from $+90^\circ$.

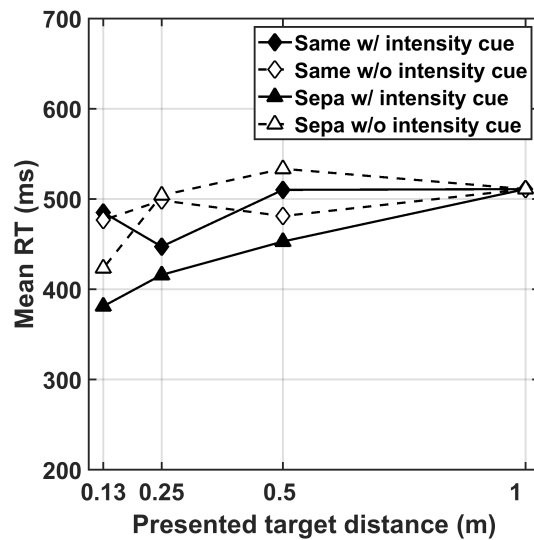


(c) Sources presented from 0° .

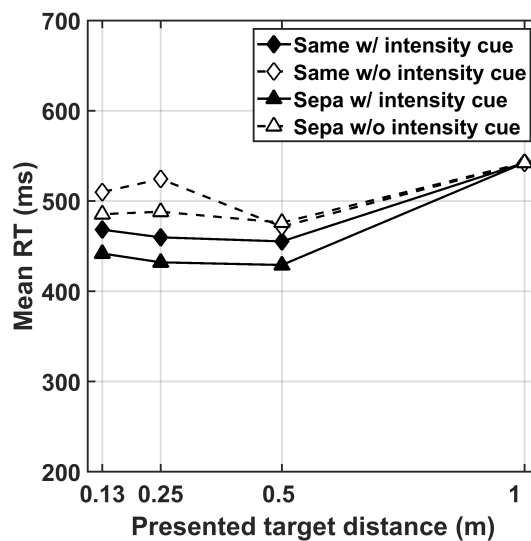
Figure B.6: Reaction time experiment results for Listener 6. (a) mean RT for sounds presented from -90° , (b) mean RT for sounds presented from $+90^\circ$, (c) mean RT for sounds presented from 0° . Full black lines are results when including the intensity cue, dashed white lines are results when excluding the intensity cue. Diamonds are results for Same Distance condition, Triangles for Distance Separation condition.



(a) Sources presented from -90° .

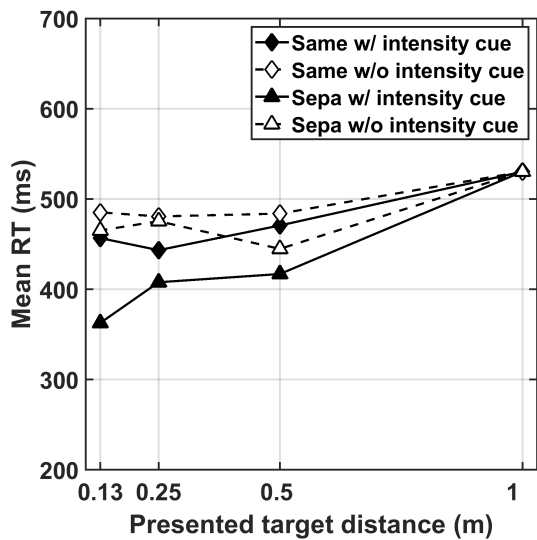


(b) Sources presented from $+90^\circ$.

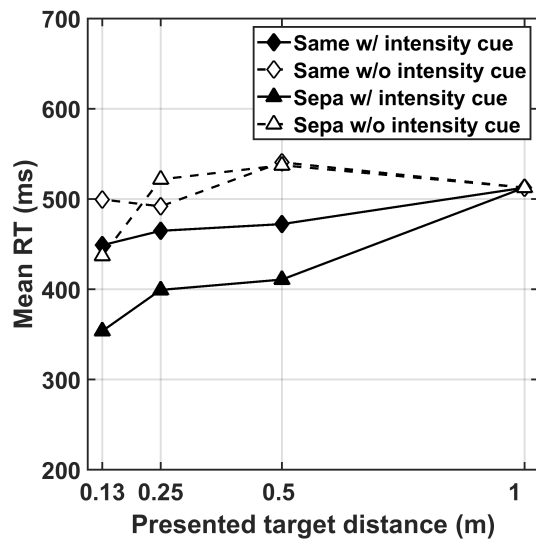


(c) Sources presented from 0° .

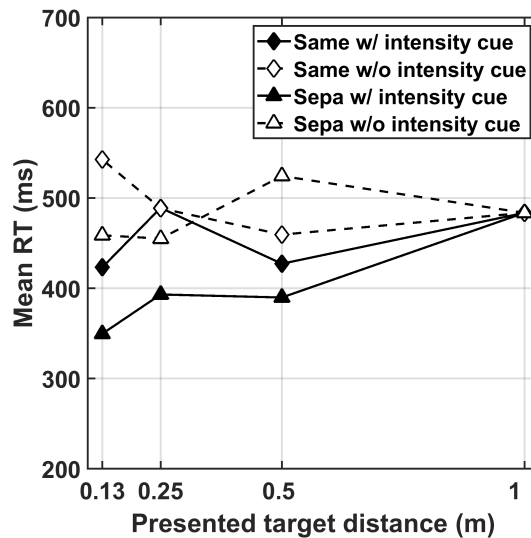
Figure B.7: Reaction time experiment results for Listener 7. (a) mean RT for sounds presented from -90° , (b) mean RT for sounds presented from $+90^\circ$, (c) mean RT for sounds presented from 0° . Full black lines are results when including the intensity cue, dashed white lines are results when excluding the intensity cue. Diamonds are results for Same Distance condition, Triangles for Distance Separation condition.



(a) Sources presented from -90° .

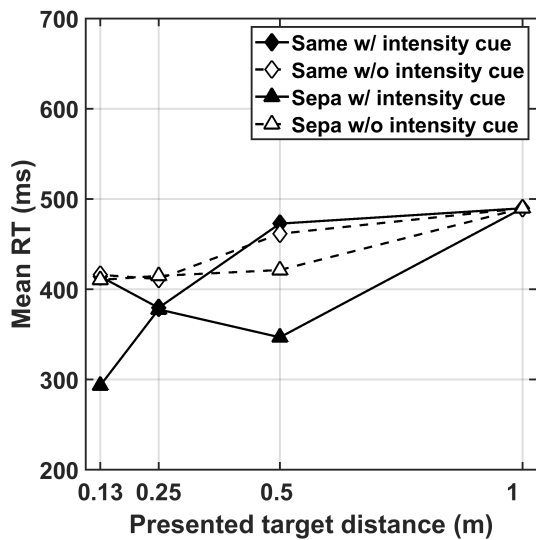


(b) Sources presented from $+90^\circ$.

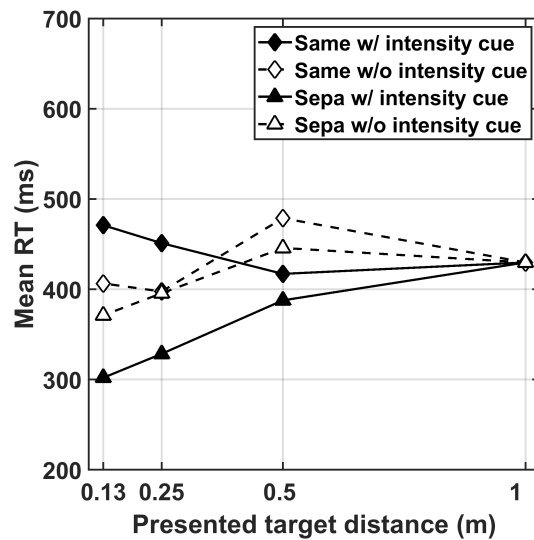


(c) Sources presented from 0° .

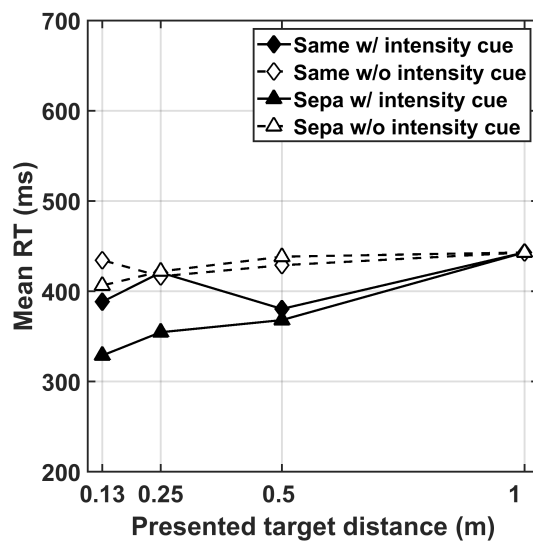
Figure B.8: Reaction time experiment results for Listener 8. (a) mean RT for sounds presented from -90° , (b) mean RT for sounds presented from $+90^\circ$, (c) mean RT for sounds presented from 0° . Full black lines are results when including the intensity cue, dashed white lines are results when excluding the intensity cue. Diamonds are results for Same Distance condition, Triangles for Distance Separation condition.



(a) Sources presented from -90° .



(b) Sources presented from $+90^\circ$.

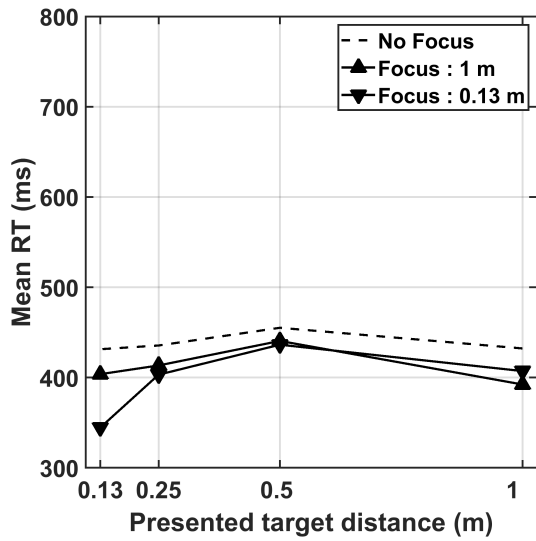


(c) Sources presented from 0° .

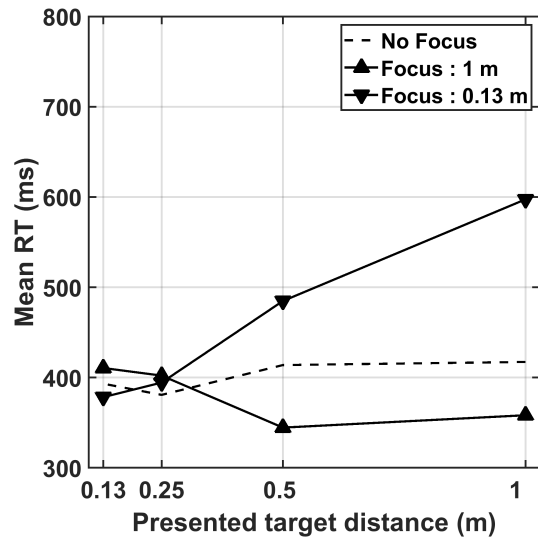
Figure B.9: Reaction time experiment results for Listener 9. (a) mean RT for sounds presented from -90° , (b) mean RT for sounds presented from $+90^\circ$, (c) mean RT for sounds presented from 0° . Full black lines are results when including the intensity cue, dashed white lines are results when excluding the intensity cue. Diamonds are results for Same Distance condition, Triangles for Distance Separation condition.

Appendix C

Top-down attention on distance - individual results

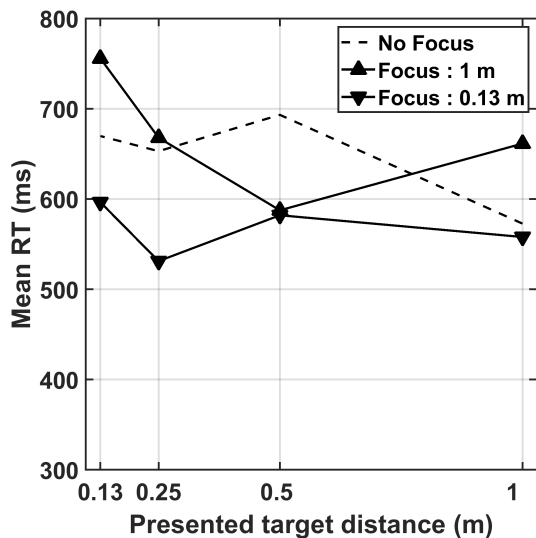


(a) Sources presented from -90° .

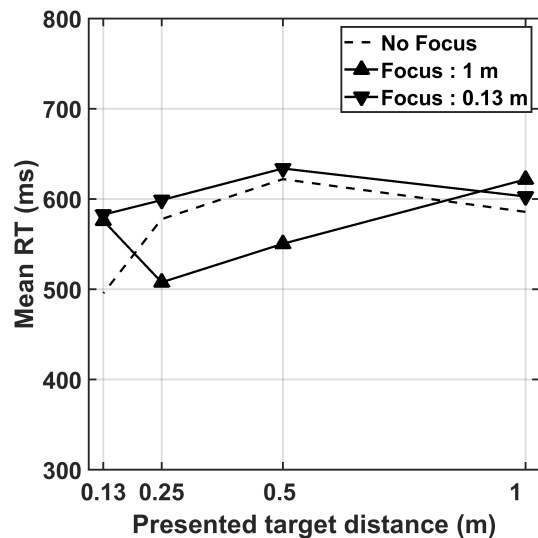


(b) Sources presented from 0° .

Figure C.1: Reaction time experiment results for Listener 1. (a) mean RT for sounds presented from -90° , (b) mean RT for sounds presented from 0° . Dashed lines are results when no focus is attempted on distance, upper triangles are results when focus is forced on 1 m, lower triangles when forced on 0.13 m.

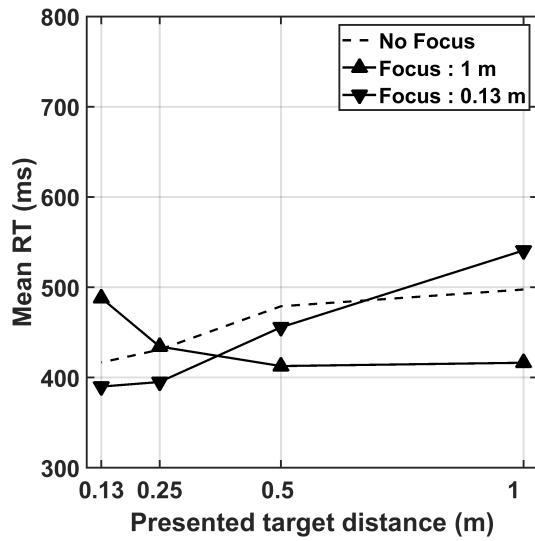


(a) Sources presented from -90° .

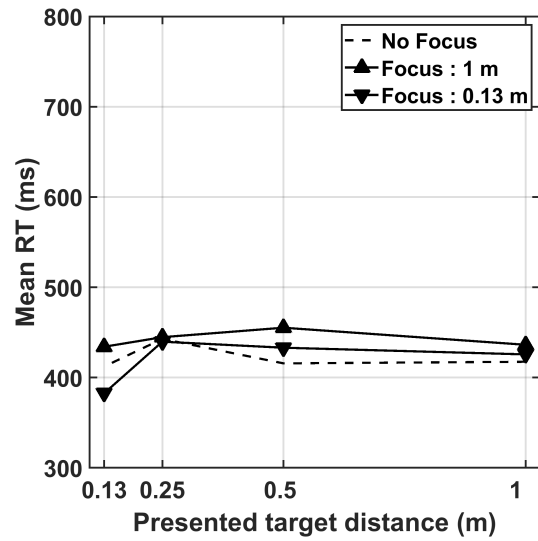


(b) Sources presented from 0° .

Figure C.2: Reaction time experiment results for Listener 2. (a) mean RT for sounds presented from -90° , (b) mean RT for sounds presented from 0° . Dashed lines are results when no focus is attempted on distance, upper triangles are results when focus is forced on 1 m, lower triangles when forced on 0.13 m.

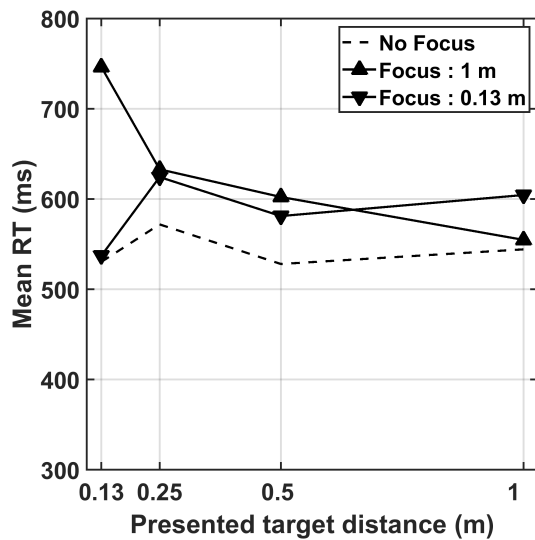


(a) Sources presented from -90° .

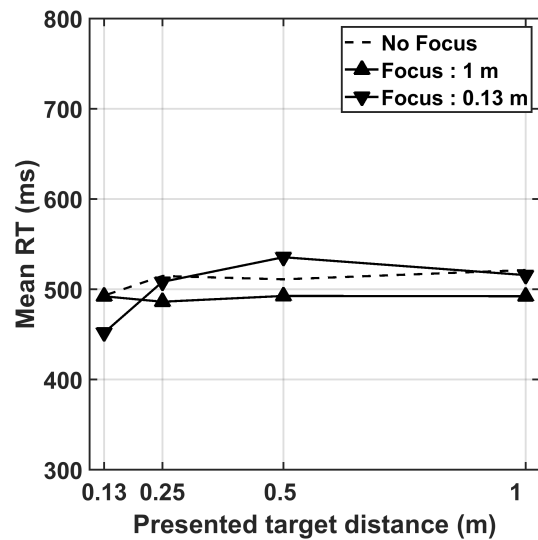


(b) Sources presented from 0° .

Figure C.3: Reaction time experiment results for Listener 3. (a) mean RT for sounds presented from -90° , (b) mean RT for sounds presented from 0° . Dashed lines are results when no focus is attempted on distance, upper triangles are results when focus is forced on 1 m, lower triangles when forced on 0.13 m.

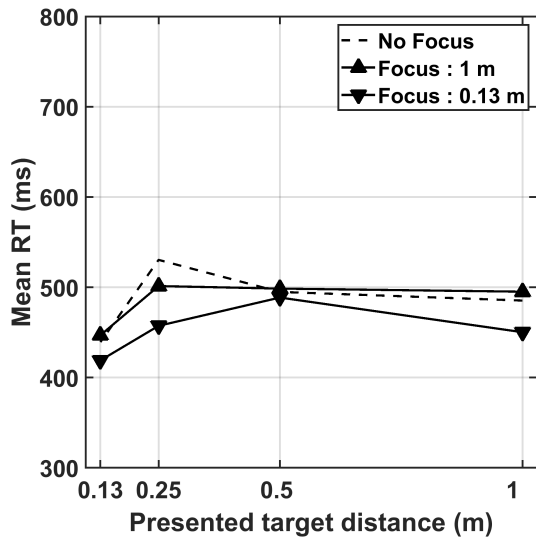


(a) Sources presented from -90° .

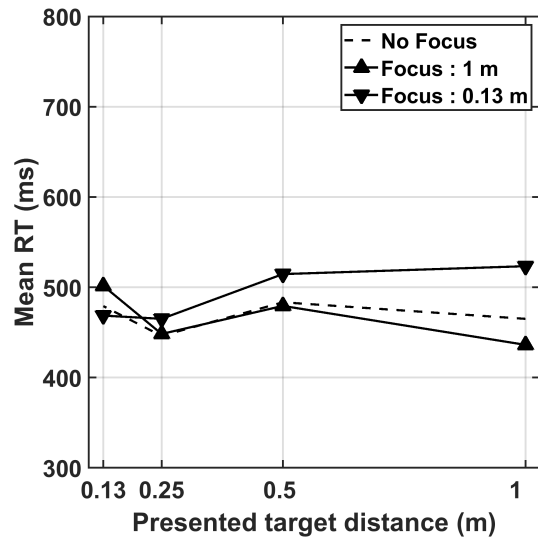


(b) Sources presented from 0° .

Figure C.4: Reaction time experiment results for Listener 4. (a) mean RT for sounds presented from -90° , (b) mean RT for sounds presented from 0° . Dashed lines are results when no focus is attempted on distance, upper triangles are results when focus is forced on 1 m, lower triangles when forced on 0.13 m.

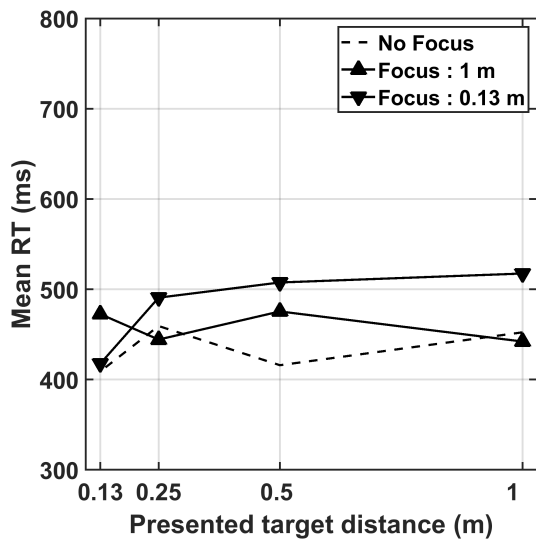


(a) Sources presented from -90° .

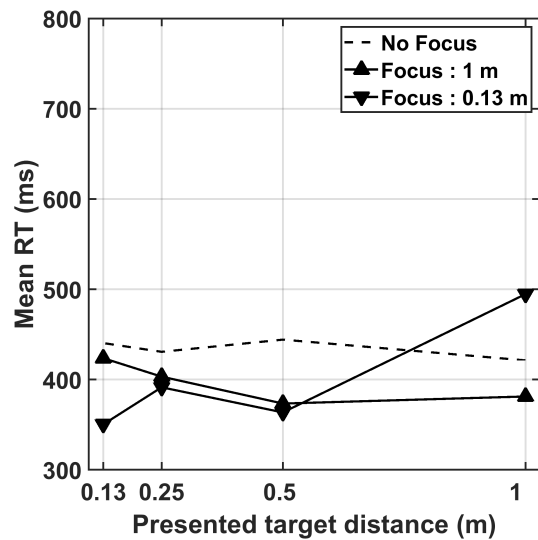


(b) Sources presented from 0° .

Figure C.5: Reaction time experiment results for Listener 5. (a) mean RT for sounds presented from -90° , (b) mean RT for sounds presented from 0° . Dashed lines are results when no focus is attempted on distance, upper triangles are results when focus is forced on 1 m, lower triangles when forced on 0.13 m.

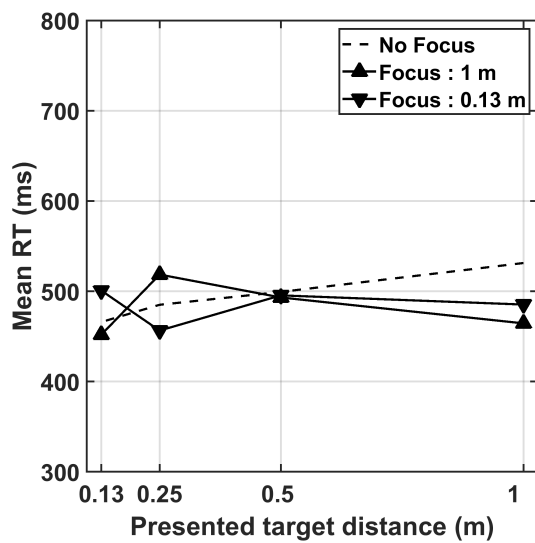


(a) Sources presented from -90° .

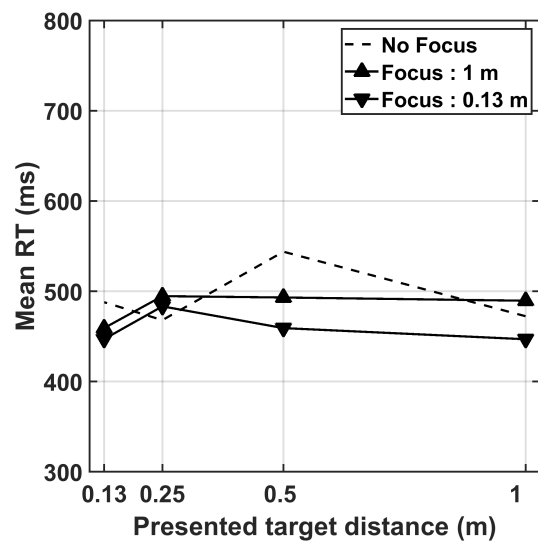


(b) Sources presented from 0° .

Figure C.6: Reaction time experiment results for Listener 6. (a) mean RT for sounds presented from -90° , (b) mean RT for sounds presented from 0° . Dashed lines are results when no focus is attempted on distance, upper triangles are results when focus is forced on 1 m, lower triangles when forced on 0.13 m.



(a) Sources presented from -90° .



(b) Sources presented from 0° .

Figure C.7: Reaction time experiment results for Listener 11. Listener 11 did not participate in the experiments presented in chapters 2 and 3. (a) mean RT for sounds presented from -90° , (b) mean RT for sounds presented from 0° . Dashed lines are results when no focus is attempted on distance, upper triangles are results when focus is forced on 1 m, lower triangles when forced on 0.13 m.

Acknowledgments

This thesis is the fruit of many people's cooperation, understanding and labour. First of all, words are too few to express my gratitude to Professor Yôiti Suzuki who gave me the opportunity of leading this exchange program in Tohoku University and of working on a master's thesis in the laboratory that he presides. I must express my deepest thanks to Professor Shuichi Sakamoto, my supervisor, for his valuable guidance and advice which pushed to me to do my best, as well as his patience throughout these two years. I am deeply grateful to Prof. Shioiri and Prof. Horio for their participation on my dissertation committee and for their excellent work and teaching skill. I am also very grateful to Dr. César D. Salvador for his helpful advice concerning my degree and support in every day life in Japan.

I wish to acknowledge Prof. Tatsuya Hirahara and Prof. Daisuke Morikawa for their contribution in gathering valuable data for the evaluation of our methods. I also wish to express my acknowledgement to Prof. Satoshi Shioiri, Prof. Ichiro Kuriki, Prof. Yoshifumi Kitamura, Prof. Akinori Ito and all the members of the Graduate School of Information Sciences for their advice. Of course, I also wish to acknowledge Prof. Cui Zhenglie, Eng. Fumitaka Saito, Prof. Jorge Treviño, and all the members of the Suzuki - Sakamoto Laboratory for their support and valuable monitoring of my work. Among them, I would especially like to thank Ryo Teraoka, a hard working doctor's student, for his great help in evaluating my research, offering valuable advice and offering valuable moments. I am grateful to Ms. Miki Onodera, whose knowledge and help for thriving in Japan I could not have lived without.

I extend my gratitude to the Ministry of Education, Culture, Sports, Science and Technology of Japan, whose financial support enabled me to lead my life and research in Japan during the past two years.

I also wish to thank all of my professors and comrades in the Ecole Centrale Lyon in France, where I studied engineering and was introduced to the possibility of leading an exchange program in Japan. It is there that I gathered knowledge of acoustics and sound

processing, was introduced to the japanese language and culture, and found the inspiration to lead my life as an engineer in these fields. A special thanks goes to Ms. Mariko Akutsu, my japanese language teacher in ECL, for her help and teaching skills.

Finally, I must express my very profound gratitude to all those who have contributed to this step in my life : my dearest parents Alexandra and Guy, and siblings Chloé and Paul for their moral support and knowledgeable advice; my comrades in the laboratory and of course my friends in Sendai.

Bibliography

- [1] C. Cherry, “Some experiments on the recognition of speech, with one and two ears.,” *Journal of the Acoustical Society of America*, vol. 26, pp. 554–559, 1953.
- [2] A. S. Bregman, *Auditory scene analysis*. Cambridge, MA, USA: MIT Press, 1990.
- [3] A. S. Bregman and J. Campbell, “Primary auditory stream segregation and perception of order in rapid sequences of tones.,” *Journal of experimental psychology*, vol. 89, no. 2, p. 244, 1971.
- [4] B. C. Moore and H. Gockel, “Factors influencing sequential stream segregation,” *Acta Acustica United with Acustica*, vol. 88, no. 3, pp. 320–333, 2002.
- [5] A. R. A. Conway, N. Cowan, and M. F. Bunting, “The cocktail party phenomenon revisited: The importance of working memory capacity,” *Psychonomic Bulletin & Review*, vol. 8, pp. 331–335, Jun 2001.
- [6] B. Scharf, “Auditory attention: The psychoacoustical approach,” *Attention*, pp. 75–117, 1998.
- [7] W. James, *The principles of psychology*. Read Books Ltd, 2013.
- [8] M. Corbetta and G. L. Shulman, “Control of goal-directed and stimulus-driven attention in the brain,” *Nature reviews neuroscience*, vol. 3, no. 3, p. 201, 2002.
- [9] A. W. Bronkhorst, “The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions,” *Acta Acustica united with Acustica*, vol. 86, no. 1, pp. 117–128, 2000.
- [10] C. L. Folk, R. W. Remington, and J. C. Johnston, “Involuntary covert orienting is contingent on attentional control settings.,” *Journal of Experimental Psychology: Human perception and performance*, vol. 18, no. 4, p. 1030, 1992.
- [11] R. W. Remington, J. C. Johnston, and S. Yantis, “Involuntary attentional capture by abrupt onsets,” *Perception & Psychophysics*, vol. 51, no. 3, pp. 279–290, 1992.
- [12] N. Asemi, Y. Sugita, and Y. Suzuki, “Auditory search asymmetry between normal

- japanese speech sounds and time-reversed speech sounds distributed on the frontal-horizontal plane,” *Acoustical Science and Technology*, vol. 24, no. 3, pp. 145–147, 2003.
- [13] R. Chocholle, “Variation des temps de réaction auditifs en fonction de l’intensité à diverses fréquences,” *L’année psychologique*, vol. 41, no. 1, pp. 65–124, 1940.
- [14] D. L. Kohfeld, “Simple reaction time as a function of stimulus intensity in decibels of light and sound.,” *Journal of experimental psychology*, vol. 88, no. 2, p. 251, 1971.
- [15] G. Z. Greenberg and W. D. Larkin, “Frequency-response characteristic of auditory observers detecting signals of a single frequency in noise: The probe-signal method,” *The Journal of the Acoustical Society of America*, vol. 44, no. 6, pp. 1513–1523, 1968.
- [16] M. Ebata, T. Sone, and T. Nimura, “Improvement of Hearing Ability by Directional Information,” *J. Acoust. Soc. Am.*, vol. 43, no. 2, pp. 289–297, 1968.
- [17] G. Rhodes, “Auditory attention and the representation of spatial information,” *Perception & Psychophysics*, vol. 42, no. 1, pp. 1–14, 1987.
- [18] G. Kidd, T. L. Arbogast, C. R. Mason, and F. J. Gallun, “The advantage of knowing where to listen,” *J. Acoust. Soc. Am.*, vol. 118, no. 6, pp. 3804–3815, 2005.
- [19] R. Teraoka, S. Sakamoto, Z. Cui, Y. Suzuki, and S. Shioiri, “Effects of auditory selective attention on word intelligibility and detection threshold of narrow-band noise,” *The Journal of the Acoustical Society of America*, vol. 144, no. 3, pp. 1838–1838, 2018.
- [20] T. L. Arbogast and G. Kidd, “Evidence for spatial tuning in informational masking using the probe-signal method,” *J. Acoust. Soc. Am.*, vol. 108, no. 4, pp. 1803–1810, 2000.
- [21] D. S. Brungart and W. M. Rabinowitz, “Auditory localization of nearby sources. Head-related transfer functions,” *J. Acoust. Soc. Am.*, vol. 106, pp. 1465–1479, Sept. 1999.
- [22] D. S. Brungart, N. I. Durlach, and W. M. Rabinowitz, “Auditory localization of nearby sources. ii. localization of a broadband source,” *The Journal of the Acoustical Society of America*, vol. 106, no. 4, pp. 1956–1968, 1999.
- [23] M. Otani and T. Hirahara, “Numerical study on source-distance dependency of head-related transfer functions,” *J. Acoust. Soc. Am.*, vol. 125, pp. 3253–3261, May 2009.
- [24] S. Okada and T. Hirahara, “Perception of approaching and retreating sounds,” *Acous-*

tical Science and Technology, vol. 36, no. 5, pp. 449–452, 2015.

- [25] P. M. Zurek, “Binaural advantages and directional effects in speech intelligibility,” *Acoustical factors affecting hearing aid performance*, vol. 2, pp. 255–275, 1993.
- [26] D. o. P. . A. Georgia State University, “HyperPhysics, by Georgia State University.” Accessed: 2019-01-08.
- [27] B. Shinn-Cunningham, “Learning reverberation: Considerations for spatial auditory displays,” Georgia Institute of Technology, 2000.
- [28] A. W. Bronkhorst and T. Houtgast, “Auditory distance perception in rooms,” *Nature*, vol. 397, pp. 517–520, Feb. 1999.
- [29] G. v. Békésy, “Über die entstehung der entfernungsempfindung beim hören. (On the origin of the sensation of distance in hearing),” *Akustische Zeitschrift*, 1938.
- [30] D. Mershon and L. King, “Intensity and reverberation as factors in the auditory perception of egocentric distance,” *Perception & Psychophysics*, vol. 18, no. 6, pp. 409–415, 1975.
- [31] D. H. Mershon and J. N. Bowers, “Absolute and relative cues for the auditory perception of egocentric distance,” *Perception*, vol. 8, no. 3, pp. 311–322, 1979.
- [32] K. U. Ingård, “A Review of the Influence of Meteorological Conditions on Sound Propagation,” *J. Acoust. Soc. Am.*, vol. 25, no. 3, pp. 405–411, 1953.
- [33] D. Brungart, “Auditory parallax effects in the HRTF for nearby sources,” in *Proc. IEEE WASPAA*, pp. 171–174, 1999.
- [34] Y. Suzuki and H.-Y. Kim, “A modelling of distance perception based on auditory parallax model,” in *Proceedings of the 16th International Congress on Acoustics and the 135th Meeting of the Acoustical Society of America*, (Seattle, USA), Acoustical Society of America, June 1998.
- [35] H.-Y. Kim, Y. Suzuki, S. Takane, and T. Sone, “Control of auditory distance perception based on the auditory parallax model,” *Appl. Acoust.*, vol. 62, pp. 245–270, Mar. 2001.
- [36] P. D. Coleman, “Failure to localize the source distance of an unfamiliar sound,” *The Journal of the Acoustical Society of America*, vol. 34, no. 3, pp. 345–346, 1962.
- [37] D. S. Brungart and K. R. Scott, “The effects of production and presentation level on the auditory distance perception of speech,” *The Journal of the Acoustical Society of America*, vol. 110, no. 1, pp. 425–440, 2001.

- [38] J.-L. Schwartz, F. Berthommier, and C. Savariaux, “Seeing to hear better: evidence for early audio-visual interactions in speech identification,” *Cognition*, vol. 93, no. 2, pp. B69–B78, 2004.
- [39] H. McGurk and J. MacDonald, “Hearing lips and seeing voices,” *Nature*, vol. 264, no. 5588, p. 746, 1976.
- [40] P. Zahorik, “Estimating sound source distance with and without vision,” *Optometry and vision science*, vol. 78, no. 5, pp. 270–275, 2001.
- [41] M. B. Gardner, “Proximity image effect in sound localization,” *The Journal of the Acoustical Society of America*, vol. 43, no. 1, pp. 163–163, 1968.
- [42] P. Zahorik, D. S. Brungart, and A. W. Bronkhorst, “Auditory distance perception in humans: A summary of past and present research,” *Acta Acust. United Ac.*, vol. 91, no. 3, pp. 409–420, 2005.
- [43] P. Zahorik and F. L. Wightman, “Loudness constancy with varying sound source distance,” *Nature Neuroscience*, vol. 4, pp. 78–73, 2001.
- [44] L. Marshall and J. F. Brandt, “The relationship between loudness and reaction time in normal hearing listeners,” *Acta oto-laryngologica*, vol. 90, no. 1-6, pp. 244–249, 1980.
- [45] M. S. A. Graziano, L. A. J. Reiss, and C. G. Gross, “A neuronal representation of the location of nearby sounds,” *Nature*, vol. 397, pp. 428–430, Feb. 1999.
- [46] A. Farnè and E. Ladavas, “Auditory peripersonal space in humans,” *Journal of cognitive neuroscience*, vol. 14, no. 7, pp. 1030–1043, 2002.
- [47] E. Canzoneri, E. Magosso, and A. Serino, “Dynamic sounds capture the boundaries of peripersonal space representation in humans,” *PloS one*, vol. 7, no. 9, p. e44306, 2012.
- [48] E. Ladavas, F. Pavani, and A. Farne, “Auditory peripersonal space in humans: a case of auditory-tactile extinction,” *Neurocase*, vol. 7, no. 2, pp. 97–103, 2001.
- [49] B. G. Shinn-Cunningham, J. Schickler, N. Kopčo, and R. Litovsky, “Spatial unmasking of nearby speech sources in a simulated anechoic environment,” *The Journal of the Acoustical Society of America*, vol. 110, no. 2, pp. 1118–1129, 2001.
- [50] D. S. Brungart and B. D. Simpson, “The effects of spatial separation in distance on the informational and energetic masking of a nearby speech signal,” *The Journal of the Acoustical Society of America*, vol. 112, no. 2, pp. 664–676, 2002.

- [51] J. Blauert, *Spatial hearing: The psychophysics of human sound localization*. Cambridge, MA, USA; London, England.: MIT Press, revised ed., 1997.
- [52] E. M. Wenzel, M. Arruda, D. J. Kistler, and F. L. Wightman, “Localization using nonindividualized head-related transfer functions,” *J. Acoust. Soc. Am.*, vol. 94, no. 1, pp. 111–123, 1993.
- [53] A. W. Bronkhorst, “Localization of real and virtual sound sources,” *The Journal of the Acoustical Society of America*, vol. 98, no. 5, pp. 2542–2553, 1995.
- [54] Y. Suzuki, F. Asano, H.-Y. K. Kim, and T. Sone, “An optimum computer-generated pulse signal suitable for the measurement of very long impulse responses,” *J. Acoust. Soc. Am.*, vol. 97, pp. 1119–1123, Feb. 1995.
- [55] D. N. Zotkin, R. Duraiswami, E. Grassi, and N. A. Gumerov, “Fast head-related transfer function measurement via reciprocity,” *J. Acoust. Soc. Am.*, vol. 120, pp. 2202–2215, Oct. 2006.
- [56] N. Matsunaga and T. Hirahara, “Reexamination of fast head-related transfer function measurement by reciprocal method,” *Acoustical Science and Technology*, vol. 31, no. 6, pp. 414–416, 2010.
- [57] 今井 悠貴, 森川 大輔, and 平原 達也, “相反法による頭部伝達関数計測に用いる超小型動電型スピーカユニットの音響特性,” *日本音響学会誌*, vol. 68, no. 10, pp. 513–519, 2012.
- [58] D. Y. N. Zotkin, J. Hwang, R. Duraiswaini, and L. S. Davis, “HRTF personalization using anthropometric measurements,” in *Applications of Signal Processing to Audio and Acoustics, 2003 IEEE Workshop on.*, pp. 157–160, Oct. 2003.
- [59] C. D. Salvador, S. Sakamoto, J. Treviño, and Y. Suzuki, “A new signal processing procedure for stable distance manipulation of circular HRTFs on the horizontal plane,” in *Proc. Spring Meeting Acoust. Soc. Jpn.*, (Yokohama, Japan), pp. 561–564, Acoustical Society of Japan, Mar. 2016.
- [60] C. D. Salvador, S. Sakamoto, J. Trevino, and Y. Suzuki, “Distance-varying filters to synthesize head-related transfer functions in the horizontal plane from circular boundary values,” *Acoustical Science and Technology*, vol. 38, no. 1, pp. 1–13, 2017.
- [61] V. R. Algazi, R. O. Duda, and D. M. Thompson, “The use of head-and-torso models for improved spatial sound synthesis,” (Los Angeles, CA, USA), Audio Engineering Society, Oct. 2002.
- [62] M. Otani and S. Ise, “Fast calculation system specialized for head-related transfer function based on boundary element method,” *J. Acoust. Soc. Am.*, vol. 119, pp. 2589–

2598, May 2006.

- [63] K.-S. Lee and S.-P. Lee, "A relevant distance criterion for interpolation of head-related transfer functions," *IEEE Trans. Audio, Speech, Language Process.*, vol. 19, pp. 1780–1790, Aug. 2011.
- [64] M. Pollow, K.-V. Nguyen, O. Warusfel, T. Carpentier, M. Müller-Trapet, M. Vorländer, and M. Noisternig, "Calculation of head-related transfer functions for arbitrary field points using spherical harmonics," *Acta Acust. United Ac.*, vol. 98, pp. 72–82, Jan. 2012.
- [65] D. G. Malham, "Rme babyface pro." Updated version of paper presented at AES UK "Second Century of Audio" Conference, London, 7 -8th June 1999.
- [66] D. S. Brungart and B. D. Simpson, "Auditory localization of nearby sources in a virtual audio display," in *Applications of Signal Processing to Audio and Acoustics, 2001 IEEE Workshop on the*, pp. 107–110, IEEE, 2001.
- [67] T. Qu, Z. Xiao, M. Gong, Y. Huang, X. Li, and X. Wu, "Distance-Dependent Head-Related Transfer Functions Measured With High Spatial Resolution Using a Spark Gap," *IEEE Trans. Audio, Speech, Language Process.*, vol. 17, pp. 1124–1132, Aug. 2009.
- [68] T. Kondo, S. Amano, S. Sakamoto, and Y. Suzuki, "Development of familiarity-controlled word-lists (fw07)," in *IEICE Society Conference research report*, vol. 107, pp. 43–48, 2008.
- [69] S. Amano, S. Sakamoto, T. Kondo, and Y. Suzuki, "Development of familiarity-controlled word lists 2003 (fw03) to assess spoken-word intelligibility in japanese," *Speech Communication*, vol. 51, no. 1, pp. 76–82, 2009.
- [70] G. Kidd Jr, C. R. Mason, T. L. Rohtla, and P. S. Deliwala, "Release from masking due to spatial separation of sources in the identification of nonspeech auditory patterns," *The Journal of the Acoustical Society of America*, vol. 104, no. 1, pp. 422–431, 1998.
- [71] D. S. Brungart, B. D. Simpson, M. A. Ericson, and K. R. Scott, "Informational and energetic masking effects in the perception of multiple simultaneous talkers," *The Journal of the Acoustical Society of America*, vol. 110, no. 5, pp. 2527–2538, 2001.
- [72] R. L. Freyman, U. Balakrishnan, and K. S. Helfer, "Spatial release from informational masking in speech recognition," *The Journal of the Acoustical Society of America*, vol. 109, no. 5, pp. 2112–2122, 2001.

List of works

Domestic Conferences

1) Florent Monasterolo, S. Sakamoto, C. D. Salvador, Z. Cui, and Y. Suzuki, "The effect of target speech distance on reaction time under multi-talker environment" in Proceedings of Auditory Res. Meeting, The Acoustical Society of Japan, Vol. 8 , No. 47, Wajima, Japan, 2018.

2) Florent Monasterolo, S. Sakamoto, C. D. Salvador, Z. Cui, and Y. Suzuki, "The effect of target speech distance on spatial auditory attention under multi-talker environment" in Proceedings of the Acoustical Society of Japan spring meeting, 1-11-4, Tokyo, Japan, 2019.

Domestic Workshop

1) Florent Monasterolo, S. Sakamoto, C. D. Salvador, Z. Cui, and Y. Suzuki, "The effect of sound source distance on reaction time under multi-talker environment" in Tohoku U-NTU Symposium: "When AI Meets Human Science", Sendai, Japan, 2018.