Perception of vertically separated sound sources in the median plane

Taeho Kim

School of Electrical Engineering

Thesis submitted for examination for the degree of Master of Science in Technology.

Espoo 7.11.2017

Thesis supervisor:

Prof. Ville Pulkki

Thesis advisor:

M.Sc. Henri Pöntynen

Aalto University
School of Electrical Engineering

AALTO UNIVERSITY
SCHOOL OF ELECTRICAL ENGINEERING

ABSTRACT OF THE
MASTER'S THESIS

Author: Taeho Kim

Title: Perception of vertically separated sound sources in the median plane

Date: 7.11.2017 Language: English Number of pages: 8+46

Department of Signal Processing and Acoustics

Professorship:

Supervisor: Prof. Ville Pulkki

Advisor: M.Sc. Henri Pöntynen

The ability of human listeners to segregate two sound sources was examined by conducting an experiment when the sources are concurrently presented from different directions in the median plane. A high-pass filtered pink noise was utilized as a sound stimulus in a free-field condition and presented as either a pair of incoherent sound sources or a single-source. Subjects responded whether they perceived sound from one or two directions. Listening tests were conducted with different directions and separation angles of sound sources. These tests consisted of two sessions: a monaural session when only the right ear was made audible, and a binaural session when both ears were audible. The results indicated that the percentage of responding "two directions" for pairwise stimuli exceeded 50% above 33.75° separation angle and reached above 70% at 67.5° separation for both sessions. However, the perceived separation showed weak correlation to the degree of separation although it increased in the binaural session. The ability to discriminate pairwise stimuli to each of two corresponding sound sources showed high statistical significance. The difference between a monaural hearing and binaural hearing was not statistically significant for the segregation of sound sources in the median plane.

Keywords: localization, median plane, vertical sources, segregation, monaural perception

Preface

I want to thank Professor Ville Pulkki and my instructor Henri Pöntynen for their great guidance and dedication on my research.

Otaniemi, 7.11.2017

Taeho Kim

Contents

Symbols and abbreviations

Symbols

H_{ec}	the response at the ear canal
H_{ff}	the response at the center of the head
H_{HRTF}	the head related transfer function
Hz	unit of frequency
c	speed of sound
p	calculated probability
r	distance to the source from the origin
α	significance level
θ, δ	azimuth
ρ	elevation
χ^2	chi-square value

Abbreviations

CMAA	concurrent minimum audible angle
ERB	equivalent rectangular bandwidth
F	center frequency in kHz
FDTD	frequency domain time difference
HRTF	head related transfer function
ILD	inter-aural level difference
ITD	inter-aural time difference
JND	just noticeable difference
LOC	lower source + center source configuration
MAA	minimum audible angle
N1, N2	the first notch, the second notch
P1, P2	the first peak, the second peak
PRTF	pinna related transfer function
SYM	symmetric pair configuration
UPC	upper source + center source configuration

List of Figures

List of Tables

1 Introduction

The ability to perceive the direction of sound has been essential for the survival of the human being, since it locates the source of a sound which can allow the listener to decide whether to seek for or to run away from the sound source. Sound localization is still a crucial function for humans because it enables the listener to precisely perceive the surrounding space, even beyond the range of our sight. Therefore, it is valuable to research how the human locates sounds as well as the attributes of sound localization, both of which have been extensively studied in psychoacoustics [1, 2, 3, 4, 5, 6].

The most common way to investigate aspects of sound localization is to use a peripheral approach. The main topics studied using this approach include the difference in perception between sounds from a source to each of two ears, also known as inter-aural difference, and how the shape of our head affects sound localization. The ability to locate a sound source in a horizontal direction is empowered by the sensitivity in perceiving inter-aural differences [1, 2, 3, 4, 6]. Localization in other directions, such as vertical or front-back directions, is mostly explained by the influence exerted by the shape of the ear, head and torso. These perceptual abilities have been investigated based on the relation between the response at an ear canal and the direct response from a sound source, which is referred to as *the head related transfer function* (HRTF) [6, 7, 8, 9].

Apart from these peripheral approaches, the spatial hearing of the human is also induced by the central auditory system. Features in sound localization, including individual differences or adaptation in hearing, cannot always be explained by inter-aural differences or HRTFs. In addition to peripheral functions, human listeners are also able to accurately localize sounds through the effect of the central auditory system.

The majority of studies in psychoacoustics have only focused on localization of a single sound source or from a single direction. In everyday life, however, the sound around us does not always originate from a single source. Although a variety of sources produce sounds from multiple directions in an acoustic environment, perception of two or more sound sources and directions has been less studied than single-sound sources. Several studies have examined the attributes of sound perception from two or multiple directions by locating sound sources in a horizontal plane [1, 5, 10, 11], but only a few studies have thus far investigated the perception of two or more sound sources in a vertical plane [11, 12].

Hence, this thesis aims to determine the ability of human listeners to segregate two sound sources when concurrently presented from different vertical directions. Specifically, the thesis investigates the segregation ability of sound sources separated at angles ranging up to 90°, the influence of binaural hearing in the vertical segregation of sound sources, and the ability to discriminate two sound sources from a single source. This will be accomplished based on listening tests in an anechoic chamber. Different combinations of directions and separation angles of the two sound sources were utilized as sound stimuli for the test. The test results will mainly be discussed using a peripheral approach, while also considering the effect of the central auditory system on the results.

1.1 Structure of this thesis

The rest of this thesis is organized as follows. Chapter 2 reviews the previous studies on sound localization. Chapter 3 describes methods and experimental set-ups used to conduct the listening tests. Chapter 4 presents the analyses carried out from the data collected in the test as well as the results of the analyses. Chapter 5 interprets and discusses the results obtained in the data analysis. The last chapter concludes the thesis by summarizing the contributions and issues emerging from this study, followed by a suggestion of topics for further research.

2 Literature review

In everyday life, multiple sounds are simultaneously heard in most cases. They might come from different directions, and the sources of those sounds can either be closely located or widely separated from each other. It would therefore be of interest to know how humans locate each of the sound sources, or how well each source is distinguished from others when perceiving multiple sounds. These two aspects of auditory perception are referred to as *sound localization* and *auditory segregation*, respectively.

This thesis assumes two sound sources vertically separated in front of the head, and examines the ability of listeners to segregate the sources presented with different directions and separation angles. In order to determine the segregation ability of sound sources separated in a vertical direction, it is necessary to understand both the sound localization in a vertical plane and the perception of multiple sound sources.

Therefore, this chapter first introduces the basic concepts and terms used in localization research (Section 2.1), then reviews the previous studies with respect to those two research subjects: *sound localization in the median plane* (Section 2.2) and *the perception of sounds from multiple directions* (Section 2.3). These two topics are followed by a review of the studies concerning *perception of sounds from multiple directions in the median plane* (Section 2.4), which is the intersection of the two subjects discussed in Section 2.2 and 2.3. The research topics discussed in Section 2.4 are the subjects most closely related to the purpose of this thesis.

2.1 Basic concepts for sound localization

In Psychoacoustics, the term "localization" refers to the perception of direction and distance of a sound source [1]. It has been known that localization is guided by several localization cues which function as combined effects in the human auditory system [1, 2, 3, 4]. Localization cues are generally categorized as *binaural cues* and *monaural cues* [4]. Binaural cues are obtained from the differences between sounds arriving at the two ears, which is referred to as the inter-aural difference. The two major binaural cues are *the inter-aural time difference* (ITD) and *the inter-aural level difference* (ILD). Monaural cues are derived from the sound perceived at one ear only or equally at both ears. The prime monaural cues are *spectral cues*, which are acquired from the dependence of the magnitude spectrum measured at the ear canal on the direction of the sound source.

2.1.1 Coordinate system used for sound localization

Localization cues can be effectively explained using a spherical coordinate system designed for locating a sound source relative to the human head. Figure 1 depicts the coordinate system for localization and the three directional planes defined by Blauert [4]. As shown in Figure 1, a sound source is localized using the distance to

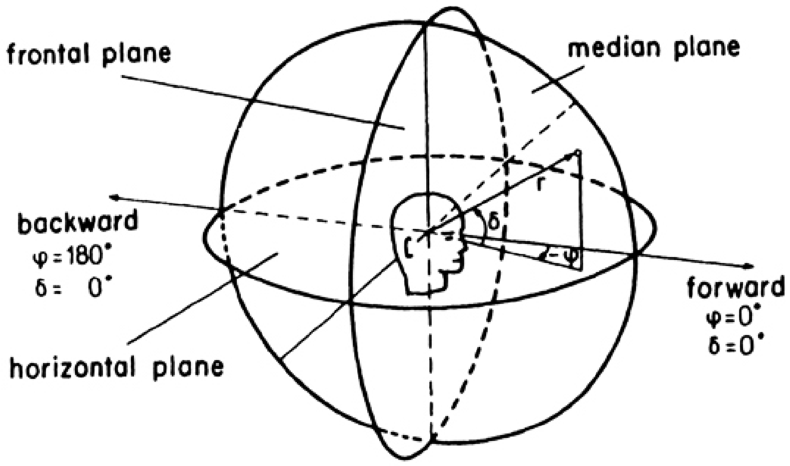


Figure 1: A head related coordinate system used for the localization. r is the distance, φ is given by the azimuth angle and δ is given by the elevation angle. Adapted from [4].

the sound source from the origin (r) and two polar coordinates: *azimuth* (φ) and *elevation* (δ). Azimuth (φ) is referred to as an angle assigning a horizontal direction and elevation (δ) is denoted as an angle for a vertical direction. For example, 0° azimuth and 0° elevation represent a sound source located straight in front of the head, and 180° azimuth and 0° elevation represent a source located directly behind the head. A sound source that lies straight above the head has a direction with 0°

azimuth and 90° elevation. The three directional planes shown in Figure 1 can also be explained by these polar coordinates. The horizontal plane covers all azimuth directions at 0° elevation, whereas the median plane covers all elevation directions at 0° azimuth. The frontal plane includes all elevation directions at 90° and −90° azimuth. These three directional planes intersect at the origin of the coordinate system.

2.1.2 Inter-aural Time Difference

The inter-aural time difference (ITD) refers to the difference in the time of arrival between sounds at the two ears, which originate from the same sound source. If the shape of a head is assumed to be circular by a view from above (see Figure 2), the difference between the paths from the source to each of two ears can be calculated as $r\theta + r\sin\theta$, where r is the radius of the head and θ is the azimuth. Since a time difference can be computed from the path difference, the ITD model can be derived as

$$\tau = \frac{r}{c}(\theta + \sin\theta) \tag{1}$$

where c is the speed of sound.

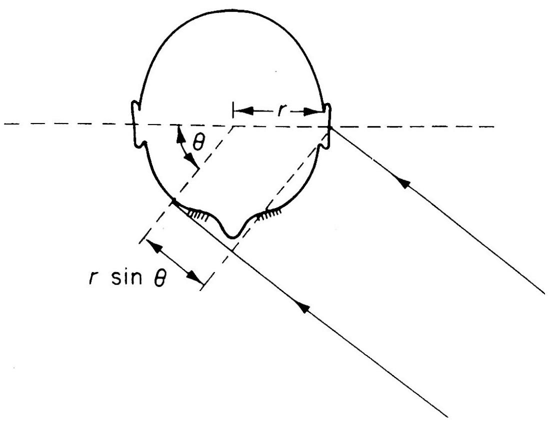


Figure 2: An illustration of ITD in terms of path difference. r is the radius of the head and θ is given by the azimuth. Adapted from [1].

This model calculates the ITD as $0s$ when a sound source lies directly in front of the head (0° azimuth) and approaches the maximum value on the straight line crossing the two ears (±90° azimuth). Assuming the average size of the human head, the maximum ITD value reaches approximately 0.6-0.7ms. Figure 3 shows the computed ITD values as a function of azimuth, compared with the measured values tested by Mills [5]. As shown in the figure, the ITD model agrees well with the real measurements.

Regarding the frequency dependence of ITD, sounds with low frequencies provide clear cues, though ambiguous cues can occur in localization for high frequency sounds.

Figure 3: Computed ITD values compared with measured ITD values as a function of azimuth. The x-axis is given by azimuth and the y-axis represents ITD in milliseconds. The dashed line with marker **x** shows the computed ITD and the straight line with marker **o** shows the measured values. Adapted from [5].

When the period of a sound is shorter than twice the maximum ITD value, the phase difference between sounds arriving at the two ears will exceed two or more cycles in the sound, causing confusion to the auditory system in determining the number of cycles corresponding to the phase difference. Considering the average size of the human head, the frequency corresponding to twice the maximum ITD is about 700-800Hz, indicating that sounds above those frequencies might provide localization ambiguity. Nevertheless, it has been known that phase differences can be detected at frequencies up to 1.6kHz, and the frequency dependence in ITD becomes significant above 1.6kHz [4, 6]. Therefore, it has been suggested that the ambiguity at high frequency sounds up to 1.6kHz may be resolved by perceiving the difference between temporal envelopes of sounds [6], or by the movements of either the head or the sound source [1].

Assuming the geometry used in the ITD model (see Figure 2), the circular shape of the head results in the same ITD value at the points with the same absolute azimuth angle symmetric to the line crossing the ears. In terms of a three-dimensional model, the shape of the head can be regarded as a sphere with two holes, and the area containing the locations with the same ITD value will form a conical surface next to each of the holes, as shown in Figure 4. This surface providing the same ITD is referred to as *the cone of confusion* [5], on the surface of which the same cues can lead to localization ambiguity. Since this model is based on the simple geometry that disregards the shape of the pinna (the outer ear), the asymmetry of the "real" shape of the head and the possibility of head movements, the localization ambiguity at the cone of confusion can be resolved from a few auditory mechanisms, including spectral cues and the effects of head movements [1, 6]. These mechanisms will be discussed in Section 2.1.4 and 2.1.5.

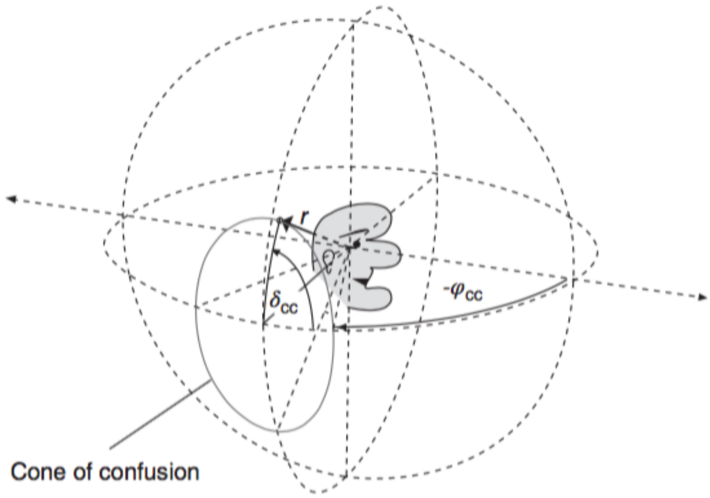


Figure 4: An illustration of the cone of confusion. φ_{cc} represents the azimuth on the horizontal plane reaching a cone of confusion area, and δ_{cc} is the angle on the cone of confusion. Adapted from [6].

2.1.3 Inter-aural Level Difference

The inter-aural level difference (ILD) refers to the difference between intensity levels of sounds at the two ears, which originate from the same sound source. In contrast to the ITD, the primary cause of level difference is the scattering of sound waves by the head rather than the path difference between the ears. Scattering of sound waves includes reflection, diffraction and shadowing, all of which are highly frequency dependent. When a sound has a wavelength longer than the size of the head, the sound wave will be diffracted around the head. Conversely, when a wavelength of a sound is shorter than the size of the head, the sound wave will be reflected and shadowed at the opposite side of the head [1]. Consequently, considering the average size of the human head, the ILDs from frequencies below 400-500Hz are nearly imperceptible, though the dependence on azimuth tends to grow with increasing frequency. This phenomenon is clearly verified in Figure 5 showing the ILD as a function of both frequency and azimuth [6].

To summarize, the ITD functions effectively at low frequencies, whereas the ILD mostly occurs at high frequencies. This frequency dependence in binaural cues appears accurately for pure tones in the horizontal plane. However, this tendency does not completely explain the localization on complex, real sounds [1].

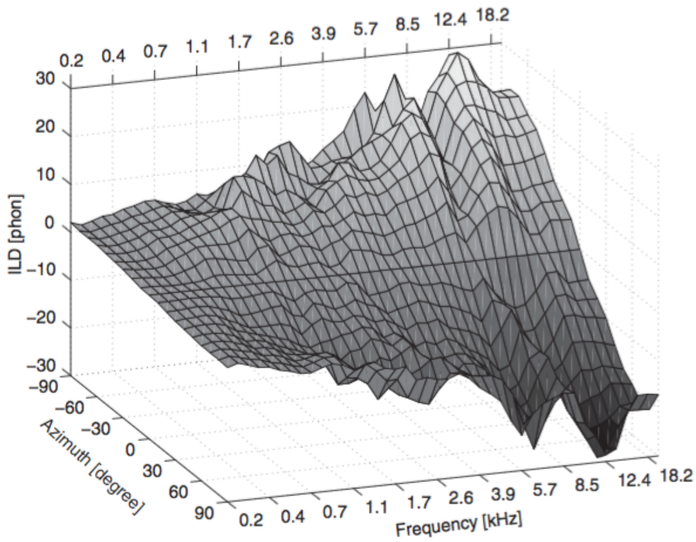


Figure 5: ILD expressed as loudness levels (phon) as a function of frequency and azimuth. Computed using an auditory model based on the difference between estimated loudness of ear-canal signals. Adapted from [6].

2.1.4 Spectral Cues

Monaural cues can be measured and derived from a sound signal arriving at the ear canal. Due to the scattering and reflection caused by the head, torso and pinna, the magnitude spectrum of the signal at the ear canal varies with the direction of a sound source. This dependence of the magnitude spectrum on the direction of the sound is referred to as *spectral cues*. As mentioned earlier, spectral cues are investigated using the transfer function of the response from a sound source at the ear canal, normalized by the direct response from the position where the center of the head might be located. This transfer function is defined as *head related transfer function* (HRTF) [1, 4, 6], which is expressed as

$$H_{HRTF}(\omega) = \frac{H_{ec}(\omega)}{H_{ff}(\omega)} \tag{2}$$

where H_{ec} is the response at the *ear canal* and H_{ff} is the *free-field* response at the point where the center of the head might be located [6].

Figure 6 shows an example set of HRTFs measured from three directions: the front-center direction ($\varphi = 0°$, $\delta = 0°$), a direction on the horizontal plane ($\varphi = 60°$, $\delta = 0°$) and a direction on the median plane ($\varphi = 0°$, $\delta = 60°$) [6]. The HRTFs at both ears in the front-center direction show similar results because of the bilateral symmetry of the human. However, when a sound source is located in a non-center position on the horizontal plane ($\varphi \neq 0°$), the HRTF at each ear is different from the other, as shown in (b). A high-frequency boost above around 1kHz is shown in the HRTF of the left ear whereas a significant drop at high frequencies can be found in the HRTF of the right ear. Since the sound source is located on the same side as the left ear ($\varphi = 60°$), the "shadowing" by the head might cause a high-frequency decline at the right ear. This difference roughly agrees with the ILD measurements shown in Figure 5. When a sound source is located on the median plane, the difference between HRTFs of both ears is small because of the symmetry, similar to the case at the front-center direction. However, comparing the front-center HRTFs (a) with median plane HRTFs (c), a difference is apparent at frequencies above around 6kHz. This spectral change is caused by different scattering conditions at the pinna induced by different source directions [1, 4, 8, 9, 13].

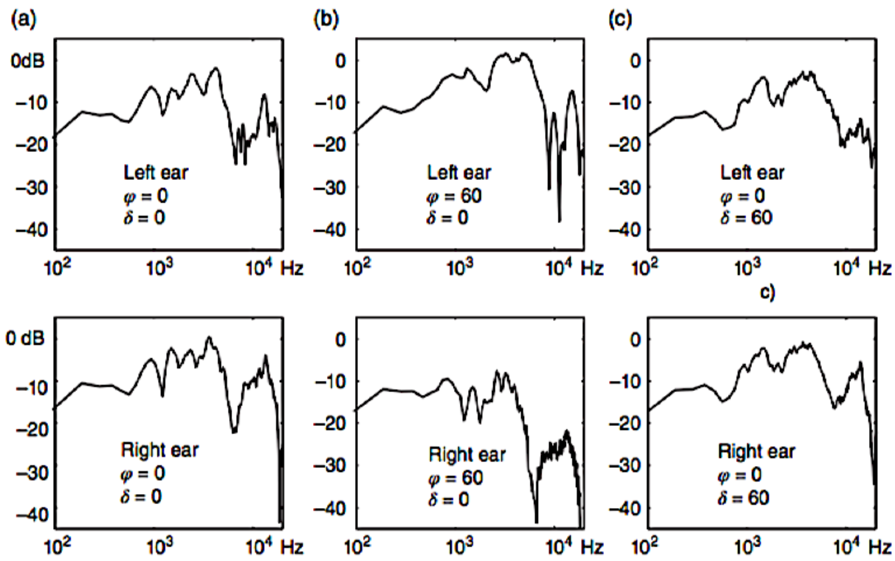


Figure 6: A set of HRTF measurements from a subject at the entrances of both ear canals. The three directions of sound source applied in this measurements are (a) $\varphi = 0°$, $\delta = 0°$, (b) $\varphi = 60°$, $\delta = 0°$ and (c) $\varphi = 0°$, $\delta = 60°$. Adapted from [6].

As described in the examples, the spectral changes in HRTFs indicate the changes in the direction of the sound source. The effects from the head, torso and pinna cause the alteration in magnitude spectrum between 500Hz and 16kHz [4, 13]. Changes in low-mid frequencies are mainly caused by the effects of the head and torso cues, whereas spectral changes above 4-5 kHz are produced by pinna cues, which are important for the localization of elevated sources [4, 8, 9, 14]. Pinna cues and the localization in the median plane will be explained in Section 2.2.

HRTFs generally appear as a frequency spectrum with peaks and notches around specific frequencies. However, the frequencies and the amplitudes of peaks and notches vary in each individual, since every human has different shape and size of the head, torso and pinna. Nevertheless, the shapes of HRTFs from many individuals show similar patterns. This similarity implies that the ability to locate a sound source is similar in every human although each individual has different sensitivity on localization. Different peripheral effects from different pinna shapes might be resolved to induce similar perception for every individual. On the other hand, the differences in HRTFs between individuals lead to the assumption that localization can be affected by learning and adaptation in the auditory system. It has been proved in a few experiments that when individual spectral cues are reshaped by a physical modification of the pinna [15], or a set of specific HRTFs is forced to each of the subjects' hearing [16], listeners continuously adapted to the different listening conditions in a few weeks, although the accuracy of localization was degraded, especially on the cone of confusion.

2.1.5 Dynamic Cues

It has been proven by a number of experiments that head movements improve localization accuracy [17]. As mentioned earlier, one of the common ideas for the role of the head movement is that it resolves the ambiguities in binaural cues, especially at the cone of confusion [1, 6, 18, 19]. When a listener horizontally moves one's head, it is assumed that a static sound source might not be localized correctly because of the changes in ITD and ILD. However, in most cases, static sources are perceived as horizontally stable in spite of the cone of confusion, which leads to the assumption that a mechanism resolving this localization error might exist in the auditory system since the magnitude of the changes in ITD and ILD due to head rotations also depend on the source location [6]. This idea implies that head movements provide localization cues in addition to binaural cues [1], which is referred to as *dynamic cues*.

A few studies showed that head movements can also reduce front-back localization errors. The ability to discriminate the front and back direction was significantly improved when subjects were allowed to move their head [20], and even small head movements clarified the front-back localization [21]. Since front-back localization is mostly influenced by spectral cues, these research induce the assumption that dynamic cues supplement spectral cues as well as binaural cues.

2.2 Sound localization in the median plane

2.2.1 Spectral cues for median plane localization

As mentioned in Section 2.1.4, it has been generally known that spectral cues are significant factors in the median plane localization, and spectral changes above 4-5kHz caused by the pinna are the major localization cue that affects the perception of elevation. This notion has been proved and established by numerous studies.

Mehrgardt and Mellert (1977) [22] measured the transfer functions of sounds from the free sound field to the ear canal entrance with nine elevation angles between 9° and 171°. The measurement within the frequency range between 500Hz and 14kHz showed that the dependence of the transfer function upon the elevation angle revealed significant spectral changes at frequencies above 5kHz.

Hebrank and Wright (1974) [23] carried out an experiment investigating the median plane localization using filtered white noises. The filter characteristics varied with low-pass, high-pass, band-pass and band-stop features having different cut-off frequencies. They reported that the median plane localization is based on spectral cues at frequencies between 4kHz and 16kHz, and suggested three spectral patterns representing the three different directions in the median plane: the "front" direction cue (0°) is represented by a 1-octave wide notch with lower cut-off frequencies at 4-8kHz and an increase above 13kHz, a 1/4 octave peak at 7-9kHz represents the "above" direction cue (90°), and a small peak at 10-12kHz with decreases above and below the peak represents the "back" direction cue (180°). They also measured the spectral patterns using artificial ears with a sound source located at different elevation angles, and verified that the spectral cues affecting the median plane localization is based on the directional filtering caused by the pinna.

In addition to Hebrank and Wright's research, a number of studies have attempted to interpret peaks and notches in a magnitude spectrum as spectral cues for elevation. Shaw and Teranishi (1968) [24] measured the pressure levels at the entrance of the ear canal and found a sharp notch near 8kHz at normal incidence directly in front of the head (0°). The frequency of the notch changed from 6kHz to 10kHz as the elevation angle moved from $-45°$ to $+45°$. Thus, the frequency of this notch and corresponding spectral pattern above 5kHz showed significant dependence on the vertical angle, as shown in Figure 7.

Butler and Belendiuk (1977) [13] carried out experiments using broadband noise bursts produced between 30° and $-30°$ elevation angle in the median plane and measured the sounds at the external ear canals of eight subjects. The analysis showed a prominent notch in the frequency response, and the center frequency of the notch was shifted from 7kHz to 5kHz as the sound source was moved from 30° to $-30°$ elevation, similar to Shaw and Teranishi's research.

By using these spectral features of the localization in vertical directions, Bloom (1977) [25] attempted to create an illusion of sound elevation by a spectral manipulation of the sound source. A sharp notch filter with varying center frequency was applied to the sound source, which was directly presented to the subjects' ear canal. Listeners detected an elevated "phantom" source from this manipulated sound presentation, supporting the idea that when conducting a spectral modification of

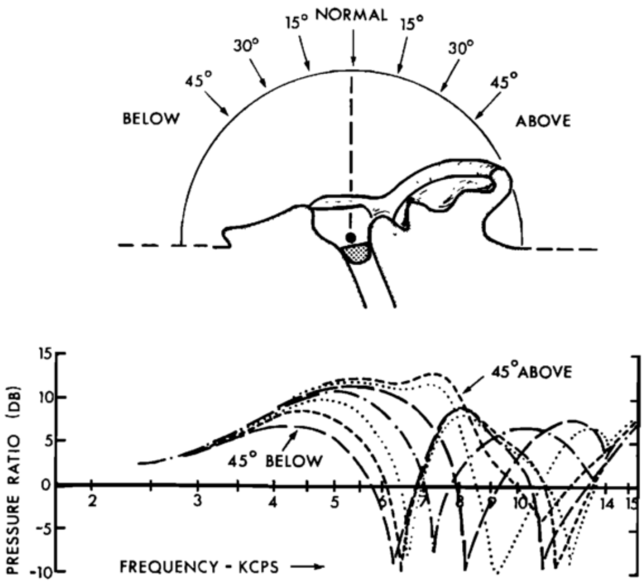


Figure 7: Frequency responses for a subject at the entrance of the ear canal from a sound source with vertical directions. Source distance: $r = 8cm$. Incidence angle: $\pm 45°$, $\pm 30°$ and $0°$ elevation. Adapted from [24].

the sound source, a sharp notch in high frequency is sufficient to create a perception of sound elevation.

Butler and Humanski (1992) [26] investigated localization in the median plane with a high-pass noise and a low-pass noise, both of which having the cut-off frequency at 3kHz. Since pinna cues operate above 4-5kHz, localization of a low-pass noise might be mainly affected by binaural cues. The result of this experiment showed that the ability to locate high-pass noise significantly exceeded the ability for low-pass noise localization in the median plane ($p < 0.01$). The perceived elevation with high-pass noise agreed well with actual elevation angle up to 60° but showed slight decrease above 60°, resulting in about 75° at the overhead direction (90°). However, localization proficiency with low-pass noise appeared to be poor above 30° elevation as shown in Figure 8. This result also confirms the dominating influence of pinna cues on the median plane localization.

Moore (1989) [27] investigated the threshold for detecting spectral peaks and notches at 1kHz and 8kHz, which can be regarded as the sensitivity for perceiving source elevation. The results of this experiment suggested that the sensitivity is enhanced by the dynamic changes of a sound spectrum, indicating movements of the sound source or movements of the head. The measured threshold for detecting a shift of the notch corresponded to the shift of 4° elevation around 0° in the median plane. This experiment supports the idea that the center frequency of the notch can be a cue for the median plane localization.

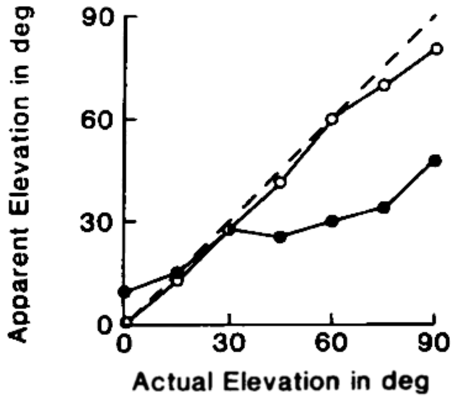


Figure 8: Perceived elevation of the sound source against the actual elevation angle in the median plane. Open circles present judgments from 3kHz high-pass noise bursts and closed circles present those from 3kHz low-pass noise bursts. Adapted from [26].

2.2.2 HRTF analysis for median plane localization

Recent studies have mostly used HRTF analysis to investigate median plane localization [8, 9, 14, 28]. The dashed lines shown in Figure 9 present examples of HRTFs for elevation angles from 0° to 180° in the upper median plane [14]. As can be seen in the figure, multiple peaks and notches are present in most of the HRTFs, which confirms the studies discussed in Section 2.2.1. When comparing these examples to the spectral patterns suggested by Hebrank and Wright (1974) [23] (see Section 2.2.1), the HRTF data show similar results to their spectral patterns. The two peaks at around 4kHz and 13kHz at 0° elevation agree well with the spectrum for "front" direction cue, and a small peak at around 12kHz as well as the two notches above and below the peak at 180° elevation correspond well with the spectrum for "back" direction cue. This early research suggests not only one prominent notch but also multiple peaks and notches for the spectral cues in median plane localization.

Regarding multiple peaks and notches in HRTFs, Ilda et al. (2007) [14] carried out an experiment to clarify which peak(s) and notch(es) in HRTFs function as directional cues for elevation. They created a parametric HRTF model by extracting peaks and notches from the measured HRTFs and by recomposing simplified HRTFs from them. For example, the solid lines in Figure 9 present parametric HRTFs recomposed from the lower two notches. Considering the frequency range of the spectral cues that contribute the perception of elevation, a peak around 4kHz was regarded as the lower frequency limit. The extracted peaks and notches were then labelled in the order of frequency: P1 and P2 for the lower two peaks, N1 and N2 for the lower two notches and so on. In order to clarify the spectral peaks and notches that provide cues for elevation, several combinations of peaks and notches were recomposed to form parametric HRTFs: N1, N2, P1, N1–N2, N1–P1, N2–P1, N1–N2–P1, as well as all peaks and notches. Each parametric HRTF from each combination was then compared to the measured HRTF in terms of localization accuracy by conducting

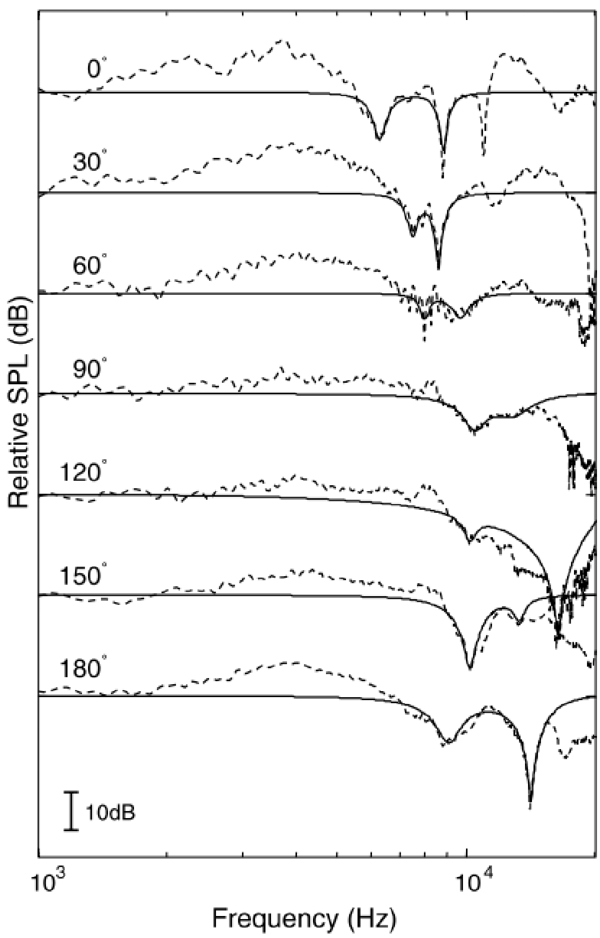


Figure 9: Examples of HRTFs for elevation in the upper median plane. The solid lines present parametric HRTFs recomposed from the lower two notches (N1,N2) and the broken lines presents the measured HRTF. Adapted from [14].

listening tests. The results showed that among all the combinations, the most similar localization accuracy to the measured HRTF is provided by parametric HRTFs with N1-N2-P1 as well as all peaks and notches. However, parametric HRTFs with N1, N2 and N1-N2 provided poor similarity to the measured HRTF.

These results suggest a few ideas for the median plane localization. Firstly, the sufficiency of the parametric HRTF with all peaks and notches implies that peaks and notches in HRTF exclusively function as spectral cues for the median plane localization. Secondly, the high similarity in the localization accuracy between the parametric HRTF with N1-N2-P1 and the measured HRTF indicates that not all peaks and notches are necessary to perceive elevation. Only one peak (P1) and two notches (N1, N2) are sufficient for the spectral cues, and a single notch (N1, N2) or notches without a peak (N1-N2) provide less accuracy in the median plane localization. Figure 10 shows an illustration of a typical peak-notch pattern of HRTFs in the median plane derived from the CIPIC HRTF database (*Algazi et al.*,2001) [29].

This simplified schematization showing only peaks and notches in HRTFs can thus represent the spectral cues in the median plane localization.

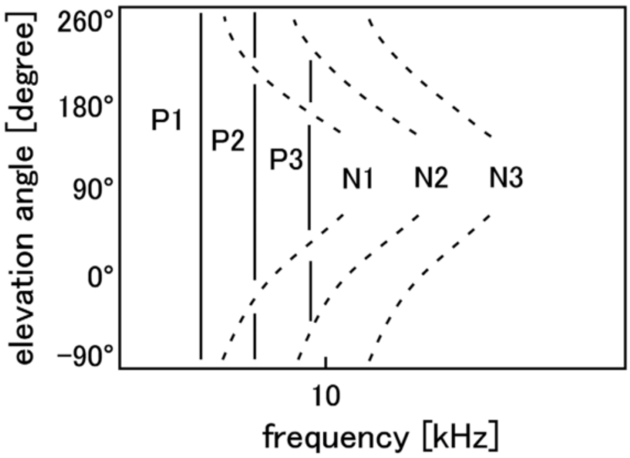


Figure 10: A schematized illustration of a typical peak-notch pattern of HRTFs in the median plane. HRTFs derived from the CIPIC HRTF database (*Algazi et al.*,2001) [29]. Adapted from [8].

As shown in Figure 10, The frequency of N1 and N2 shifts as the elevation angle changes, which shows that these spectral notches operate as potential elevation cues. However, P1 is shown to be independent of the elevation angle although it was verified to be essential for the perception of elevation in this test. Furthermore, the fact that poor localization accuracy was obtained from parametric HRTFs without any spectral peaks, such as HRTFs with N1, N2 and N1-N2, supports the need of P1 for an essential element of the median plane localization. The contribution of this spectral peak might be, according to Ilda et al. [14], the reference information that helps the auditory system to analyze N1 and N2. It will be easier to analyze the spectral information when a frequency band independent of the elevation angle exists, and the frequency of a notch is determined by the distance from the reference frequency band.

In addition to P1, the second notch (N2) also appeared as an essential element in the spectral cues for elevation. N2 is newly added to the single spectral notch mentioned in previous studies. The necessity of the second notch for median plane localization is proved from the test results showing that the parametric HRTF with P1-N1-N2 provides higher localization accuracy than HRTF without N2 (P1-N1).

In order to validate the role of the pinna as the main factor of these essential peaks and notches, Takemoto et al. (2012) [8] measured *the pinna related transfer functions* (PRTF) using *the frequency domain time difference* (FDTD) technique and compared with HRTFs of the same subject. The peak-notch pattern was almost identical among both HRTFs and PRTFs for all the subjects, although their spectral structures were slightly different in detail, as shown in Figure 11.

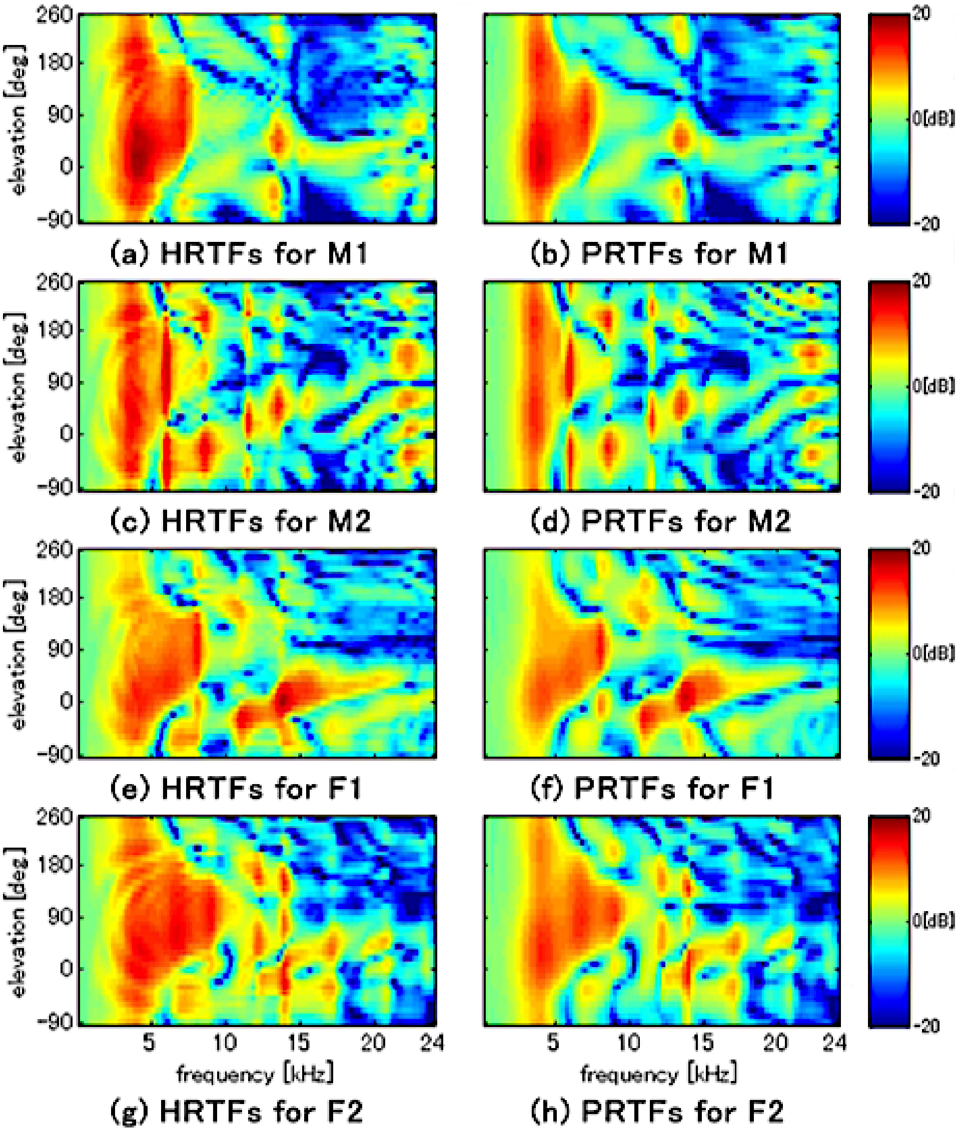


Figure 11: HRTFs and PRTFs in the median plane for four subjects. Adapted from [8].

2.2.3 Effects of head movements on the median plane localization

Localization in the median plane can be supported by head movements. When the head rotates, the spectral information at the ear will be altered continuously, and these spectral changes can enhance the sensitivity for perceiving elevation of a sound source, as mentioned in Section 2.2.1 [27]. To examine the effect of horizontal head rotation on the median plane localization, Perrett and Noble (1997) [30] carried out a listening test comparing 60° head rotation with motionless conditions using sound stimuli with low-pass, broadband and high-pass noise. The error rate in the perceived elevation was significantly reduced from low-pass noise, broadband noise and 1kHz high-pass noise when subjects repeatedly rotated their head. However, the results for 2kHz and 4KHz high pass noises, where the spectral cues operate for source elevation, did not show significant effect under the condition of head rotation. Moreover, the effect of rotation on the perception of source elevation was highly significant ($p < 0.001$) with low-pass noises filtered below 4kHz. These unexpected results can be interpreted that horizontal head movements improve median plane localization at low frequencies though those movements give little effect on pinna cues. When the head directs towards non-zero azimuth, and the elevation angle of the sound source changes, each path length from the source to each of the two ears will be changed differently. This difference in the change of each path length for each ear will thus produce different ITDs when the elevation angle of the source changes. Therefore, the dynamic cues created from horizontal head movements create low-frequency ITDs, and these interaural differences enable median plane localization even at lower frequencies below 4kHz.

2.3 Perception of sounds from multiple directions

Although the majority of studies in sound localization are based on the perception of sounds from a single source, the sound reaching the ears barely originates from a single source in everyday life. Thus, it is necessary to investigate the perception of sounds from multiple directions in the environment. Studies related to multiple sound sources can be categorized to two subjects: *the auditory grouping* and *the auditory segregation*, meaning how multiple sound sources are perceived as a grouped object or separate objects with respect to spectral attributes, temporal structures and a degree of separation. This chapter will mainly focus on the auditory segregation since this thesis studies the separability of two sound sources.

When two sounds simultaneously reach the ears, the auditory event differs by the coherence of the sources. Two entirely coherent sound sources, if their levels and times of arrival are not significantly different, are perceived as one "phantom" source located between two sound sources. This phenomenon is called *summing localization*, which is widely utilized for stereophonic reproductions. However, summing localization largely depends on the degree of coherence between two sources. According to Chernyak and Dubrovsky's experiment (1968) [31], the area where the auditory event appears becomes wider when the degree of coherence decreases. In other words, the phantom source that emerged when two sources are fully coherent becomes a wider auditory image as the two sources become less coherent. The widening of the auditory event in the horizontal plane causes the increase of *lateralization blur* [4], meaning that the sound sources are not clearly localized as a phantom source nor two separate sources. When the sound sources become fully incoherent, the auditory event then finally appears as two separated images.

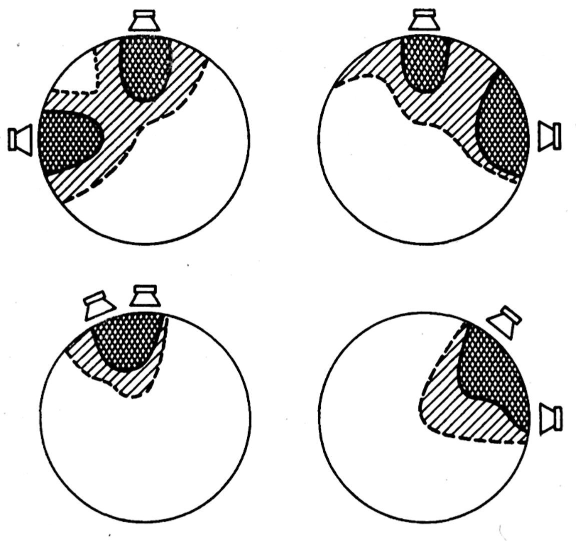


Figure 12: The auditory area when two incoherent broadband noises are presented from different direction in the horizontal plane. Adapted from [32].

The degree of separation also influences auditory segregation. Damaske (1968) [32] carried out an experiment using two incoherent broadband noises with varying azimuth directions. Figure 12 shows the diagrams describing a few test results by Damaske. The diagrams indicate the range of area where components of auditory events were found by listeners, and different shading within the circles represents different frequency of auditory events found. The upper two diagrams show the auditory events from two largely separated sound sources, and the lower ones give the results with closely located sound sources. It can be seen in the upper diagrams that the auditory events are mostly separated when the degree of separation is high. However, the auditory events from two close sound sources are fused into a single area although the sound sources were incoherent with each other. Therefore, it is necessary to investigate the precise relation between the degree of source separation and corresponding separation of auditory event, which can be referred to as the resolution of auditory space.

The relative resolution of auditory space was measured by Mills (1958) [33]. The precision of localization was investigated by measuring *just noticeable differences* (JND) of localization angle, which is denoted as *the minimum audible angle* (MMA). By conducting listening tests, MMAs were measured at five source directions ranging between 0° and 90° azimuth. A pure tone pulse was presented at a given azimuth, then presented again with a little deviation to the right or left of the reference azimuth angle. This process was repeated with increasing deviation until the subject could discriminate the difference in azimuth. Thirteen tone pulses that range between 250Hz and 8kHz were used for each of five source directions.

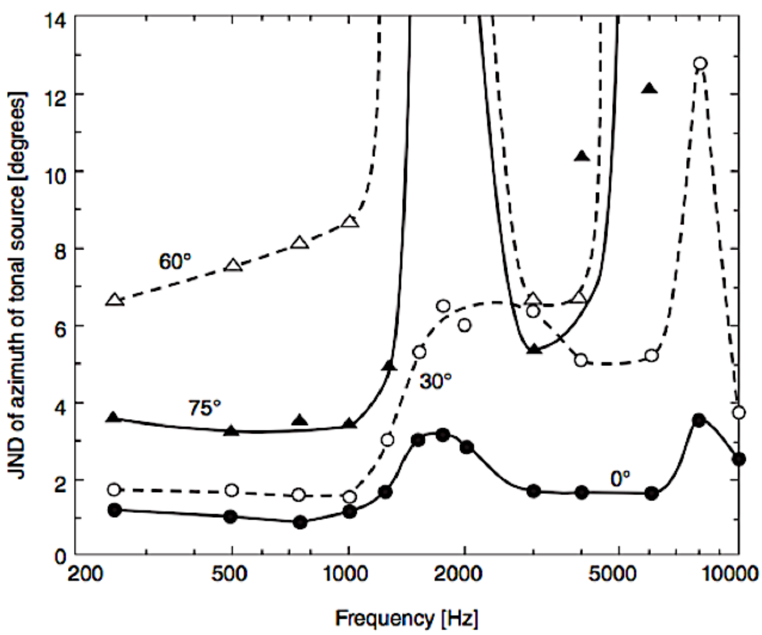


Figure 13: MMA at four azimuth angles (0°, 30°, 60°, and 75°) in the horizontal plane as a function of frequency. Adapted from [33].

The test result is depicted in Figure 13. The MMA of each azimuth angle shows a minimum value between 250Hz and 1kHz and steeply increases to a maximum above 1kHz. At frequencies between 3kHz and 6kHz, MMA meets another minimum then raises again. The minimum values below 1kHz are relatively constant, showing $1-2°$ at $0°$ and $30°$ azimuth, and $3-4°$ at $75°$ azimuth. These results represent the high sensitivity in horizontal plane localization at certain conditions.

The MMA is, however, measured from successive presentations of sound stimuli. Since the MMA is based on the detection of a minimum *change* in localization, it is not fully sufficient to examine the resolution of auditory space regarding the presence of coexisting sounds in the environment. In order to investigate the auditory perception of simultaneous sounds, Perrott (1984) [34] measured the MMA for concurrent stimuli, which is defined as *the concurrent minimum audible angle* (CMAA). He presented two tones of different frequencies at five positions in the horizontal plane. A pair of sound sources were separated around a midpoint which is given between $0°$ and $67°$ azimuth. Subjects were asked to detect the separation of a pair of tones by indicating whether the higher frequency tone is positioned at the right or left among the pairs, while the separation angles were varied between $3.3°$ and $46°$. The threshold of separation angle given at 75% of the correct answers was determined as the CMAA for each source position. As a result, the CMAA at the front ($0°$) was measured as $4-10°$, which was much larger than the MMA at that position ($1-2°$). When the sources were located around $67°$ azimuth, CMAAs increased significantly showing $30-45°$ separation. This test was carried out with various frequency differences of the two tones, and CMAAs tend to decrease with increasing frequency difference.

Divenyi and Oliver (1989) [35] conducted a similar experiment using various sound sources including frequency-modulated tones, amplitude-modulated tones, pure tones and noises. Since stimuli with broader spectra were utilized in the tests, the two sound sources could involve a common spectral region. The experiment showed that as the overlapping spectral region became wider between two concurrent stimuli, the spatial resolution became less efficient. In other words, the ability to segregate two sound sources was weakened by significant spectral overlapping between the sources.

Best et al. (2004) [11] examined the ability to segregate a pair of concurrent broadband noises with a complete spectral overlap in their long-term spectra. Therefore, little effects from the frequency separation of sound sources influenced the perception of spatial separation in this study. The sound stimulus used in this test was a pair of broadband noise bursts that were presented as either spatially coincident or separate sources. Tests were carried out with a headphone presentation using the virtual auditory space created from individual HRTFs. Pairs of horizontally separated sound sources were presented at five source locations between $0°$ and $90°$ azimuth in the horizontal plane. The separation angle was varied between $0°$ (spatially coincident) and $\pm 63°$. Subjects were asked to respond whether the stimuli were perceived as one or two sound sources. Figure 14 shows the percentage of perceiving two source locations as a function of separation angle between the sources. As shown in the left column in the figure, the auditory segregation for horizontally separated sources was still effective although the frequency difference was absent between two sound sources. In addition, a larger separation was required to segregate the sounds as the source location was moved towards lateral positions, similar to Perrott's experiment results [34].

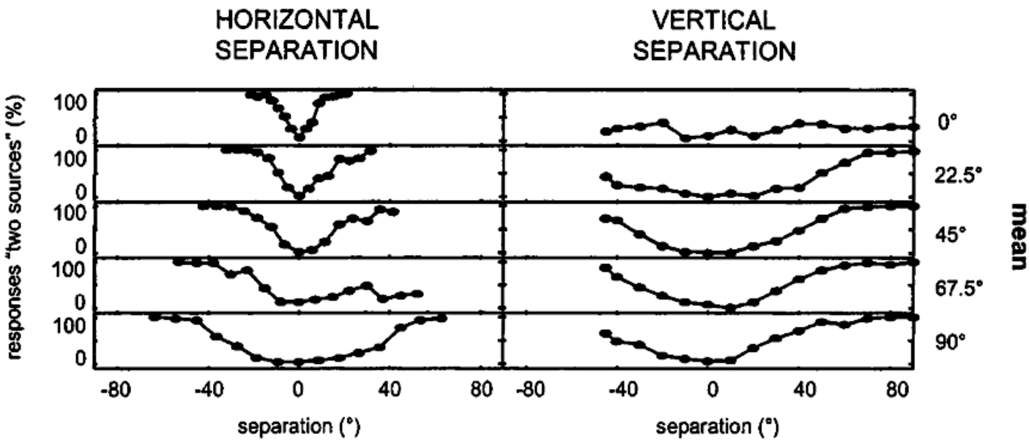


Figure 14: Mean response percentages of perceiving two source locations as a function of separation angle between two concurrently presented broadband sound sources. The left-hand column shows the data from horizontally separated sources and the right-hand column shows the data from vertical separation. Each row represents the location of a source-pair in the horizontal plane. Adapted from [11].

These studies are based on the perception of two successive or concurrent sound sources. However, only a few have researched the perception of three or more sound sources. Santala and Pulkki (2011) [36] investigated the perception of various spatial distributions of sound sources. Twenty-one different combinations of sound stimuli were presented in random order using thirteen evenly distributed sound sources covering between $-90°$ and $90°$ azimuth in the horizontal plane. Subjects were asked to identify the speakers that were emitting the sound among fifteen speakers for each stimulus combination, including two unused speakers at $-105°$ and $105°$ azimuth. The histograms showing the results for twenty-one stimulus combinations are shown in Figure 15. As shown in the figure, the results indicated that the perception of complex spatial distributions of sound sources was mostly inaccurate, especially when more than three sources were emitting sound. Cases 2-7 in the figure show that the perceived width of a sound array was narrower than the actual width of the array. On the other hand, a gap between sound sources was perceived to be wider than the actual gap widths (cases 9-12). Nevertheless, narrow gaps in wide sound stimuli were often undetectable, as can be seen in cases 8, 14, 15 and 17.

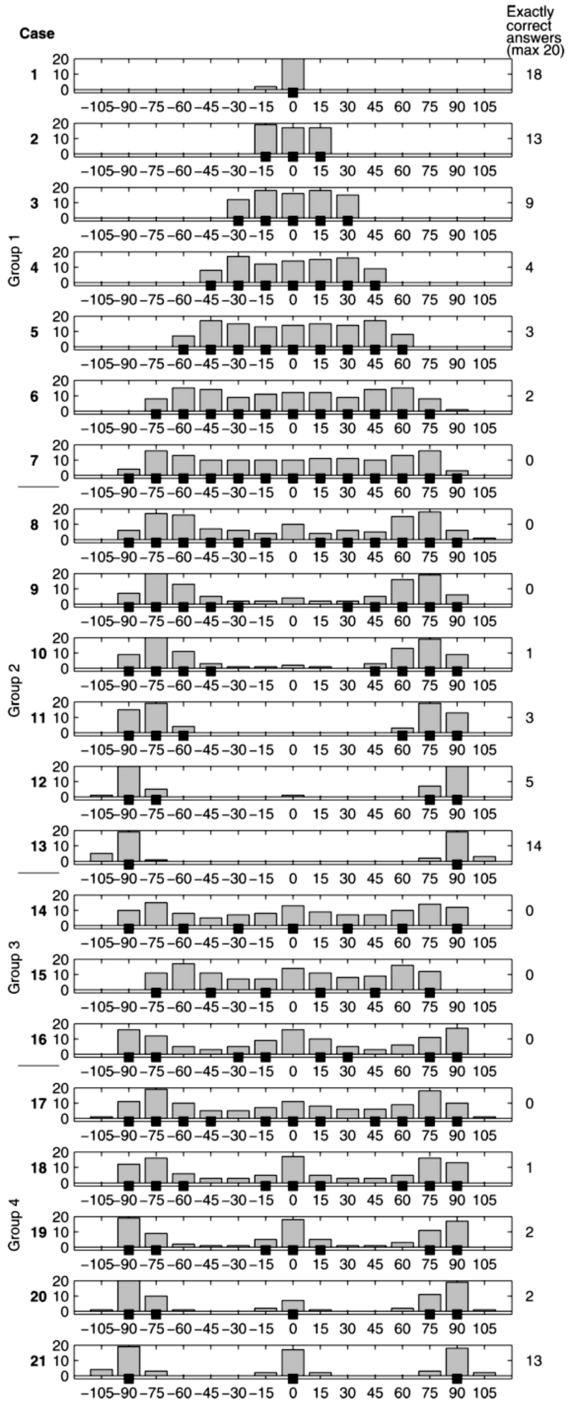


Figure 15: Perception of 21 different stimulus combinations presented as histograms. X axis represents the speaker position in azimuth. Black boxes indicate the loudspeakers emitting sound, and the height of the gray bar shows the number of loudspeakers selected by subjects as emitting sound. The number of exactly correct answers is shown on the right side of each histogram. Adapted from [36].

2.4 Perception of sounds from multiple directions in the median plane

Since this thesis studies the perception of separated sound sources in the median plane, this section is closely related to the main subject. However, only a few studies were found relevant to the subject, and those studies are reviewed below.

The MAA for the median plane was measured by Perrot and Saberi (1990) [10]. A 400Hz click train was used for sound stimuli and presented from two speakers in a vertical array that consisted of ten loudspeakers. The separation angle between the speakers was 0.46°. A sound stimulus was played from a randomly selected speaker among the array, followed by the second stimulus from another random speaker after the 500ms interval. The task for the subjects was to indicate whether the second stimulus was located above or below the first stimulus. The threshold was determined by a forced-choice adaptive procedure: three successive correct responses resulted in a 5.7cm decrease between the stimuli whereas an increase of equivalent distance was applied to an incorrect response. As a result, the MMA of 3.65° was obtained from vertically separated sound sources in this measurement. This result suggests that the sensitivity of detecting a change in the median plane localization is weaker than the MMA in the horizontal plane, according to the same measurement carried out from a horizontal array (0.97°) as well as the measurement from Mills (1958) [33].

Best et al. (2004) [11] also investigated the segregation ability for vertical pairs of broadband noise by utilizing the identical test setup used for the horizontal segregation mentioned in Section 2.3. For each of the five source locations between 0° and 90° azimuth, a pair of vertically separated sound sources was presented with different separation angles ranging between −45° and 90° in the virtual auditory space. Sound stimuli were presented as either spatially coincident or separate sources, and subjects were asked to select whether the stimuli were perceived as one or two sources, just as the tests for the horizontal segregation. As shown in the second to fifth rows of the right-hand column in Figure 14, subjects tend to easily segregate two sources when the separation angle is above about 40°, and when the sources are located at lateral positions. The horizontal location of the sound stimulus causes binaural cues to be applied in localization, and those binaural cues supplement the ability to locate vertical directions, as mentioned in section 2.2.3. In this way, the ability to segregate vertically separated sources might be empowered by binaural cues. However, the response data from the sources located in 0° azimuth, which indicate the segregation of sound sources in the median plane, might have little influences from binaural cues as shown in the first row of the right-hand column in Figure 14. Localization would mainly be dominated by spectral cues in the median plane. The results from 0° azimuth showed that the rate of perceiving "two sources" was mostly less than 50%, and no such trends in the curve were found to be significant. Best et al. suggested that spectral cues become inefficient for vertical segregation when two cues are summed at each ear.

Ferguson and Cabrera (2005) [37] examined the ability to localize two vertically separated sound sources that consisted of a high-frequency noise and low-frequency noise. A pair of noises were either synchronous or asynchronous with each other.

The location of each source was varied with five vertical positions and the spectral gap between the noises ranged between zero and six octaves. Subjects were asked to report the position of each source separately, but they often found it difficult to perceive two different locations and failed to segregate a pair of low-frequency noise and high-frequency noise. Although the frequency difference between the sources was apparent, co-location or unification occurred in 38% of the responses with synchronous sound pairs. However, the segregation ability was improved when the sound stimulus was presented with an asynchronous sound pair of a low-frequency noise and high-frequency noise.

Bremen et al.(2010) [12] carried out an experiment using two broadband sound sources which had different temporal structures. Two sources were simultaneously presented in the median plane with different separation angles that ranged between 15° and 120°, and the differences in sound levels between them ranged from -13dB to 7dB. Subjects were asked to identify the location of one specific noise among the two, and the perceived localization was affected by sound level differences and separation between the sources. The statistical result of this experiment showed two response modes under limited conditions with level differences less than 5dB and spatial disparity larger than 45°. That is, no sound source dominated the perceived location among the pairs under such conditions. This bimodal response implies that subjects could simultaneously perceive two different locations of vertically separated broadband noises when spatial disparity was sufficiently large.

2.5 Summary of the literature review

This chapter presented the basic terms and concepts of sound localization and reviewed various studies regarding two research subjects: localization in the median plane and perception of sounds from multiple directions. Median plane localization is mainly guided by spectral cues with a peak and two notches and supported by dynamic cues. The minimum audible angle for concurrent sources (CMAA) at the front is about $4-10°$ in the horizontal plane. The combination of these two subjects constituted the research topic that is most closely related to the purpose of this thesis, which is the auditory segregation in the median plane. This research subject was presented in Section 2.4.

The studies reviewed in Section 2.4 showed various experiments using two vertically separated sound sources under different conditions. However, these diverse tests regarding auditory segregation did not yield consistent results, and some of the results partly disagreed with each other. Therefore, to clarify the ability to segregate a pair of sound sources in the median plane, the experiments presented in Chapter 3 will apply limited test conditions, such as partial broadband sounds that the perception of which might be affected by pinna cues only, a free-field listening environment in a *real* acoustic space, a pair of incoherent sources with the same sound level and long-term spectrum, as well as the restriction of head movements for avoiding the support of dynamic cues. The results of these tests will be compared with those from previous studies in Chapter 5.

3 Methods

The purpose of this experiment was to determine the ability of listeners to segregate two sound sources, which were concurrently emitted from different directions in the median plane. The directions and separation angles of two concurrent sound sources were varied by different combinations. While conducting the listening test, stimuli with a single-sound source were also included along with two concurrent sound sources in order to prevent the test results from becoming biased towards the responses that would indicate only two separate sources. Consequently, this stimuli setup also enabled the thesis to investigate listeners' ability to discriminate between two concurrently played sound sources and a single-sound source.

Each test consisted of two sessions: *a monaural session* and *a binaural session*. The monaural session was carried out to determine listeners' segregation ability that would exclude any binaural cues. In addition, subjects were not allowed to move their head during both sessions in order to avoid dynamic cues as well.

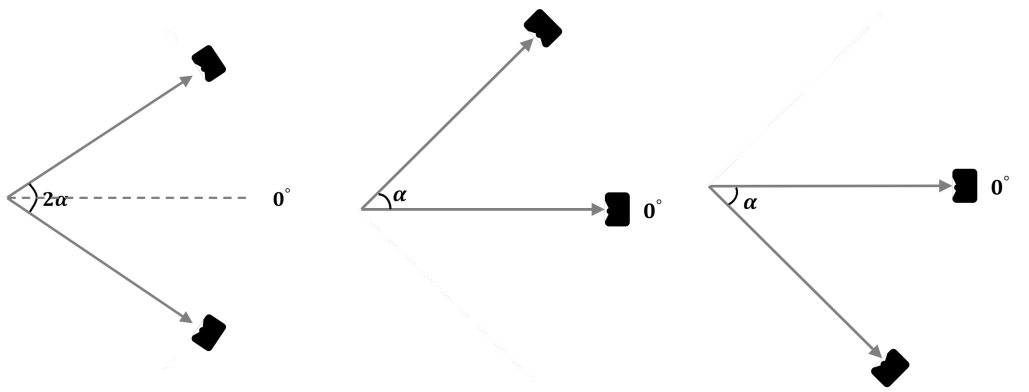
3.1 Test subjects

Fourteen subjects with normal hearing participated in the listening test. Subjects included twelve males and two females aged between 24 and 44 years. All the participants were staff members or students from *Aalto University*: ten from *The Department of Signal Processing and Acoustics* and four from *The Department of Media*. Since every participant did not have information about the purpose of this experiment, brief instructions were given to each subject before the listening test.

3.2 Sound stimuli

The sound stimuli used for the listening test comprised four configurations: three configurations of pairwise sound stimuli and a configuration of single-source stimuli. The three pairwise stimuli configurations consisted of symmetric pairs, upper+center direction pairs and lower+center direction pairs. Symmetric pairs (SYM) were configured with two sound sources symmetrically distributed above and below 0° elevation. Upper+center pairs (UPC) had one sound source at 0° elevation angle and the other from above the 0° elevation angle (the "upper" source). Lower+center pairs (LOC) had, similar to the UPC, a center source (0° elevation) and a "lower" source with a negative elevation angle (see Figure 16). The elevation angles in SYM were ±11.25°, ±22.5°, ±33.75° and ±45°. Upper sources and lower sources in UPC and LOC consisted of sources with 11.25°, 22.5°, 33.75°, 45° and −11.25°, −22.5°, −33.75°, −45° elevation angles, respectively. Accordingly, the number of sound stimuli totalled 21, including 4 stimuli for each of the three pairwise stimuli configurations and 9 single-source stimuli applied to all the sound directions set up in this experiment.

The sound signal utilized for the stimuli was a high-pass filtered pink noise. Low frequency elements were excluded in order to focus on the pinna cues that function for median plane localization. Continuous pink noise signals were generated and



(a) symmetric pair (SYM) (b) upper+center (UPC) (c) lower+center (LOC)

Figure 16: Stimuli configuration for sound source pairs

filtered using MAX7 software, and 100ms smooth ramps were applied at onsets and offsets of the signals. For high-pass filtering, each signal was passed through a bi-quad digital filter with a cut-off frequency of 3kHz. The sound pressure level of each single-source stimulus was measured from the listening position and aligned to 70dB SPL, A weighted. Stimuli with sound pairs consisted of two incoherent sound signals, and the amplitude of each signal was scaled by a factor of $1/\sqrt{2}$, resulting in the summed sound pressure level having the same value as the that of single-source stimuli. Consequently, the overall sound pressure level was identical for all the sound stimuli. In the listening test, however, each sound stimulus was presented using varying sound pressure levels. The level was altered by factors of -3dB, 0dB and 3dB in random order between trials to avoid possible bias that can be caused by stimuli repetition.

3.3 Apparatus and experimental setup

The listening test was carried out in an anechoic chamber equipped with multiple loudspeakers. Nine loudspeakers were mounted on a vertical arc in the front-center of the chamber. Sound sources were equidistantly located from the listening position, and equally distributed with the distribution interval of 11.25° covering elevation angles between $-45°$ and $45°$ (see Figure 17). All the loudspeakers were the same active monitors with flat frequency responses between 58Hz - 20kHz (±2dB).

The listening spot was located in the center of the anechoic chamber. At the listening position, a chair was tightly fixed to the floor, facing toward the sound sources. In addition, a headrest was attached to the chair in order to let the subjects stabilize their head and body. Since head movements were not allowed during the listening test, a head tracking device was utilized to monitor the head movements of each subject. For each test session, the pitch, yaw, and roll angles of head orientations were recorded every 10 milliseconds, and the detected movement angles were below 1° for all the subjects based on the record.

In order to provide monaural listening conditions in the monaural sessions, ear protectors were used to insulate one ear of each of the subjects. Every participant

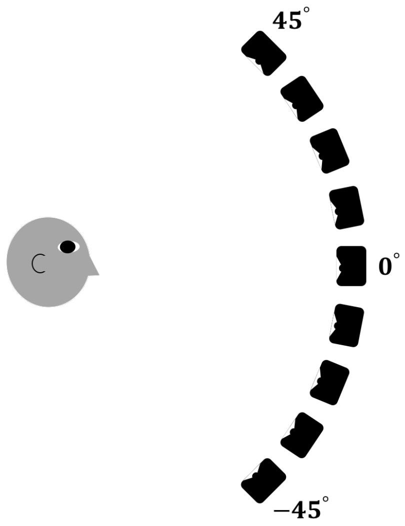


Figure 17: The arrangement of sound sources. All sound sources are equidistantly located from the listener's position and equally distributed with distribution interval of 11.25°. The elevation angles of 9 sound directions are 45°, 33.75°, 22.5°, 11.25°, 0°, −11.25°, −22.5°, −33.75° and −45° respectively.

wore both an earplug and a single-sided earmuff on the left ear while carrying out the monaural session.

In order to let the subjects control the stimuli and answer the questions, a remote control device was utilized during the tests. A tablet PC was laid on a fixed stand next to the chair, which could mutually communicate with the computer through a WI-FI connection.

3.4 Test Procedure

During the listening test, subjects controlled the stimuli for themselves by starting, stopping and proceeding to the next stimulus. This self-control was conducted using the user interface installed in the remote device. For each stimulus, subjects were requested to answer the question whether the given stimulus is heard from one direction or two directions.

Two sessions were carried out for each listening test, one was *a binaural session* and the other was *a monaural session*, as mentioned earlier. The order of the sessions was evenly distributed to the subjects in order to avoid bias resulting from experience in the previous test session. For the 14 subjects, 7 subjects were tested in the binaural session first and the monaural session later, and the remaining subjects were tested conversely. The stimuli used for both sessions were identical regardless of the usage of ear protectors. Since every stimulus was repeated seven times for each session, a total of 147 stimuli presentations (7×21 stimuli) were played in random order to each subject for each session.

Table 1: Normality test results

	Kolmogorov-Smirnov			Shapiro-Wilk		
Configuration	Statistic	df	Sig.	Statistic	df	Sig.
SYM (monaural)	.165	56	.001	.929	56	.003
UPC (monaural)	.132	56	.016	.948	56	.017
LOC (monaural)	.144	56	.005	.945	56	.013
Single (monaural)	.215	126	.000	.871	126	.000
SYM (binaural)	.156	56	.002	.919	56	.001
UPC (binaural)	.150	56	.003	.932	56	.004
LOC (binaural)	.139	56	.009	.936	56	.005
Single (binaural)	.315	126	.000	.708	126	.000

4 Results

4.1 Data analysis

The test data analyzed for the experiment corresponded to the percentage of answering "two directions" among seven answers per each stimulus. These data were used for both pairwise stimuli and single-source stimuli. Accordingly, the data for pairwise stimuli indicated frequencies of correct answers from the subjects, whereas the data for single-source stimuli showed frequencies of *wrong* answers, i.e. false alarms.

Before the statistical analysis, normality tests were conducted for each stimulus configuration. Table 1 shows the results from *the Kolmogorov-Smirnov test* and *the Shapiro-Wilk test*. Since the null-hypothesis of both tests assumes the normal distribution of the data, none of the stimuli configurations was confirmed to be normally distributed at the significance level of 0.05 ($\alpha = 0.05$).

Therefore, a number of non-parametric tests were carried out for statistical analyses of all the test data. *The Spearman's rank-order correlation* method was used to analyze the correlation between the test data and the separation angles of each stimulus configuration. *The Friedman test* was used to examine differences between a pairwise stimulus and two corresponding single-source stimuli for each pairwise stimulus configuration. When the result from the Friedman test was statistically significant, *the Wilcoxon signed rank test* for pairwise comparison was carried out for post-hoc tests. Since three comparisons were conducted in post hoc tests, the significance level of 0.05 was adjusted to 0.017 ($0.05 \div 3$) by *the Bonferroni correction*. The Wilcoxon signed rank test was also applied for comparisons between test results from the monaural session and the binaural session.

4.2 Overall trend

Figure 18 shows the response data from both the monaural and binaural sessions plotted as a function of separation angles. The plots describe the response percentages for selecting "two directions" in three pairwise stimuli configurations. As shown in

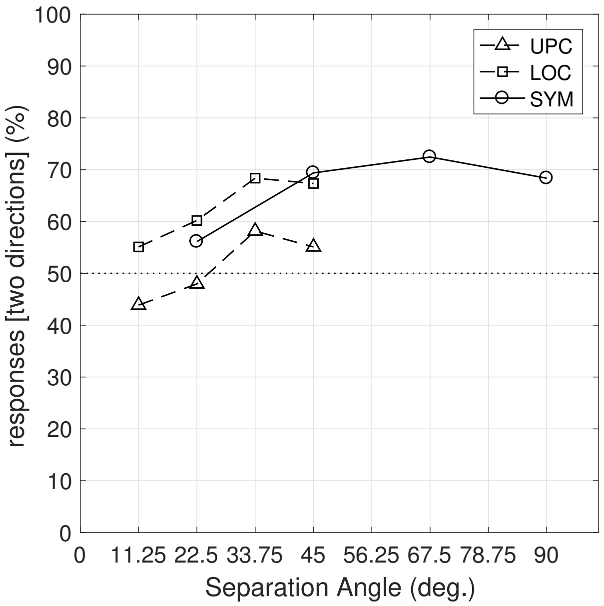
Figure 18a, the monaural session had response data ranging between 43.9% and 72.5%, exceeding 50% above a 33.75° separation angle and reaching the maximum at 67.5°. The correlation between the separation angles and the response data tended to be very weak, as shown in Table 2. The binaural session provided response data ranging between 26.5% and 76.5%, exceeding 50% above a 33.75° separation angle and reached the maximum at 67.5° (see Figure 18b). When excluding the minimum separation angle (11.25°), however, the data ranged between 48% and 76.5% at the separation angles between 22.5° and 90°, showing the range of response percentages similar to that of the monaural session. The correlation coefficients between the response data and the separation angles in the binaural session were relatively weak, though they were higher than the coefficients from the monaural session. However, the figure shows a steep rise between 11.25° and 22.5° followed by a gentle slope above 22.5°. When the correlation coefficients for the binaural session were measured with the separation angles above 22.5°, which excludes 11.25°, the coefficients were 0.114 and 0.070 at UPC and LOC, respectively. These correlation coefficients are much weaker than those found when including an 11.25° separation angle.

Table 2: *Spearman's Correlation Coefficient* for the correlation between the separation angles and the response data.

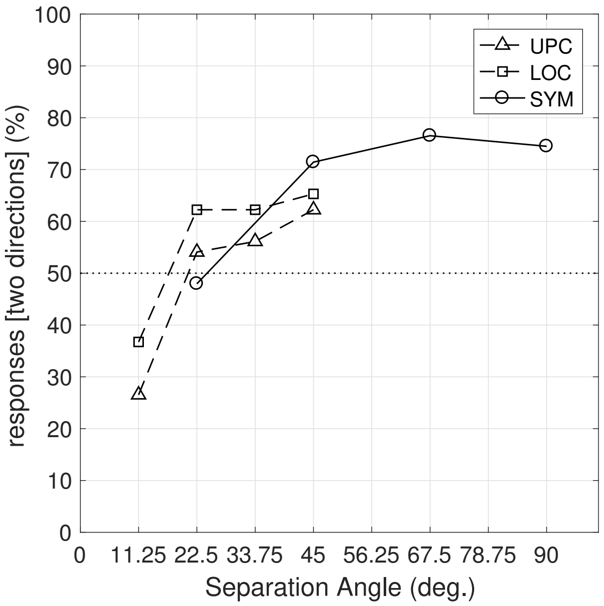
Configuration	Monaural Hearing	Binaural Hearing
SYM	.158	.361
UPC	.166	.408
LOC	.167	.341

Table 3: The average and standard deviations of response duration for subjects to answer each question for each take (sec).

	Monaural Hearing		Binaural Hearing	
Configuration	Mean	Std.	Mean	Std.
SYM	15.6	9.21	12.1	7.2
UPC	9.5	7.4	8.5	4.7
LOC	10.4	10.2	7.2	3.4
Single-source	14.2	11.9	10.3	5.5



(a) Monaural Hearing



(b) Binaural Hearing

Figure 18: The response data of answering "two directions" in terms of separation angle. The upper plot shows the results from the monaural session and the lower plot from the binaural session.

When comparing the response data from UPC with LOC, LOC showed higher mean percentages of responding "two directions" than UPC for all the separation angles in both sessions. Statistic comparisons between the data from these two configurations were carried out with the Wilcoxon signed rank test. The statistic difference showed conflicting significance values between the monaural session and the binaural session. The difference was statistically significant in the monaural session ($p = 0.02$), whereas it appeared to be insignificant ($p = 0.11$) in the binaural session at the significance level of 0.05.

The response duration spent for subjects to answer each question was recorded during the tests. Although the duration data varied considerably by subjects and configurations, the overall tendency can be observed in the average duration data. As can be seen in Table 3, subjects tend to spend more time in answering during the monaural session than during the binaural session. Comparison of each stimuli configuration to one another revealed that the result could be attributed to the longer period of answering time with symmetrical pairs than with both upper/lower+center pairs.

4.3 Comparison between the perception of sound pairs and corresponding single sources

To determine the ability of listeners to discriminate a pairwise stimulus from each of two single-source stimuli, statistical comparisons were carried out between the response data from pairwise stimuli and two single-sources corresponding to the pairs.

Figure 19 shows the response data from each pairwise stimuli configuration for each session along with the response data of single sources that correspond to each configuration. Since the response data from single sources consisted of the responses that answered "two directions", they present the response percentages of *incorrect* answers. As a result, this statistical analysis was carried out by comparing between "correct answers" in response to pairwise stimuli and "incorrect answers" arising from each of two single-source stimuli.

The results of these statistical comparisons are given in Table 4. For all comparisons between pairwise stimuli and corresponding single-source stimuli (the Friedman test), the difference in every stimuli configuration was statistically significant at $p < 0.05$. Post-hoc tests were also conducted between pairwise stimuli and each corresponding single-source stimulus (the Wilcoxon signed rank test). Comparison between two single sources was not taken into account. The difference in every pairwise comparisons for each stimuli configuration was statistically significant at $p < 0.017$ when applying the Bonferroni correction.

Table 4: Statistical differences between pairwise stimuli and corresponding single-source stimuli. Friedman test was conducted to compare a pairwise stimulus and two corresponding single-source stimuli, and Wilcoxon signed rank test was conducted for post-hoc pairwise comparisons.

	Friedman test		Wilcoxon signed rank test (Sig.)		
Configuration	χ^2	Sig.	pair:upper	pair:center	pair:lower
SYM (monaural)	64.663	< 0.001	< 0.001	n/a	< 0.001
UPC (monaural)	42.378	< 0.001	< 0.001	< 0.001	n/a
LOC (monaural)	56.708	< 0.001	n/a	< 0.001	< 0.001
SYM (binaural)	84.910	< 0.001	< 0.001	n/a	< 0.001
UPC (binaural)	59.879	< 0.001	< 0.001	< 0.001	n/a
LOC (binaural)	60.694	< 0.001	n/a	< 0.001	< 0.001

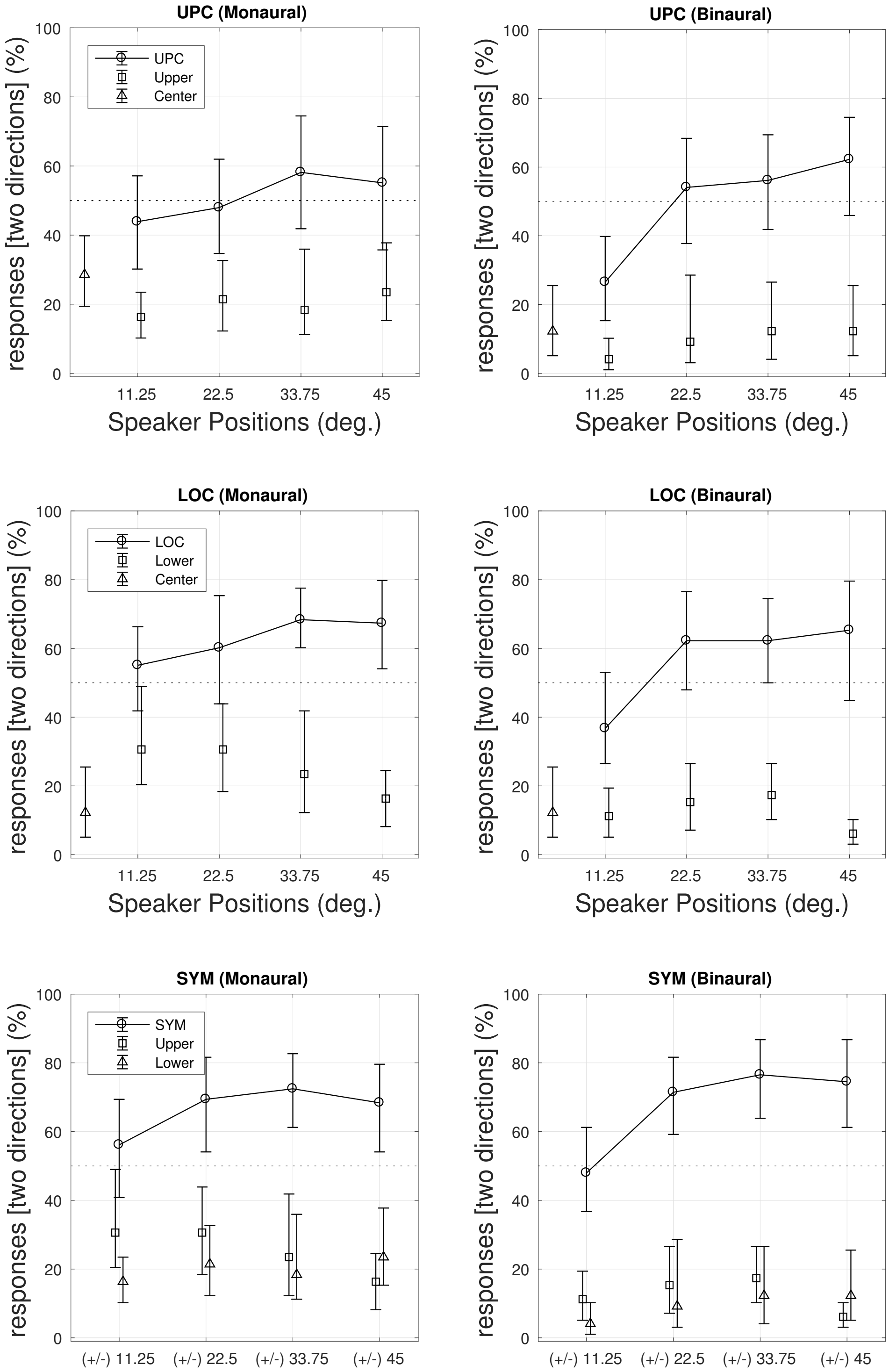


Figure 19: The response data of each pairwise stimuli configuration and two corresponding single-source stimuli. In UPC and LOC graphs, center speaker stimulus is located in 0° elevation. x coordinates of UPC and LOC plots indicate separation angles.

4.4 Comparison between monaural hearing and binaural hearing

Comparisons between the response data from the monaural session and the binaural session are described in Figure 20. The Wilcoxon signed rank test indicated that none of the pairwise stimuli configurations showed a statistically significant difference between monaural hearing and binaural hearing at $p < 0.05$. The P-values in SYM, UPC and LOC were 0.805, 0.773 and 0.195, respectively.

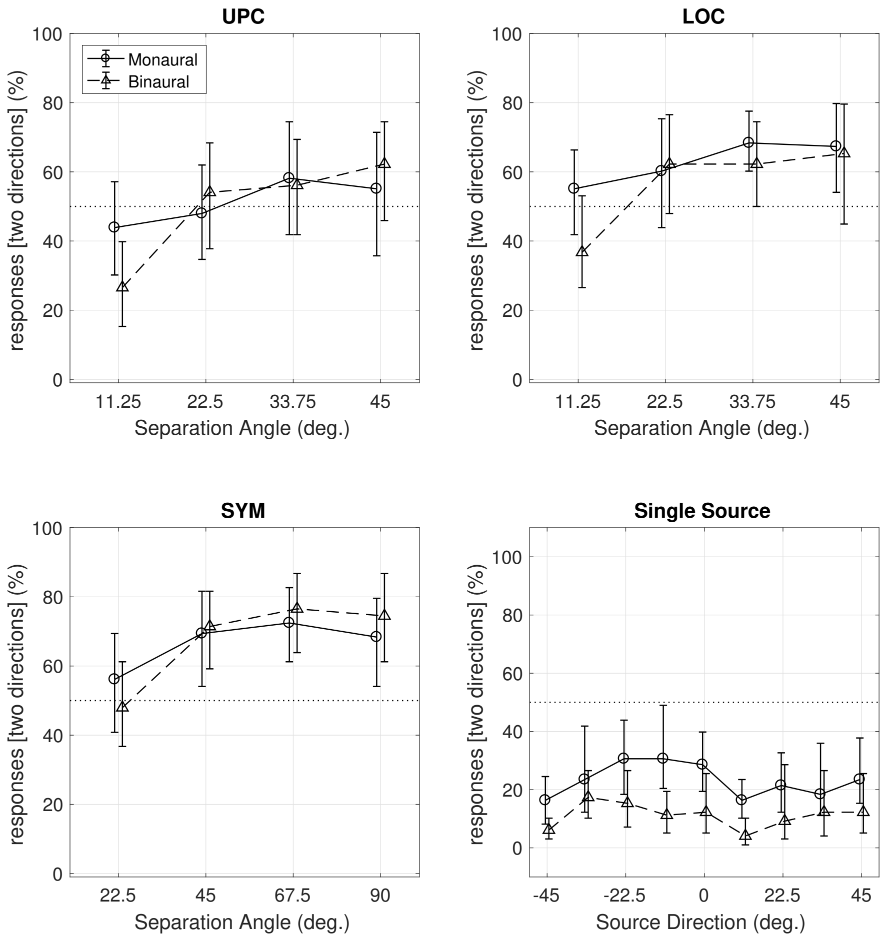


Figure 20: Comparisons between the monaural session and the binaural session.

5 Discussion

5.1 The segregation of pairwise stimuli

In this study, the response percentage indicating the segregation of two sources exceeded 50% above the separation of 33.75° for all configurations in both monaural hearing and binaural hearing. Furthermore, the response reached nearly 70% above the separation angle of 45° for symmetric pairs (SYM) in both sessions. When the criterion for measuring a CMAA by Perrott (1984) [34] is applied, which is 75% of the correct answer for pairwise stimuli, only one specific case satisfies this criterion among all the test conditions in this test: 76.5% of the correct answer at the separation of 67.5° for symmetric pairs (SYM) in binaural hearing session. However, since the sound stimuli used in Perrot's experiment were pairs of pure tones which are different in their frequencies, it is inappropriate to strictly apply that criterion to this test that used a partial broadband signal for the sound stimulus. As mentioned in Section 2.3, the auditory spatial resolution deteriorates with decreasing frequency difference between two tones, or with increasing spectral overlap between two sources [34, 35]. Therefore, it is applicable to compare the results of this test with Best et. al's test [11], since both tests used a pair of broadband noises with full spectral overlap for the sound stimuli.

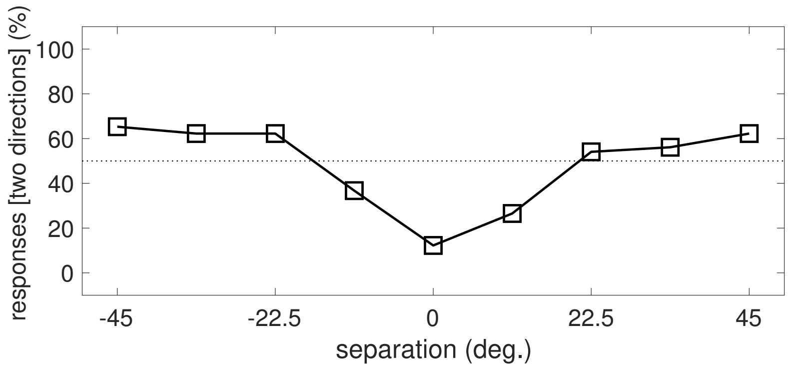


Figure 21: The combined response rate of answering "two directions" from UPC and LOC in the binaural session. The x-axis represents the elevation angle relative to the center source.

Figure 21 shows the combined response data from UPC and LOC in the binaural session. This plot is easily comparable to the plots by Best et. al (see Figure 14). When this plot is compared to the uppermost plot in the *left* column of Figure 14, which is the response data from horizontal separation at the front (0° azimuth), the perception of *vertical* separation in this test showed poorer results than those from horizontal separation. This result is predictable because the effects of binaural cues are restricted in the median plane localization. However, when compared to the uppermost plot in the *right* column of Figure 14, which shows the median plane localization that this thesis examines, the test results by Best et. al differed from the results of this experiment. The response rates were much poorer than those from this test, and no such trend could be found in the plot by Best et. al although the result of this test showed an increasing trend with respect to the separation angle, as shown in Figure 18 and Figure 21. The disagreement between the test results from this experiment and Best et. al's experiment can be explained in terms of differences in the test conditions. The biggest difference may be the listening environment because this experiment was carried out in a *real* acoustic space using loudspeakers whereas the *virtual* auditory space (VAS) was used for the experiment by Best et. al. The duration of sound stimuli might be another reason for the differences because Best et. al's experiment utilized short broadband noise *bursts* as sound stimuli whose duration was limited to 150ms, whereas unlimited stimulus duration was applied in this experiment.

It is interesting to note that the mean response rate of UPC was lower than LOC for all conditions. A possible assumption for this result might be that "upper" sources were localized more sensitively than "lower" sources. The higher localization sensitivity for an upper source compared to a center or lower source could lead to the dominant perception of the upper source, resulting in a pair of sources to be misperceived as a single-source. This assumption may be explained by the increase in the rate-of-change of notches in HRTFs (N1, N2) which function as spectral cues for the median plane localization. Figure 22 shows an example of HRTFs from the CIPIC HRTF database [29]. The notches are shown as the dark lines on both sides of the 2D plot. As can be seen in the plots, notches decrease moderately below 0° elevation, then fall down steeply above around 0°. That is, a direction from the "lower" source has a lower rate-of-change while it is higher in a direction from the "upper" source for the notches in this HRTFs, which might indicate the increasing localization sensitivity. However, this specific HRTF does not represent the overall trend of the changes in their notches. The diversity of HRTFs in every individual makes it difficult to clearly explain a psychoacoustic phenomenon by means of HRTFs.

Another possible explanation for this result is that the listening environment influenced the perception of pairwise sources. Although sounds from "upper" sources encountered little obstacles to reach one's ear(s), sounds from "lower" sources could not directly arrive at the ear(s) since they were scattered and reflected by the legs, arms and body of the subject as well as the remote device that lied next to the subject. The scattering and reflection of a sound can weaken the clarity of the sound, resulting in a biasing localization to the upper source. This assumption can also explain the decrease in the response rate at 90° for symmetric pairs (SYM) in both sessions (see Figure 18). the lowermost sound source at $-45°$ elevation would be more scattered and reflected than any other sources, which could cause a slight decrease in the response rate at 90°.

The auditory segregation of a pair of concurrent sound sources can be regarded as a simultaneous localization of two sources. From this approach, a localization from two directions can be expressed as two HRTF plots overlapping each other. Figure 23 shows examples of overlapped HRTF plots from two directions in the median plane. The coloured rectangular areas in the plots indicate the critical bandwidths around the center frequencies of the two notches from each of the two HRTFs. The critical bandwidth is derived from *the equivalent rectangular bandwidth* (ERB) value defined by Glasberg and Moore (1990) [38], of which the equation is expressed as

$$ERB = 24.7(4.37F + 1) \tag{3}$$

where F is a center frequency in kHz.

As shown in Figure 23a, the critical bandwidths of four notches do not overlap each other even for the smallest separation angle tested in this experiment (11.25°). The separation between each critical bandwidth indicates that all the notches can be distinctively perceived at the ears in this case, and the interpretation of multiple notches are not affected by the frequency resolution of the inner ear. Therefore, this plot implies that two concurrent sources might be simultaneously localized with less disruption at the separation of 11.25°. Although the ability to segregate sound sources in the median plane turns out to be relatively sensitive by these peripheral measurements, the data from this example do not fully agree with the test results of this thesis, especially for the binaural session. This difference will be discussed further in Section 5.3.

One of the interesting phenomena in this approach might be the case when two notches overlap each other. As shown in Figure 23c, the second notch (N2) in the HRTF at −33.75° overlaps the first notch (N1) in the HRTF at 33.75°. In this case, the segregation might be influenced by the depths of the notches. The localization can be biased to the direction of the source showing the deeper notch that dominates the other. However, this plot shows the depths of the overlapping notches similar to each other, which could detriment the localization accuracy for both directions. Since the elevation angle where two notches overlap may vary by every individual, this phenomenon did not appear from the mean response rates in the test result.

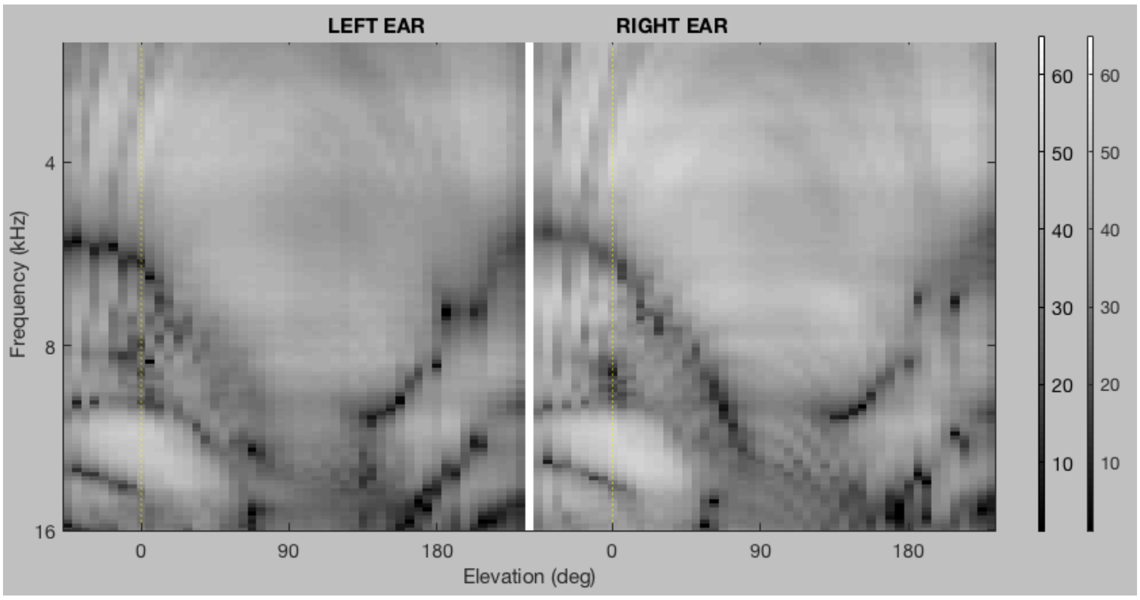


Figure 22: An example of 2D HRTF plots showing the responses at both ears as functions of elevation and frequency. Adapted from the HRTF data of the subject "50" in the CIPIC HRTF database [29].

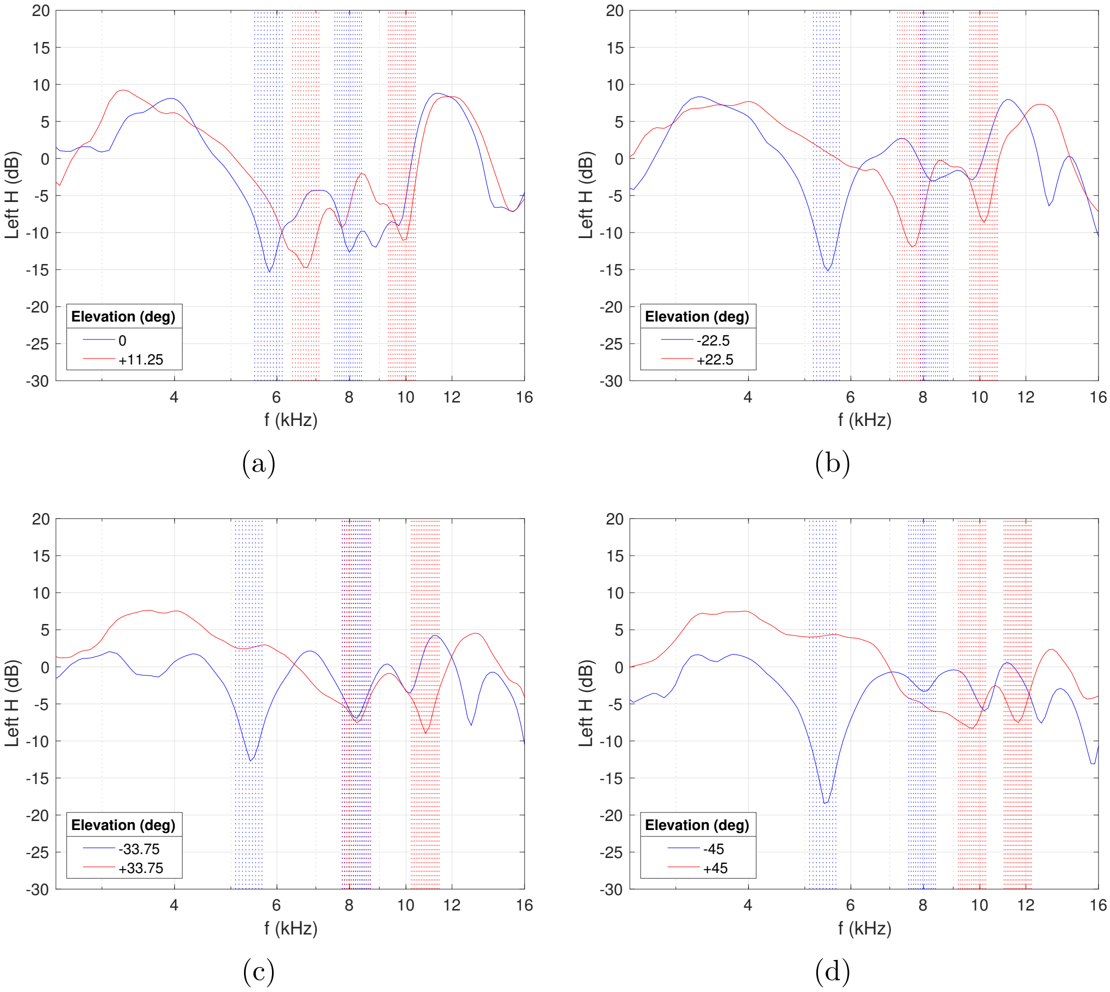


Figure 23: Examples of overlapped HRTF plots from two directions in the median plane. The coloured rectangular areas in the plots indicate the critical bandwidths around the center frequencies of the two notches from each of the two HRTFs.(a) shows HRTFs from 0° and +11.25° elevation, (b) shows HRTFs from −22.5° and +22.5° elevation, (c) shows HRTFs from −33.75° and +33.75° elevation and (d) shows HRTFs from −45° and +45° elevation. Adapted from the HRTF data of the subject "50" in the CIPIC HRTF database [29].

5.2 Discrimination between a pairwise stimulus and corresponding single sources

Table 4 and Figure 19 indicate the comparison between the rate of correct answers from a pairwise stimulus and incorrect answers from each single-source which corresponds with the pairwise stimulus. This comparison determines the validity of correct responses from pairwise stimuli by identifying the significance of difference between the response rates from the two cases. In other words, it examined whether the responses of "two directions" for pairwise stimuli showed the actual segregation of two sources, or the subjects tended to just answer for "two directions" in all the cases. As shown in Table 4, the test result showed statistically significant differences for all the comparisons, which indicate the validity of the response rates in this experiment. However, these results must be interpreted with caution because they cannot confirm the assumption that subjects actually heard sounds from two different *directions*. This comparison result rather confirms that subjects could distinguish between a pair of sounds and a sound from a single-source whether they detected the difference in the number of directions or just different timbre from each of the sound sets. It is possible to assume that subjects could select "two directions" because a pairwise stimulus sounded different from a single-source stimulus although they could not segregate two sound sources. The summation of sounds from two different directions will create different spectral information compared to the single-source at the pinna. Therefore, the argument whether the perception of this spectral information can lead to the perception of two different directions or a different timbre remains an issue for further research.

5.3 The influence of the binaural hearing on the segregation in the median plane

The statistical comparison between the monaural session and the binaural session found no significant difference for all the stimulus configurations (see Section 4.3). However, it is interesting to note a few findings concerning the differences of binaural hearing compared to monaural hearing in this experiment.

One interesting finding is that the mean response rates for the smallest separation angle (11.25°) for both UPC and LOC were much lower in the binaural session than the monaural session. The response rates in the binaural session were 26.5% and 36.7% whereas the rates in the monaural session were 43.9% and 55.1% for UPC and LOC, respectively. Since the rate-of-change between 22.5° and 90° in both sessions were similar to each other (see Section 4.1), these low rates for 11.25° separation in the binaural session are found to be particular among other response rates. Moreover, these results disagree with the assumption regarding the overlapped HRTF plot shown in Figure 23a, which implies that the perceived auditory event is affected by something other than peripheral effects (see Section 5.1). However, they seem to be consistent with the research by Damaske (1968) [32]. A test result from Damaske showed that the auditory event from two closely located sound sources in the horizontal plane were fused into a single auditory space, as mentioned in

Section 2.3. Since Damaske used two incoherent broadband noises for sound stimuli and tested with binaural hearing, which is similar to the experiment in this thesis, the result from Damaske can be regarded as a reference for the reason why this phenomenon appeared in the test. The experiment by Best et. al (2004) [11], which used incoherent broadband sources and binaural hearing as well, also supports this phenomenon. As shown in Figure 14, the response rates were remarkably low at lower separation angles. The ability to segregate two concurrent sound sources was weakened from closely located sources, similar to the experiment by Damaske. Although the results of these studies were obtained from sound sources located in the horizontal plane, these findings may help to understand some results from this thesis. This phenomenon may be concerned as an effect from the central auditory system than a peripheral effect.

In addition to the lower response rates for the smallest separation angle, higher rates were found for wider separation angles above 45° in the binaural session compared to the rates in the monaural session. That is, the ability to segregate two sound sources with binaural hearing was weaker for the closely located sources and stronger for the widely located sources in comparison with the monaural hearing. This trend can be confirmed by the correlation (a.k.a. the rate-of-change) between the response rate and the separation angle. As shown in Table 2, the correlation coefficient for binaural conditions was much higher than monaural conditions for all the stimuli configurations. Concerning the limited allowance of head movements applied in this experiment, this increase of correlation in the binaural hearing compared to the monaural hearing can hardly be explained as an influence from binaural cues as well as dynamic cues, but it might be more affected by central processing.

Another interesting finding is the perception of single-source stimuli. As shown in Figure 20, the rates of answering the incorrect responses were much lower in the binaural session than the monaural session, meaning that single-source stimuli were more correctly responded in the binaural session than the monaural session. When these data from both sessions are statistically compared with each other, the difference between the response rates of each session was highly significant ($p < 0.001$, from the Wilcoxon signed rank test). In other words, a single sound source might be more clearly heard with both ears than with only one ear, even when subjects were not allowed to move their heads. This finding is in agreement with the experiment results by Butler et. al (1990) [39] in which the subjects localized the sound source in the median plane more accurately with binaural hearing than with monaural hearing. Supposing the fact that the head movement was strictly restricted and thus almost no influence from binaural cues could be present in this experiment, it is interesting to observe that a summation of two nearly identical sounds by central processing can provide clarity to the sound perception.

6 Conclusion

This study examined the ability of human listeners to segregate two concurrently presented sound sources in the median plane. It was shown in the results that a pair of vertically separated sound sources can be moderately segregated when the separation between the sources is sufficiently wide. The segregation ability in the monaural hearing condition showed no significant difference from the binaural hearing condition. The rate-of-change for the segregation was relatively weak with respect to the degree of separation, although it increased in the binaural hearing condition. In addition, it was not difficult for the listeners to discriminate a pair of sound sources from each of the corresponding two single sources.

6.1 Further study

Further work needs to be carried out to clarify a correct response of pairwise stimuli to be the perception of two different directions. It might be done using several broadband sound signals with the same long-term spectrum but applied with different temporal structures. A random presentation of these stimuli would enable the subjects to discriminate the perception of two different *directions* from other possible perceivable factors in the experiment.

References

[1] Brian C J Moore. An Introduction to the Psychology of Hearing. *Boston Academic Press*, 3:413, 2003.

[2] John C. Middlebrooks and David M. Green. Sound Localization by Human Listeners. *Annual Review of Psychology*, 42(1):135–159, 1991.

[3] D.W Grantham. Spatial hearing and related phenomena. In *Hearing*, chapter 9, pages 297–345. Academic Press, 1995.

[4] J Blauert. *Spatial Hearing: The Psychophysics of Human Sound Localization*. MIT Press, 1997.

[5] A William Mills. Auditory localization(binaural acoustic field sampling, head movement and echo effect in auditory localization of sound sources position, distance and orientation). *Foundations of modern auditory theory.*, 2:303–348, 1972.

[6] Ville Pulkki and Matti Karjalainen. *Communication Acoustics: An introduction to speech, audio and psychoacustics*. Wiley, 2015.

[7] V. R. Algazi, R. O. Duda, D. M. Thompson, and C. Avendano. The cipic hrtf database. In *Proceedings of the 2001 IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics (Cat. No.01TH8575)*, pages 99–102, 2001.

[8] Hironori Takemoto, Parham Mokhtari, Hiroaki Kato, Ryouichi Nishimura, and Kazuhiro Iida. Mechanism for generating peaks and notches of head-related transfer functions in the median plane. *The Journal of the Acoustical Society of America*, 132(6):3832–41, 2012.

[9] Kazuhiro Iida, Yohji Ishii, and Shinsuke Nishioka. Personalization of head-related transfer functions in the median plane based on the anthropometry of the listener's pinnae. *The Journal of the Acoustical Society of America*, 136(1):317–33, jul 2014.

[10] David R. Perrott and Kourosh Saberi. Minimum audible angle thresholds for sources varying in both elevation and azimuth. *The Journal of the Acoustical Society of America*, 87(4):1728–1731, 1990.

[11] Virginia Best, André van Schaik, and Simon Carlile. Separation of concurrent broadband sound sources by human listeners. *The Journal of the Acoustical Society of America*, 115(1):324–336, 2004.

[12] Peter Bremen, Marc M. van Wanrooij, and A. John van Opstal. Pinna cues determine orienting response modes to synchronous sounds in elevation. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 30(1):194–204, jan 2010.

[13] Robert A. Butler and Krystyna Belendiuk. Spectral cues utilized in the localization of sound in the median sagittal plane. *The Journal of the Acoustical Society of America*, 61(5):1264–1269, 1977.

[14] Kazuhiro Iida, Motokuni Itoh, Atsue Itagaki, and Masayuki Morimoto. Median plane localization using a parametric model of the head-related transfer function based on spectral cues. *Applied Acoustics*, 68(8):835–850, aug 2007.

[15] P M Hofman, J G Van Riswick, and A J Van Opstal. Relearning sound localization with new ears. *Nature neuroscience*, 1(5):417–21, 1998.

[16] Elizabeth M. Wenzel, Marianne Arruda, Doris J. Kistler, and Frederic L. Wightman. Localization using nonindividualized head-related transfer functions. *The Journal of the Acoustical Society of America*, 94(1):111–123, jul 1993.

[17] Ira J. Hirsh. Masking of Speech and Auditory Localization. *International Journal of Audiology*, 10(2):110–114, jan 1971.

[18] H Wallach. The role of head movements and vestibular and visual cues in sound localization. *Journal of Experimental Psychology*, 27(4):339–368, 1940.

[19] Simon Carlile and Johahn Leung. The Perception of Auditory Motion. *Trends in Hearing*, 20:1–19, 2016.

[20] Frederic L. Wightman and Doris J. Kistler. Resolution of front–back ambiguity in spatial hearing by listener and source movement. *http://dx.doi.org/10.1121/1.426899*, apr 1999.

[21] EA Macpherson. Head motion, spectral cues, and wallach's 'principle of least displacement' in sound localization. *Principles and applications of spatial hearing*, pages 103–120, 2011.

[22] S. Mehrgardt and V. Mellert. Transformation characteristics of the external human ear. *The Journal of the Acoustical Society of America*, 61(6):1567–1576, jun 1977.

[23] Jack Hebrank and D. Wright. Spectral cues used in the localization of sound sources on the median plane. *The Journal of the Acoustical Society of America*, 56(6):1829–1834, dec 1974.

[24] E. A. G. Shaw and R. Teranishi. Sound Pressure Generated in an External-Ear Replica and Real Human Ears by a Nearby Point Source. *The Journal of the Acoustical Society of America*, 44(1):240–249, jul 1968.

[25] P. Jeffrey Bloom. Creating Source Elevation Illusions by Spectral Manipulation. *Journal of the Audio Engineering Society*, 25(9):560–565, sep 1977.

[26] Robert A. Butler and R a Humanski. Localization of sound in the vertical plane with and without high-frequency spectral cues. *Perception & psychophysics*, 51(2):182–186, 1992.

[27] Brian C. J. Moore, Simon R. Oldfield, and Gary J. Dooley. Detection and discrimination of spectral peaks and notches at 1 and 8 kHz. *The Journal of the Acoustical Society of America*, 85(2):820–836, feb 1989.

[28] Vikas C. Raykar, Ramani Duraiswami, and B. Yegnanarayana. Extracting the frequencies of the pinna spectral notches in measured head related impulse responses. *The Journal of the Acoustical Society of America*, 118(1):364–374, jul 2005.

[29] V.R. Algazi, R.O. Duda, D.M. Thompson, and C Avendano. The CIPIC HRTF database. In *Proceedings of the 2001 IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics (Cat. No.01TH8575)*, number October, pages 99–102. IEEE, 2001.

[30] Stephen Perrett and William Noble. The effect of head rotations on vertical plane sound localization. *The Journal of the Acoustical Society of America*, 102(4):2325–2332, 1997.

[31] RI Chernyak. Pattern of the noise images and the binaural summation of loudness for the different interaural correlation of noise. In *Proc. 6th Int. Congr. Acoust.*, volume 3, 1968.

[32] Peter Damaske. Subjektive untersuchung von schallfeldern (subjective investigation of sound fields). *Acustica*, 19:199–213, 1968.

[33] A. W. Mills. On the Minimum Audible Angle. *The Journal of the Acoustical Society of America*, 30(4):237–246, apr 1958.

[34] David R. Perrott. Concurrent minimum audible angle: A re-examination of the concept of auditory spatial acuity. *The Journal of the Acoustical Society of America*, 75(4):1201–1206, apr 1984.

[35] Pierre L. Divenyi and Susan K. Oliver. Resolution of steady-state sounds in simulated auditory space. *The Journal of the Acoustical Society of America*, 85(5):2042–2052, may 1989.

[36] Olli Santala and Ville Pulkki. Directional perception of distributed sound sources. *The Journal of the Acoustical Society of America*, 129(3):1522–1530, 2011.

[37] Sam Ferguson and Densil Cabrera. Vertical localization of sound from multiway loudspeakers. *AES: Journal of the Audio Engineering Society*, 53(3):163–173, 2005.

[38] Brian R. Glasberg and Brian C J Moore. Derivation of auditory filter shapes from notched-noise data. *Hearing Research*, 47(1-2):103–138, aug 1990.

[39] Robert A Butler, Richard A Humanski, and Alan D Musicant. Binaural and Monaural Localization of Sound in Two-Dimensional Space. *Perception*, 19(2):241–256, apr 1990.