HELSINKI UNIVERSITY OF TECHNOLOGY
Faculty of Electronics, Communications and Automation
Department of Signal Processing and Acoustics

Olli Santala

Perception of Spatially Distributed Sound Sources

Master's Thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in Technology.

Espoo, May 22, 2009

Supervisor: Matti Karjalainen
Instructor: Ville Pulkki

HELSINKI UNIVERSITY OF TECHNOLOGY

ABSTRACT OF THE MASTER'S THESIS

Author:	Olli Santala
Name of the thesis:	Perception of Spatially Distributed Sound Sources
Date:	May 22, 2009 **Number of pages:** 72
Faculty:	Electronics, Communications and Automation
Professorship:	S-89
Supervisor:	Prof. Matti Karjalainen
Instructor:	Docent Ville Pulkki

Sound localization studies have mostly been concentrating on the localization of a single source. Nevertheless, there are studies on the perception of several simultaneous sound sources in spatial conditions. A large number of those experiments have been done using headphones, but also loudspeakers have been used. It has been found out that spatial width perception is affected for example by signal loudness, frequency and temporal length.

In this thesis, perception of spatial sound was investigated by conducting two listening tests. The focus was on the resolution of directional perception of spatially distributed sound sources. The tests were performed in an anechoic chamber using 15 loudspeakers that were placed in the horizontal plane equidistant from the listener.

In the first listening test, various sound source distributions such as sound sources with varying widths and wide sound sources with gaps in the distribution were used. The subjects were asked to distinguish which loudspeakers emit sound according to their own perception. Results show that small gaps in the sound source were not perceived accurately and wide sound sources were perceived narrower than they actually were. The results also indicate that the resolution for fine spatial details is worse than 15 degrees when the sound source is wide.

In the second listening test, noise signals with different bandwidths as well as sine waves divided to the loudspeakers were used as stimuli. These were presented to the subjects using loudspeaker combinations with different loudspeaker densities. Two loudspeaker combinations at a time were presented and the task of the subjects was to discriminate which of the two shown combinations was used in producing the latter of the two sound events. The results indicate that the perception accuracy decreased as the loudspeaker density increased. Also, the bandwidth of the noise signals affected the perception accuracy.

Keywords: Spatial sound, Psychoacoustics, Perception, Listening tests

TEKNILLINEN KORKEAKOULU DIPLOMITYÖN TIIVISTELMÄ

Tekijä:	Olli Santala
Työn nimi:	Äänilähteen tilajakauman havaitseminen
Päivämäärä:	22.5.2009 **Sivuja:** 72
Tiedekunta:	Elektroniikka, tietoliikenne ja automaatio
Professuuri:	S-89
Työn valvoja:	Prof. Matti Karjalainen
Työn ohjaaja:	Dos. Ville Pulkki

Äänen suunnan havaitsemisen tutkimukset ovat paljolti keskittyneet yhden äänilähteen tapaukseen. Useamman samanaikaisen tilassa sijaitsevan äänilähteen havaitsemisesta on kuitenkin myös tutkimuksia. Suuri osa noista tutkimuksista on tehty kuulokkeilla, mutta kaiuttimiakin on käytetty. On havaittu, että äänen leveyden havaitsemiseen vaikuttavat esimerkiksi äänenvoimakkuus, taajuus ja ajallinen pituus.

Tässä diplomityössä tilaäänen havaitsemista tutkittiin tekemällä kaksi kuuntelukoetta. Painopiste oli tilassa hajautetusti sijaitsevien äänilähteiden suuntien havaitsemisen tarkkuudessa. Kokeet suoritettiin kaiuttomassa huoneessa, jossa 15 kaiutinta oli asetettu horisontaalitasoon, kaikki samalle etäisyydelle koehenkilöstä.

Ensimmäisessä kuuntelukokeessa käytettiin erilaisia laajalle jakautuneita äänilähdekokonaisuuksia. Mukana oli esimerkiksi yksittäinen leveä äänilähde, jonka leveys vaihteli tapauksesta toiseen sekä leveitä äänilähteitä, joiden jakaumassa oli aukkoja. Koehenkilöiden tehtävänä oli kussakin tapauksessa oman havaintonsa mukaan erottaa, mitkä kaiuttimet lähettivät ääntä. Tulosten mukaan äänilähteessä olleita pieniä aukkoja ei havaittu täsmällisesti ja leveät äänilähteet havaittiin kapeampina kuin ne oikeasti olivat. Tulokset viittaavat myös siihen, että tilajakauman yksityiskohtien havaitsemisen tarkkuus on huonompi kuin 15 astetta, kun äänilähde on leveä.

Toisessa kuuntelukokeessa testiääninä käytettiin kohinasignaaleja eri kaistanleveyksillä sekä siniaaltoja, jotka jaettiin eri kaiuttimiin. Näitä ääniä esitettiin koehenkilöille käyttämällä eri kaiutinyhdistelmiä, joiden kaiutintiheys vaihteli. Kaiutinyhdistelmiä esitettiin kaksi kerrallaan, ja koehenkilöiden tehtävä oli erottaa, kumpi kahdesta kosketusnäytöllä kuvatusta yhdistelmästä oli käytössä jälkimmäisessä testiäänessä. Tulosten mukaan havaitsemisen tarkkuus pieneni, kun kaiuttimien tiheys kasvoi. Myös kohinasignaalien kaistanleveys vaikutti havaitsemisen tarkkuuteen.

Avainsanat: Tilaääni, psykoakustiikka, havaitseminen, kuuntelukokeet

Acknowledgements

This Master's thesis and the research in it has been done in the Department of Signal Processing and Acoustics. The project was funded by the Emil Aaltonen Foundation.

I want to thank my instructor docent Ville Pulkki for all the valuable ideas and guidance throughout the whole process. I would also like to thank my supervisor professor Matti Karjalainen for the useful comments on the thesis.

My gratitude also goes to my co-workers in the department, particularly Tapani Pihlajamäki, Marko Takanen, Marko Hiipakka, Toni Hirvonen, Mikko-Ville Laitinen, Jukka Ahonen and Jan Oksanen. They all helped me in one way or another.

Finally, I would like to thank my family and friends for all the support during my studies, and special thanks go to my dear Johanna.

Otaniemi, May 22, 2009

Olli Santala

Contents

Abbreviations

ANOVA	Analysis of variance
ERB	Equivalent rectangular bandwidth
HRTF	Head-related transfer function
IACC	Interaural cross-correlation
ILD	Interaural level difference
ITD	Interaural time difference

List of Figures

List of Tables

Chapter 1

Introduction

Humans are able to perceive the direction of sound whether it comes from the front, back, sides, above or below the listener. This is practical in everyday life because the field of vision is more limited and thus, sounds can provide information of the things that cannot be seen. Typically, when a person hears an interesting sound, he turns his head towards the direction where he assumes the sound source is. This way, sound information is often used as an aid to the visual sense.

Sound localization has been studied actively for decades. Therefore, a lot is known about the subject. Localization is very accurate in front of the listener and less accurate at the sides and behind the listener. However, when there are many sound sources present simultaneously, the situation is different. Then, the localization task becomes much more complicated and it is also more challenging to study.

When the sounds that can be heard outdoors are considered, a single sound source could be a car driving on a quiet street. Even with eyes closed, it would be quite easy to accurately localize the moving car. If there were – in addition to the car – a bicyclist, two kids laughing, a bird singing in the tree and an airplane flying in the sky, the auditory discrimination task would be more challenging. However, the localization of these auditory events would be aided by the fact that they are rather different and therefore they would not be confused to each other. On the other hand, if there would be a dozen cars at an intersection on the street driving to different directions, the localization of all those cars with mere help of hearing would be very difficult because the sounds of the cars are very similar to each other.

1.1 Aim of the thesis

In this thesis, spatial hearing of humans is studied. The aim is to find out more about the capability to perceive details in sound source groups that are spatially complex. In order to obtain knowledge about this issue, two listening tests are conducted. The situation that the test subjects are facing in the tests could be compared to the above-mentioned situation where there are many cars on the street. This is because the sounds in the listening test are very similar to each other and they are emitted simultaneously from many directions.

The procedure of the listening tests is presented conceptually in Figure 1.1. There, the test subject sits at the center of a room and loudspeakers are placed equidistant from the subject. A number of loudspeakers emit sound and the subject perceives a certain combination of sound events, as presented in the lower right corner of the figure. Then, the subject responds according to his or her perception using a touch-screen. Finally, the response is recorded to a computer that is outside the room.

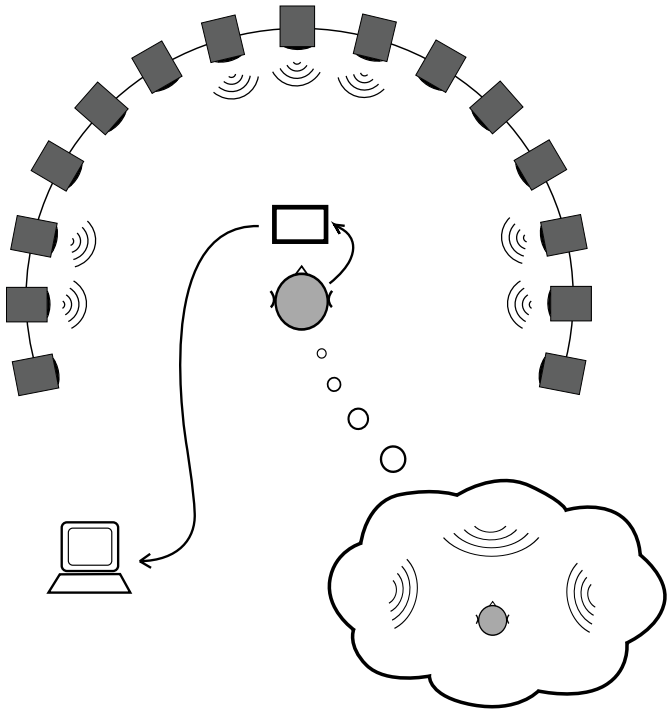

Figure 1.1: A conceptual presentation of the listening test procedure in this thesis.

In the first listening test, different loudspeaker combinations are used to produce spatially distributed sound sources. Sound source width is an important variable. By asking the test subjects what they perceived, information about the subjective perceptions of the test cases

is gathered. Spatially distributed sound sources are also used in the second test. There, the subjects are asked to discriminate which of the two presented spatially different stimuli is produced using a certain loudspeaker combination. It is also studied how the frequency range of the stimuli affects the perception.

The results from the listening tests are used as background information in developing multichannel sound systems to correspond better and more accurately to the needs of human hearing. This way, the reproduction could be done more effectively and the sounds that are played back with the sound system could be perceived as intended. A such recently proposed method for spatial sound reproduction is Directional Audio Coding (DirAC) (Pulkki, 2007) that has been developed at Helsinki University of Technology in the Department of Signal Processing and Acoustics.

Moreover, the results may be used to develop, adjust and improve computer models that predict the behavior of human hearing. Models of that kind are developed because, compared to listening tests, they are a faster, cheaper and more convenient way to test different sound reproduction methods. They are also used to model and study human hearing on a more general level. Research on such modeling has been carried out also in the Department of Signal Processing and Acoustics (Hirvonen, 2007), (Pulkki and Hirvonen, 2009).

1.2 Outline of the thesis

The thesis begins with an introduction on sound in Chapter 2 where it is discussed as a physical phenomenon. Spatially distributed sound sources are also briefly introduced. In Chapter 3, the psychoacoustics of sound, the physical properties of human hearing and spatial hearing are discussed. Some of the concepts that are needed to understand how the perception of spatial sound works are explained.

General know-how for conducting listening tests are introduced in Chapter 4. Then, research in the field of perception of spatially distributed sound sources is reviewed in Chapter 5. Both chapters are essential in the planning and conducting of the listening tests in this thesis as they form the basis that is needed in order to be able to do those tests.

The main emphasis in the thesis is on the listening tests that were conducted. As said, the two experiments concentrated on spatial distribution perception. They are presented and discussed in detail in Chapter 6. Finally, conclusions of the whole thesis and ideas of what could be done on the subject in the future are presented in Chapter 7.

Chapter 2

Sound

This chapter introduces a few physical properties of sound. Sound in space and noise signals are also discussed.

In a physical sense, sound is wave motion in the air or some other medium and it is caused by mechanical vibration. A sound can be presented in the frequency domain or in the time domain. In the frequency domain, the sound has a spectrum in which the sound is divided into frequencies. In the time domain, the sound has a waveform. In the case of a harmonic sound, the shape of the waveform repeats over and over, and wavelength refers to a distance between two sequential points that are in the same phase. Wavelength can be calculated with the equation

$$\lambda = c/f \tag{2.1}$$

where c is the speed of sound and f is frequency. The most basic type of sound is a sine wave, which consists of a single wave with a certain frequency and amplitude.

2.1 Reflection, absorption and diffraction

When a sound wave encounters a solid object such as a wall, it is partly reflected back and partly absorbed into the wall. The reflection is similar to the original sound. The reflected sound appears to come from an image source inside the wall, and the distance to the wall is the same for both the original source and the image source. The reflection is a familiar

phenomenon from the situation where one shouts towards a rock far away and the sound is reflected back and perceived by the shouter a moment later.

The absorbed part of the sound wave turns into heat in the material. Thus, the amount of sound wave's energy is decreased and the reflection is not as loud as the original sound is. As a generalization, hard surfaces reflect sound waves whereas soft surfaces absorb them.

When an obstacle comes in the way of sound waves, diffraction may happen. In this phenomenon the waves bend around the obstacle so that the sound reaches behind it. This happens most effectively at frequencies where the obstacle is about the size of the wavelength. For shorter wavelengths, mostly reflection and absorption occurs, and for longer wavelengths, the obstacle might not affect at all. Also, when sound waves go through a hole, such as a doorway or a keyhole, they pass it and spread out beyond it.

2.2 Sound in space

The acoustical properties of surrounding areas constitute significantly to the sound. Roughly speaking, the areas can be divided into two different cases: free field and enclosed spaces. Typically, a free field is found outdoors, and it is present when the sound source can be considered a point source and there are no reflecting surfaces nearby. By definition sound pressure is proportional to $1/r$, where r is distance (Rossing et al., 2002), and this means that sound pressure is halved when distance doubles. Enclosed spaces are normal human built spaces such as a living room, an office, a concert hall etc. There, the acoustical situation is much more complex than outdoors, as there are several reflecting and absorbing surfaces.

The sound that arrives to the listener in a room consists of direct sound, early reflections and reverberation. The direct sound travels the shortest possible route from the source to the listener and is the first to reach the listener. Then, the same sound is reflected from various surfaces such as ceiling and walls, and these reflections together form the early sound. After this, the sound travels more complex paths through many reflecting surfaces from all directions and this forms the reverberant sound. There may be thousands of reflections arriving in a time span of a couple of seconds. An impulse response in an enclosed space as a function of time is illustrated in principle in Figure 2.1. There, an impulse occurs in a room at time point zero. After time interval t_0 it arrives to the listener, and thereafter comes the first reflection. Finally, the other reflections and reverberation arrive.

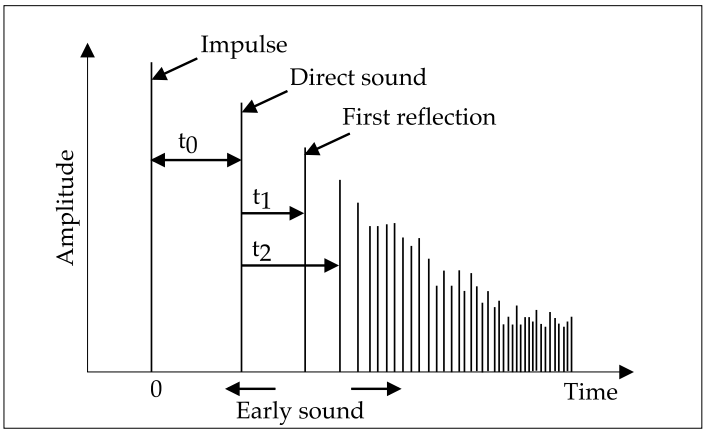

Figure 2.1: An impulse response of a room. (Adapted from Karjalainen (1999))

When a listener sits in a concert hall, he usually pays attention to the reverberation time. Even though one measure is not enough in describing the whole space, the reverberation time is a crucial factor in the quality of the hall. It is defined as the time that takes for the sound level to decrease by 60 dB (Rossing et al., 2002). If the reverberation time is too short, the space sounds dry, and if it is too long, it may distract the listening of the direct sound.

When listening to music in a concert hall, it is preferred that there is a pleasant reverberation. This is achieved by selecting various suitable materials to the walls. All materials have their characteristic reflection and absorption coefficients and the materials are used accordingly to make the hall more reflective or absorptive. In normal rooms, more absorption is wanted, because then the overall sound level doesn't become distracting.

An anechoic chamber is a room with very high absorption. It is a special room used in many acoustical measurements and research whenever it is necessary to exclude the influence of room acoustics. A free field can be achieved in an anechoic chamber in addition to outdoors.

All the surfaces of the anechoic chamber are made as absorptive as possible so that any sound would not be reflected back from them. Typically they are covered with special cones that are made of absorptive foam. The floor is often a net made of steel in the middle height of the room. This is because even the bottom of the room can then be covered with absorptive material. There is also extensive isolation of outside sounds and vibrations. All this causes the room to be very quiet, with background noise levels much lower than in a normal room.

2.3 Sound sources in space

On the scope of this thesis, it is reasonable to discuss sound sources according to their spatial attributes.

The most fundamental case is a single, point-like sound source. There, the sound is perceived to come from a certain point in space. However, the reflections and the reverberation of a room can make the sound appear to have spatial dimensions.

A sound source can also be wide and still be interpreted as a single sound source. Examples of such in our surroundings are thunder and waves of the ocean. Also, when very loud music is played in a building, the whole building vibrates and acts as a sound source. In these cases, not only the source itself is wide but also the auditory perception is wide and not point-like.

In addition to a single sound source, there can naturally be many sound sources present simultaneously. Then, they can be said to be spatially distributed. If the sound sources are very different, i.e., a guitar playing and a person singing, they are discriminated from one another quite easily. Multiple similar sound sources, on the other hand, can unite and form a single sound event. A large crowd or the leaves of a tree that rustle in the wind are good examples of this. They are spatially distributed, and even though the individual sound sources could be pointed out one by one, the perceived sound is a single spatially wide sound event.

2.4 Noise signals

Noise signals are random signals that typically sound like a hiss and can be heard e.g. on the radio frequencies that have no program. Also the above-mentioned waves of the ocean sound very much like noise. Noise is generally considered as unwanted sound. Yet, there are different kinds of noise signals and they have characteristics that make them useful in some testing procedures.

White noise can be seen as a sort of basic noise: it has equal energy density in all frequencies. This means that there is always the same power per certain bandwidth in any part of the signal and thus the spectrum is flat. The spectrum of white noise is illustrated in principle in Figure 2.2.

Pink noise has the same power in each octave or one third of an octave. Therefore, it is more closely related to human hearing system because humans do not perceive absolute changes on the frequency scale but rather relative changes in bandwidth are perceived equally large. For example in the case of a sine wave, a shift in frequency from 120 Hz to 160 Hz is perceived as big a change as from 1200 Hz to 1600 Hz. This is further discussed with the topic of critical bands. Because of its practical relation to human hearing, pink noise is popular when doing listening tests or some other research. The spectrum of pink noise is illustrated in principle in Figure 2.3. It can be seen that compared to the spectrum of white noise, the power density falls off when the frequency increases.

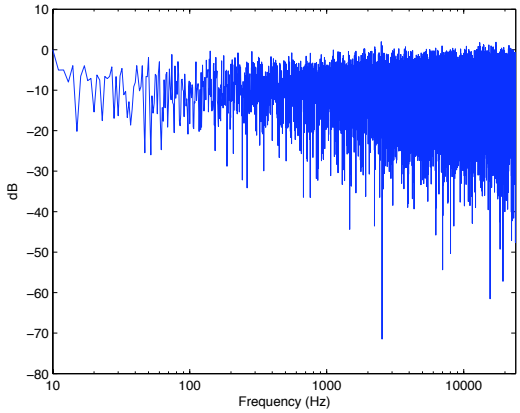

Figure 2.2: Spectrum of white noise.

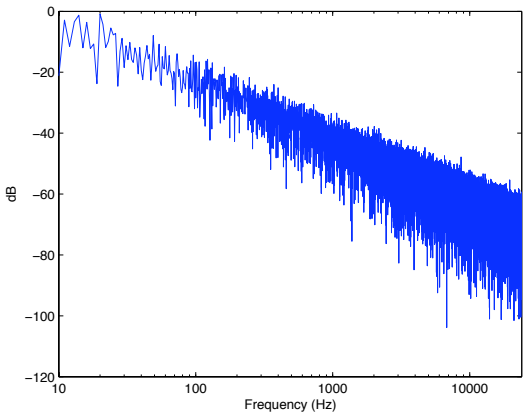

Figure 2.3: Spectrum of pink noise.

Chapter 3

Psychoacoustics

This chapter presents psychoacoustical concepts and attributes. These are related to the way how humans perceive sounds in general. A simplified introduction of the structure of the human ear is presented. Many of the topics are related to spatial hearing. They are important to understand in order to be able to examine the results of the listening tests in this thesis.

Spatial hearing is highly aided by the fact that humans have two ears. Other factors that help in spatial hearing are the shapes of the external ears, head and torso. Because they are asymmetric the sounds that arrive from different directions are shadowed differently and reflected for example from the shoulders. Sounds that arrive from the front and from the back are easily confused to each other but the pinna aids by affecting them differently. Also helpful is moving of the head as then the directional cues change.

3.1 Structure of the human ear

The human ear can be structurally divided into three parts: the external ear, the middle ear and the inner ear. These are illustrated in Figure 3.1.

The external ear simply transmits the signals further into the ear. It consists of three parts: the pinna, the ear canal and the tympanic membrane. The pinna is the visible part of the ear and helps in directional hearing by filtering sounds depending on their directions. The ear canal amplifies signals around 3 - 4 kHz because at those frequencies it can be seen as a linear resonator. The tympanic membrane, also called the eardrum, changes the wave

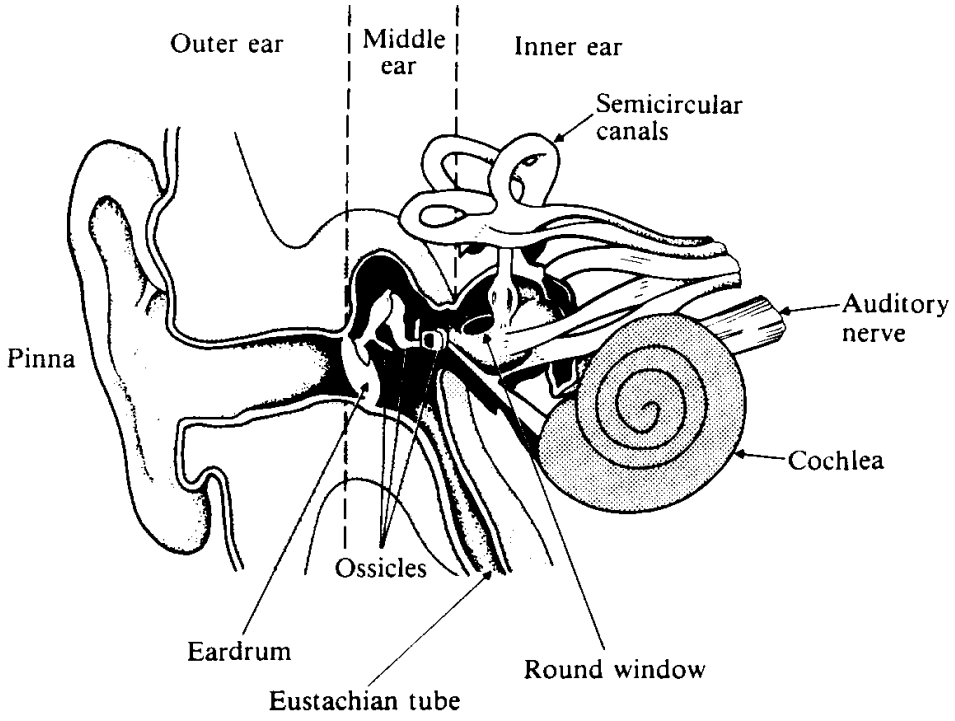

Figure 3.1: The structure of the human ear. Only some relevant parts are presented and named. The drawing is not to scale – the middle ear and the inner ear have been enlarged. (Adopted from Rossing et al. (2002))

signals in the air into mechanical vibrations in the ossicles.

The middle ear consists mainly of the ossicles. They are three bones called malleus, incus and stapes, which are actually the smallest bones in the human body. They conduct the mechanical vibrations further into vibrations in the liquid of the inner ear. This impedance matching of the middle ear is essential because otherwise the signal would be significantly weaker in the liquid. The Eustachian tube is a connection between the middle ear and the oral cavity and does pressure equalization when needed.

The most important part of the inner ear is the cochlea. Also semicircular canals are located in the inner ear, but they are not related to hearing but are necessary for balance. The cochlea is a shell-shaped organ about 35 mm in length. A detailed cross-section of it is illustrated in Figure 3.2. Inside the cochlea there is the basilar membrane, which is essential in hearing. On the basilar membrane is the organ of Corti that has numerous inner and outer hair cells attached to it. The hair cells are also in contact with the tectrorial membrane above the basilar membrane. This structure continues throughout the cochlea so that there are about

20 000 - 30 000 hair cells in total (Karjalainen, 1999).

The stapes in the middle ear makes the liquid inside the cochlea vibrate, which causes a traveling wave in the basilar membrane. The traveling wave causes movement between the tectorial membrane and the basilar membrane and this moves the hair cells. Then, the hair cells send neural spikes via the auditory nerve to be processed in the brains. There, the information about the location and amplitude of the traveling wave is processed. Finally, the whole complex process leads to perception.

Different parts of the basilar membrane react to different frequencies of sound waves – the beginning reacts to high frequencies and the end to low frequencies. This is because the basilar membrane has different width, mass etc. along its length. The width changes so that the basilar membrane is widest in the end.

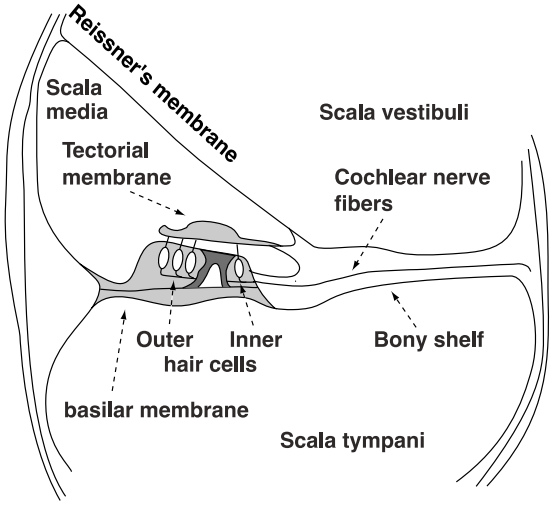

Figure 3.2: Cross-section of cochlea. (Adapted from Karjalainen (1999))

The frequency range of human hearing reaches approximately from 20 Hz to 20 000 Hz. Typically, when a person gets older, the perception of higher frequencies deteriorates so that already frequencies of 10 000 - 15 000 Hz may not be perceived. Also, very loud short sounds or long-lasting unwanted noise may permanently damage hair cells and thus cause deterioration of hearing.

Binaural hearing refers to listening with two ears, whereas monaural hearing refers to listening with one ear only. The fact that humans have two ears at the opposite sides of the head plays a very important role in spatial and directional hearing. This is discussed in more detail later in this chapter.

3.2 Auditory attributes of sound: Pitch, timbre and duration

The human auditory system perceives different attributes of sound on a relative rather than on an absolute scale. This is why common physical measures such as frequency or decibels do not directly correspond to subjective interpretations of sounds.

A harmonic sound has a frequency spectrum where there is a fundamental frequency and harmonic components that are higher in frequency. These affect to the perception of that sound. Pitch is a measure closely related to the fundamental frequency, as they both try to rate sounds on a scale from low to high. However, pitch concentrates on subjective perception of sound whereas in the case of harmonic sounds, the spectrum and the fundamental frequency are measurable attributes. As the fundamental frequency of a harmonic sound increases, the pitch also increases, but not as rapidly. The unit that is used for the measurement of pitch is called mel. The mel scale is constructed by having test subjects judge when they perceive a doubling of pitch. Up to about 1 000 Hz the doubling of frequency approximately equals to the doubling of pitch, but after that the frequency must be increased much more than pitch, as they have almost logarithmic relationship. The Bark scale and Equivalent Rectangular Bandwidth (ERB) scale are related to pitch, and this is discussed later in more detail.

Timbre is an attribute of a sound that has to do with time-dependent frequency spectrum of sound. When a note of same pitch and loudness is played with two different instruments, they are distinguished from one another because the timbre of the notes is different. Timbre cannot be directly measured with a physical attribute, but the formants of a sound are a crucial part in forming it. As mentioned above, the formants can be seen in the spectrum of a sound.

The subjective duration of sound is very close to actual duration of sound. For signals longer than a few hundred milliseconds the subjective and physical durations match, but for signals shorter than 100 ms the subjective duration is longer. Also, if a very short sound is repeated with equally short pauses in-between, the pause is perceived as being shorter than the sound.

3.3 Auditory attributes of sound: Loudness

Loudness represents subjective sound level. The perception of loudness level is highly dependent on frequency. Equal-loudness contours that are presented in Figure 3.3 show how loud different sound pressure levels are heard at different frequencies. These curves are measured using pure tones, i.e., single frequency at a time. They were first determined in 1933 (Fletcher and Munson, 1933) and updated in 1956 (Robinson and Dadson, 1956).

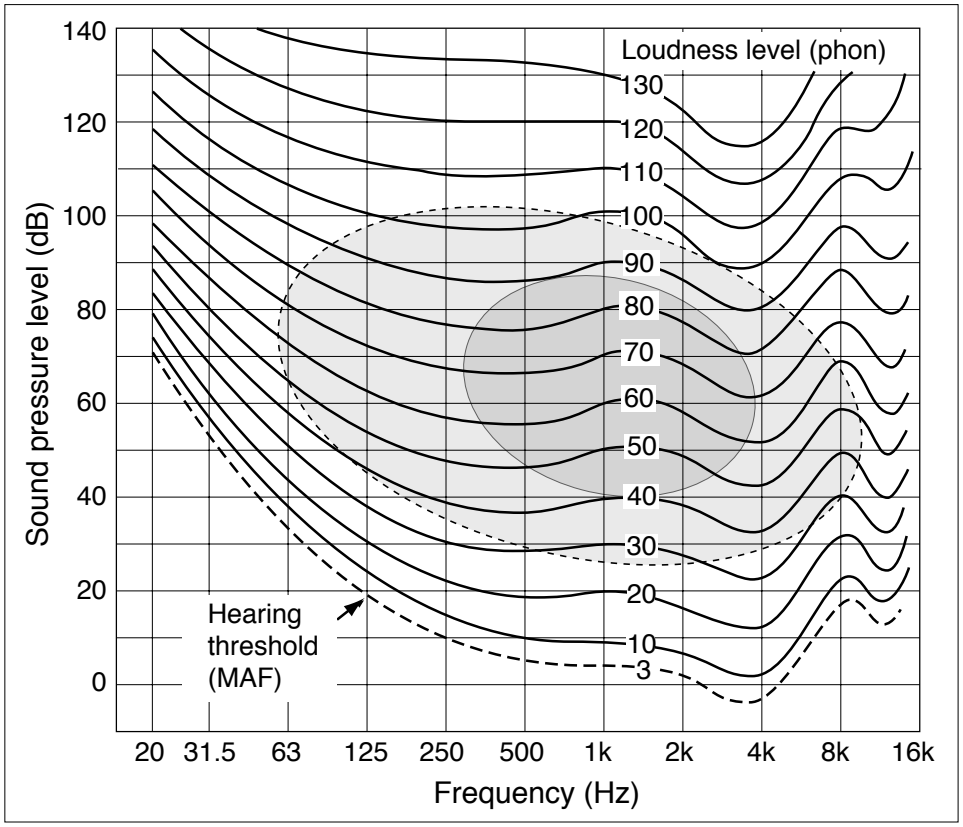

Figure 3.3: Equal loudness level contours for pure tones in phons. (Adapted from Karjalainen (1999))

The unit of loudness level is phon, which is anchored to 1000 Hz, meaning that for 1000 Hz a certain sound pressure level in decibels corresponds to the same value in phons. In Figure 3.3, along a single phon curve, human always perceives the pure tones equally loud, but the decibel value changes with frequency. For example, 30 phons in 1000 Hz is 30 dB, but 30 phons in 63 Hz is approx. 60 dB.

Also, some interesting features of the human hearing can be seen in Figure 3.3. Hearing

is most sensitive in the area from 3000 to 4000 Hz, which is caused by the ear canal's first resonance. Overall, the sensitive part of hearing is along the lines from a few hundred Hz to 5 or 6 kHz. These frequencies contain the most important parts of speech when concerning intelligibility. Consequently, the human hearing works very well with speech. Lastly, it should be noted that the lowest curve in the figure is the hearing threshold beyond which humans cannot hear the tones, and for very high sound pressure levels, the measurements cannot be made because they would be unpleasant to the test subjects.

Because of the fact that perception of loudness doesn't directly correspond to sound pressure level, a few weighting filters for decibel measurements are constructed. The usual ones are called A, B, C and D-weightings, illustrated in Figure 3.4. The most commonly used is A-weighting, but all of the options have their advantages. A-weighting tries to mimic equal loudness level contours with a parabolic curve. The low frequencies are attenuated, whereas around 2 kHz, there is a very slight amplifying, and further at high frequencies again a small attenuation.

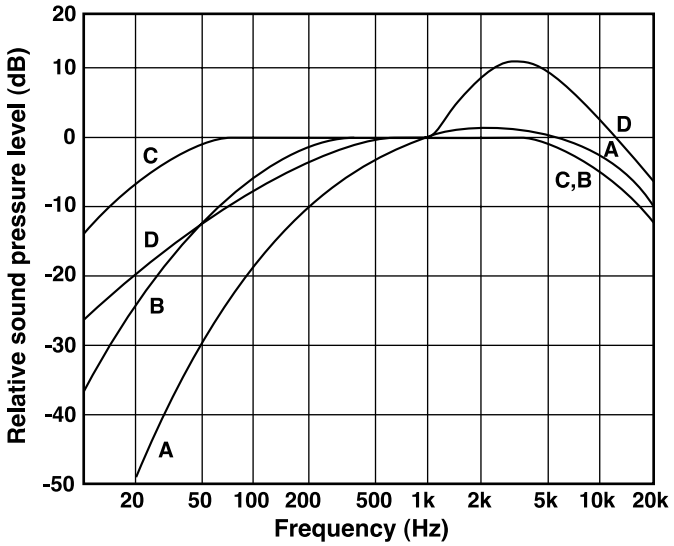

Figure 3.4: Weighting curves for sound level. (Adapted from Karjalainen (1999))

3.4 Critical bands and equivalent rectangular bandwidth

When a broadband signal is presented to a listener, it is processed in hearing mechanisms so that parts of the signal close to each other in frequency are treated as one entity. The width of one entity is defined by critical bands (Zwicker et al., 1957). This is related to the

basilar membrane, where the hair cells that are close to each other are functioning together.

For example, at the center frequency of 1000 Hz the width of the critical band is 160 Hz. This does not mean that critical bands are "locked" to certain places – rather, they are formed around any centre frequency. Critical bands are wider at higher centre frequencies, but the relationship is not linear. Rather, the width of the critical band is almost constant up to about 500 Hz and after that it starts to increase.

The width of a critical band is typically measured using noise signals. Reference sound, narrow-band noise is compared to an adjustable noise signal whose bandwidth is altered, as noise pressure level is kept constant. The noise signal is adjusted to match the subjective loudness level of the reference sound. Then the bandwidth is adjusted wider, and for some time the loudness level does not rise. The point when the loudness level starts to increase is critical bandwidth. This measurement is done for several selected centre frequencies and eventually a thorough view of dependence between frequency and width of critical band can be achieved.

The Bark scale tells the pitch of the sound with the help of critical bands. It is defined so that one Bark equals to one critical band. The point of a scale like this is that it is more closely related to psychoacoustics and the basilar membrane than the Hertz scale. The Bark scale can be linearly compared to the length of the basilar membrane so that ideally one Bark is always a leap of same length along it.

A more accurate method to match the relationship with the basilar membrane and frequency is Equivalent Rectangular Bandwidth (ERB). It is measured very much like critical bandwidth except that there is masking noise on both sides of the test signal in order to prevent off-band listening, i.e., to prevent the other parts of the signal from affecting to the listening. The bandwidth of the unmasked part of the signal can be adjusted, and modifying this further we get the ERB-rate scale which is comparable to the Bark scale.

An ERB-rate scale indicates that one ERB equals about 0.9 mm on the basilar membrane. The bandwidth of one ERB on a certain center frequency can be calculated with the formula

$$ERB = 24.7(4.37f + 1) \tag{3.1}$$

where ERB is the width of the ERB band and f is center frequency in kHz (Glasberg and Moore, 1990). From (3.1) we get a formula that approximates the ERB-rate scale

$$R_{erb} = 21.4 \log_{10}(4.37f + 1) \tag{3.2}$$

where R_{erb} is the number on the ERB-rate scale and f is frequency in kHz (Glasberg and Moore, 1990).

3.5 Masking

Masking is a phenomenon where, for example, a weaker sound becomes inaudible because a louder sound prevents it from being perceived. The masking effect may depend on frequency, intensity and even time. It is quite common that when two or more sounds are present at once, one may mask the others. In general, masking increases the threshold of hearing for other sounds.

White noise masks sinusoidal tones of all frequencies if they are adequately weaker in intensity. The relationship between noise and masking level is approximately linear - when noise level is raised 10 dB, the masking level i.e. the hearing threshold is also raised 10 dB.

When narrow-band noise is used, a tone with frequency very close to the centre frequency of the noise is masked most effectively. In respect to the centre frequency, tones of higher frequency are masked more effectively and farther away than tones of lower frequency. The louder the narrow-band noise is, the wider range of frequencies it masks.

A sound can also mask sounds that do not occur simultaneously with it in time – a weaker sound might not be perceived slightly before or after the masking sound. This is called backward masking and forward masking. Forward masking may last up to 150 - 200 ms after the masking sound. The masking sound must be long enough and the masked sound must be short and adequately quieter than the masking sound in order for the masking to happen. Backward masking is less noticeable as it is effective only 5 - 10 milliseconds before the masking sound. Yet, the phenomenon is interesting because even though the ear picks up the masked sound first, it is not perceived because the masking sound overrides it.

3.6 Localization

Localization defines the human's ability to perceive the direction of the sound source and localize the sound event at the position of the sound source. The accuracy of localization varies between directions around the head. Biologically thinking, the purpose of the localization is to draw one's attention and turn the face to the correct direction and then further analyze the situation visually. This proposes that accurate localization is not essential in all directions.

It has been found out that in the horizontal plane the accuracy is best directly in front of the listener (Blauert, 1997). The results of two extensive studies on the subject are shown in Figure 3.5. The task of the listeners was to place a sound-emitting loudspeaker to four positions: exactly at the front, behind, left and right. The test signals were 100 ms white-noise pulses. As said, at the front the task is performed well. At the rear the localization is also accurate even though there is a little more deviation. However, at the sides the listeners placed the loudspeaker notably closer to the front, about 10 degrees. The results are slightly different when using different types of signals, but the main principle remains the same. When using tones of a single frequency, the results vary. Especially when a sound signal from the side is around 1.5 kHz in frequency, localization becomes very difficult. This is because the wavelength is about the size of the head and thus the localization cues are vague.

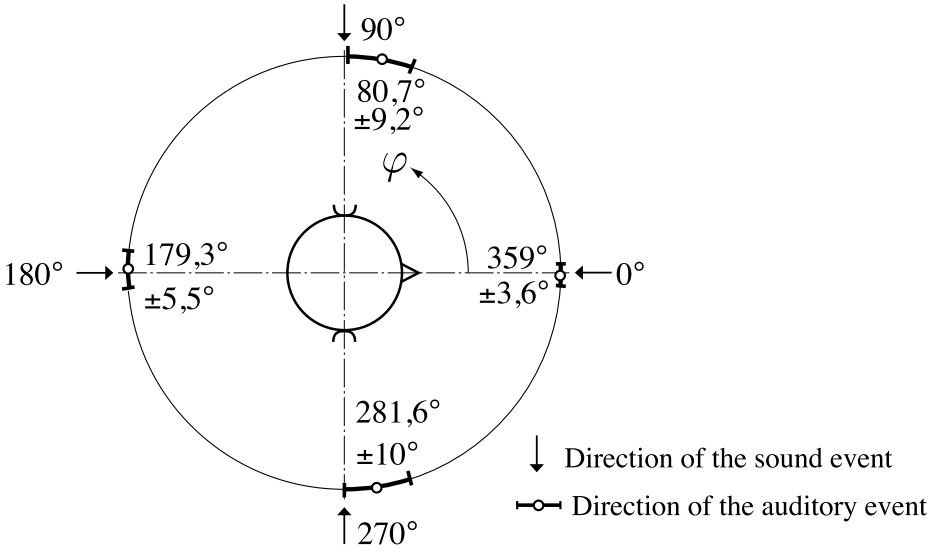

Figure 3.5: Direction-dependent localization. (Adapted from Blauert (1997))

Localization blur refers to inaccuracy of localization, more precisely defined as the smallest change in an attribute of a sound event that causes a change in the location of the auditory event. In the above-mentioned experiments the localization blur was found to be larger at the sides than in front of the listener. Localization blur can also be tested for example by altering the position of a loudspeaker and asking the listener when the change in location is just noticeable. In general, the localization blur increases as the displacement from the forward direction to either side is increased (Blauert, 1997).

Lateralization is a specific sort of localization used in context with headphone listening (Yost and Gourevitch, 1987). When the sounds from both sides are merged into one sound event, it is heard inside the head. This is called lateralization and it is one-dimensional along the axis of the ears. Lateralization can be artificially adjusted by altering the temporal difference or loudness level difference between the two ear signals. The reason why sounds are usually heard in a limited area inside the head is because the directional cues and reverberation are not natural.

3.7 Distance cues

Perceiving the distance of a sound source seems to be more challenging than perceiving direction, even though there are various distance cues that help in the task. These cues include room attributes - reverberation and reflections, particularly early reflections - timbre and spectral properties of a familiar signal, filtering caused by air absorption and overall loudness. The last-mentioned is an important cue, but it easily leads to false estimations, as the actual sound level of the source is often not known and thus it might simply be an attribute of the source signal rather than indicating sound source distance.

When the test signal is more familiar, such as human speech at normal loudness, the distance of the auditory event is quite close to the actual distance of the sound source. However, even for familiar signals, there are misleading situations. An experiment was made with four different signals: a human shouting, speaking normally, speaking with a low voice and whispering (Gardner, 1969). They were presented from various distances in the front direction. As said, for both normal speech and low voice the judgements were accurate. More interestingly, auditory event of shouting was consistent with distance but was always judged to be a little farther away than the sound source. Whispering was perceived to be closer than the sound source, and moreover, the perception remained close even though the actual sound source was further away. These results are, at least partly, affected by the

association of an actual situation where shouting usually happens far away and whispering happens close.

3.8 Precedence effect

The precedence effect (Wallach et al., 1949), also called the law of the first wavefront, is the effect of directional hearing that makes us localize the sound event according to the direction of the sound that arrives first. This is practical because the first sound is usually the direct sound and thus the localization of sound source is done correctly.

There are actually a few different perceptual effects that all contribute to the precedence effect (Litovsky et al., 1999). These include summing localization, fusion of sound sources, law of the first wavefront and lag discrimination suppression.

In a normal room, a single person speaking is perceived as one even though there are numerous reflections from the surfaces of the room in addition to the direct sound. Instead of being individual sounds, the reflections contribute to the overall loudness or timbre of the sound and to the overall sense of space. The perceived location is at the leading source. This is called fusion of sound sources and happens with reflections arriving approximately 1.5 - 40 ms after the direct sound. The reflections that arrive less than approximately 1.5 ms after the direct sound influence the direction perception together with the direct sound. This is called summing localization. At time values larger than 40 ms, the sounds are heard separately, the latter ones being perceived as the echoes of the first one.

Lag discrimination suppression affects during the fusion of sound sources. Then, any changes in the location or some other property of a sound that arrives directly after the leading source are hard to discriminate.

3.9 Interaural time difference and interaural level difference

Humans can perceive sounds coming from different directions and localize the auditory event accordingly. The most important cues in accomplishing this are interaural time difference (ITD), interaural level difference (ILD) and monaural cues. The concept of ITD and ILD was first hypothesized as early as 1907 (Lord Rayleigh (a.k.a. J. W. Strutt 3rd Baron of Rayleigh), 1907). The main cause for these attributes is the fact that the ears are located

at the opposite sides of the head, which causes time differences and shadowing. Because of this, the ear input signals are different from each other (except, for example, in the case of cone of confusion, discussed later) and they change when direction of sound changes thus making it possible to distinguish signals coming from different directions.

When the sound travels towards the listener, it arrives first at the ear closer to the sound source and then travels around the head to the opposite ear. This results in time difference, i.e. delay, in the ear signal, and that delay is called ITD. This kind of situation is illustrated in Figure 3.6. There, a sound comes from 45° to the left. The left ear receives the sound signal first and right ear slightly later, which causes a time difference in the ear inputs. A maximum time difference would occur when a sound comes exactly from the side of the head.

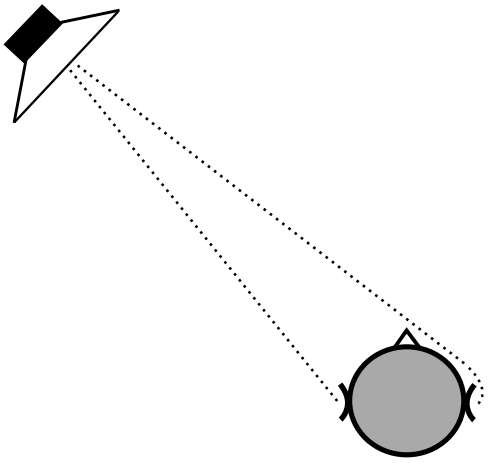

Figure 3.6: A sound arriving to the listener from 45° to the left.

The head shadowing causes the signals to be weaker in the opposite ear. Also, the reflections from pinna and shoulders affect certain frequencies of the signal in one ear. These effects together cause ILD of the ear input signals.

The cone of confusion, illustrated in Figure 3.7, represents a group of positions in which the sound incidence angle to the ear is the same and therefore the interaural cues are identical. This means that if a sound comes from one point in a cone of confusion, the sound may be localized coming from any of the positions within the same cone of confusion. Exact localization is difficult, but head rotation is very helpful since it changes the interaural cues. Also, spectral information may help in the task.

The simplest forms of lateral localization with interaural differences can be experimentally demonstrated by means of headphones and simple signals. When the signals at left and right

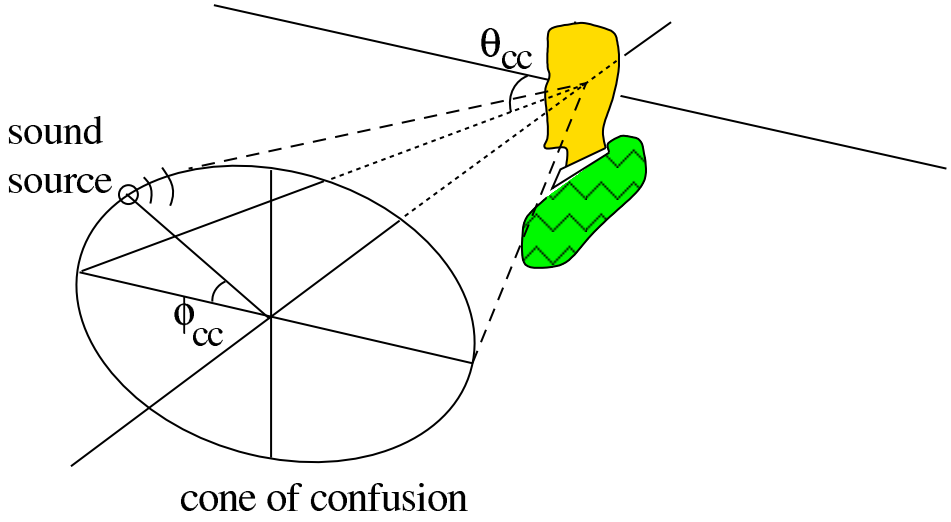

Figure 3.7: Cone of confusion. Within this cone, the interaural cues are the same and this causes localization difficulties. (Adopted from Pulkki (2001))

ears are identical, the auditory event appears in the median plane. When the sound pressure level is kept constant and the listener hears the same signals at both ears, only temporally displaced, the auditory event is shifted towards the ear that receives the first signal. On the other hand, when there is no temporal displacement but the amplitudes are different, the auditory event is shifted towards the ear that receives the louder signal. In natural listening situations, usually both of these effects are present and affecting the lateral localization.

Both ITD and ILD affect the localization at all frequencies, but according to traditional point of view ITD cues are more important at lower frequencies, approximately below 1.5 kHz, whereas ILD cues dominate above 1.5 kHz (Wightman and Kistler, 1992). This is due to the fact that the shadowing effect caused by the head becomes more significant only when the wavelength is about the diameter of the head or smaller than that. Thus, at high frequencies the level differences in the ear signals are more significant and at low frequencies they are relatively small. Nevertheless, ILD cues are present and affecting the localization also at low frequencies. ITD cues, on the other hand, are based on phase differences in the ear signals. At high frequencies, ITD is larger than the duration of one cycle of a signal and therefore phase difference cannot be used. Instead, ITD is then based on signal envelope.

Interaural cues give high-resolution information, as just noticeable difference in ITD is about 10-20 μs (Klumpp and Eady, 1956), and in ILD about 0.5 - 1.0 dB (Mills, 1960).

Complex sounds that contain various frequencies are localized with the help of interaural

cues in a very similar way as simple sounds, but moreover, it seems that certain frequencies are dominating the localization. This is discussed in more detail in Chapter 5.3.1.

3.10 Head-related transfer function

Head-related transfer function (HRTF) is a way to describe how a sound signal is transferred and modified when it travels from the source into the ear canal. It takes in account the effects of the shoulders, the shape of the head and the pinna (Møller et al., 1995). It is defined separately for both ears.

An HRTF is usually measured in an anechoic chamber with a loudspeaker placed a few meters away from a test subject. The loudspeaker is moved around the subject - alternatively, the subject rotates himself. Small microphones are placed in both ears of the subject – they may be at the beginning of the ear canal, in the ear canal or at the eardrum. The responses to signals are measured from several directions so that a complete sphere is covered.

Measuring HRTFs is a time-consuming process. Every individual has his or her own HRTF because the shape and measures of the head and torso are different, and so the measurement should be done for each test subject individually. A test subject has to be at the same spot until the response from approximately one hundred sound directions in a free field have been measured. This is why it is very common to use a specific artificial head, also called a dummy head. Typically, a dummy head consists of a head and upper part of the torso and its measures represent those of a typical human being. Small microphones are placed at the position of the ear canals, and they are surrounded by artificial external ears. The HRTF that is measured in this way is a sort of average HRTF for all the people, which is practical but also produces less optimal sound direction perceptions. For example, a sound may be perceived coming from behind when it actually comes from the front. Nevertheless, a spatial effect is achieved.

3.11 Multimodal perception

Even though human senses are usually discussed and treated separately in literature, they are not independent. Natural situations are often interfering with many senses. This means that hearing - especially spatial hearing - can be affected not only by auditory but also by visual or even tactile information simultaneously. This is called multimodality or multi-

modal perception (Warren, 1999). An everyday example of this effect can be demonstrated with a television: even when there is only one loudspeaker emitting sound on the side of the television, the newsreader's voice is heard in the direction of the lips. In general, visual information actually helps in directional hearing, as can be seen from the fact that localization is accurate in the visual area but much more inaccurate outside it.

Multimodality causes certain challenges when performing listening tests. In the case of headphone listening, the spatial information of sound is likely to be inconsistent with the surrounding room. In other words, visual and auditory information do not match, and this causes the listener to conclude that the sound is artificial and doesn't belong to that particular space. This is the case when there are no loudspeakers visible. However, if loudspeakers are placed visibly in suitable positions, the sound from the headphones may actually feel like coming from the loudspeakers. The listener is more likely to think that the sound is genuine if he can imagine it coming from a particular source.

Also, when performing for example a test of the effects of loudness on distance perception (discussed earlier in chapter 3.7), there should be many loudspeakers placed on various distances from the listener. This way, the visual information helps to believe that the sound may come from far or near - even though the actual test signal may always come from the same loudspeaker as only the loudness varies.

Chapter 4

How to conduct a listening test

Before a report is finished with analyzed results of a listening test, an extensive amount of work has to be done. Careful preparation, planning and beforehand testing is the key to success in listening tests. This chapter reviews some commonly known methods as well as some details that should be considered when conducting a listening test.

The task of conducting a listening test is very challenging and detailed because all the variables of the different parts of the listening test have to be taken into account. In order to gain results that can be generalized and truly do answer to the questions asked, the experimental procedure must follow a widely accepted standard or some common methods. Also, when reporting the results, it is important to describe all the methods and equipment used so that the experiment is reproducible and the results can be compared with those from another experiments.

There are a number of things that have to be considered when planning a listening test. Appropriate test signals must be constructed, the reproduction system and listening room must meet certain requirements, and the subjects have to be selected to suit the needs of the test (Bech and Zacharov, 2006). Also of importance are the things that are not directly in relation to the presentation of the stimuli: the formation of the question to the test subjects and the answering method, i.e., the response format with which the subject quantifies the auditory impression.

The goal is to make the stimuli so that only the attribute of interest is altered between the cases. Because this is rather difficult, also other attributes may change and therefore it is very important to ask a question that makes the test subject concentrate on the attribute of

interest.

It is not possible to directly measure the actual sensory response or perception of a test subject. Instead, the subject has to express his or her perception in some way and therefore the collected results are always influenced by unwanted additional variables.

The question of suitable subjects depends on the desired generalization of the test results. If the results are supposed to represent the population in general, the subjects should be selected randomly from the whole population. These subjects do not have any specific knowledge of the field of the listening test and are therefore called naive subjects. So-called experienced or expert subjects have previous experience in the field of the listening test or have participated in similar listening tests before. These kind of subjects are usually selected when certain knowledge is needed in order to be able to perform the test or when the focus is on a detailed analysis of some particular aspect. When using experienced subjects, the test results are not general but rather imply how well or how accurately is it possible to hear the test stimuli.

4.1 Methods for measurements

So-called classical psychophysical methods were first described by physiologist Gustav Fechner in 1860 (Goldstein, 2002). These were the original methods that were used to measure the stimulus-perception relationship. Often it is desired to measure the smallest amount of stimulus energy that is needed for the perception of a stimulus – for example the smallest loudness level in which a sine wave is perceived. This is called the absolute threshold, and Fechner described three methods for performing this: the method of limits, the method of adjustment and the method of constant stimuli.

In the method of limits, the stimulus parameter value is increased or decreased, and after each presentation of the stimulus the test subject is asked to indicate whether he or she perceived it or not. When this is done a number of times, half of the time increasing and half decreasing the stimulus, the absolute threshold is obtained as the mean value of the points where the perception changed.

The method of adjustment is much like the method of limits, but in it the stimulus parameter value is adjusted in a continuous manner, and it may be adjusted by the test subject himself. The value is adjusted until the stimulus is only barely heard by the subject, and this value is taken as the absolute threshold. Again, this should be done a number of times so that more

reliable average value is obtained.

In the method of constant stimuli, a number of stimulus parameter values are chosen and presented to the subject a number of times. As in the method of limits, the subject indicates whether he or she perceived the stimulus or not. Then, the percentage of perceptions is calculated and typically plotted in a graph where a curve is fitted to the data points. The estimated threshold is the point in the curve where 50% of the stimuli are perceived. For the results to be reliable, the number of repetitions is usually quite high.

Naturally, the absolute threshold is not the only attribute that is measured. The difference threshold is the smallest difference between two stimuli that a person can perceive. Two stimuli with different parameter values are presented to the test subject, the standard stimulus and a comparison that is being changed. Quite logically, large differences are perceived easily, but small differences are harder to perceive. Interestingly, the ratio of the difference threshold to the standard stimulus is constant, so when the parameter value of the standard stimulus increases, so does the difference threshold (Goldstein, 2002).

Scaling procedures can be divided into direct scaling and indirect scaling methods (Bech and Zacharov, 2006). The above-mentioned difference threshold is an indirect scaling method. Direct scaling methods such as magnitude estimation, ratio comparison and cross-modality matching are used to find out a scale that relates the perceived magnitude to the physical stimulus, e.g. perceived loudness level to the level in decibels (Yost, 1994).

Matching procedure is a slightly more complex method as there the two stimuli are usually different in many attributes. The task of the test subject is to adjust a certain one of these stimulus attributes so that the two stimuli are perceived as equal in that attribute (Yost, 1994).

The method, response format and the whole user interface can be seen as a link between the subject's sensory perception and the results obtained. Therefore, the correspondence between the perception and response should be clear to the subject. The subject must feel that the available response scale is appropriate in quantifying the sensation and also, he or she must know how to use the user interface. These issues are solved by choosing a suitable response format and by familiarizing the subject with the whole test procedure prior to the test. Often a training session is held prior to the actual test.

The methods described are not the only ones that are used in psychophysical studies but provide a firm basis. Nowadays, more and more detailed aspects of human hearing are studied and for some experiments, more complex testing methods are needed. However,

the simpler the method, the more likely it tests the desired phenomenon. Because of this, the classical methods are always very convenient and widely used.

4.2 Selection of testing environment

The space where the listening test is performed must be suitable for the test in question and meet certain requirements. By controlling the key characteristics of the listening space, a stable acoustical environment should be ensured. Typically these characteristics are background noise level and spectrum, reverberation time and minimization of external disturbances such as sounds and vibration (Bech and Zacharov, 2006). There are recommendations and technical specifications of suitable listening rooms for various situations.

An anechoic chamber is an appropriate space for many listening tests, including sound localization tests, as the acoustics of it are very well controlled. There, the reverberations and background noise are at the minimum, and external disturbances are usually avoided by effectively isolating the chamber from other parts of the building. This makes it possible to present test stimuli in an undisturbed space. However, it must be noted that the listening situation in an anechoic chamber is not very natural. Therefore, it is important to use such a space only when it is justified.

4.3 Loudspeaker and listener positioning

The loudspeaker setup and the position of the listener depend mainly on the listening room and the selected reproduction method. There are some recommendations that apply to all situations, whether the case is monophonic, stereophonic or multichannel reproduction. It is important to avoid unwanted effects such as standing waves caused by conflicts between the room measures and the loudspeaker positions. The loudspeaker height should be 1.2m from the ground to the loudspeaker's listening axis. This is because the listener's ears are assumed to be at the same height. The listening position should be over 1.5m from the side or back walls of the listening room (Bech and Zacharov, 2006).

Chapter 5

Perception of spatially distributed sound sources

Localization in general has been studied extensively for decades. Most of the studies have concentrated on the localization of a single sound source. Some of those studies were discussed earlier in Chapter 3. This chapter concentrates on the perception of several simultaneous sounds in spatial conditions. Several studies on the subject are reviewed and discussed – especially research on spatially distributed and spatially wide sound sources.

When auditory spaces such as concert halls or auditoriums are discussed, certain terms are typically used to describe different aspects of the acoustics of that space. Among them, a few terms contribute to auditory spatial impression – these spatial attributes are apparent source width, listener envelopment and intimacy (Cabrera et al., 2004). Apparent source width is an attribute that a concert hall has if the music performed in it is perceived to emanate from a wider source than the actual visual width of the source (Beranek, 1996). It is related to the overall spaciousness and is affected by the early reflections of the listening space. Listener envelopment is an attribute of a listener's impression of the strength and directions from which the reverberant sound seems to arrive. When the reverberant sound is perceived as arriving at a person's ears equally from all directions, the listener envelopment is rated highest (Beranek, 1996). Intimacy refers to the impression that the music played in a concert hall sounds like it is played in a small concert hall (Beranek, 1996). Perceived spatial width, as used in this thesis, is a different concept than the apparent source width since the latter is more dependent on the reflections of the listening space. In the listening tests of this thesis the perceived spatial width is studied in an anechoic chamber.

5.1 Multiple simultaneous sound sources

As said, a large portion of localization studies have only used single sound sources in their experiments. When multiple sound sources are present at the same time, localization task becomes more challenging. A commonly known phenomenon related to this is the cocktail-party problem (Cherry, 1953). There, two or more persons are speaking simultaneously and the listener has to recognize what one of them is saying. The name, cocktail-party problem, refers to a situation where one could face such a discrimination task.

The effect of multiple simultaneous masking sounds on speech intelligibility was studied by having target speaker at the front and masking sounds in different spatial locations (Hawley et al., 2004). The experiments were made using headphones. The results indicated that when the task was performed binaurally, speech intelligibility was better than with monaural listening. The difference in intelligibility was higher when there were more masking speakers. This effect did not occur when the masking sounds were noise signals. Also, speech intelligibility was better when the masking sounds were spatially distributed on the sides than when all sounds were at the front. Overall, the results suggested that in this task, binaural hearing was more efficient than monaural hearing.

5.2 Perceived spatial width and distribution using headphones

In the field of spatial sound, perceived spatial width is an interesting issue. The first studies on the general subject of sound source wideness were performed in 1926 (Boring, 1926). Many aspects of the subject have been studied and many are yet to be studied. A large number of experiments on spatial width have been done using headphones (Mason et al., 2005). There, the situation is different from loudspeaker experiments, as in the case of headphones the perceptions of sound sources are localized inside the head, whereas with loudspeakers the perceptions are localized outside the head.

In a headphone experiment, effects of signal loudness level and duration were studied (Perrott and Buell, 1981). Broadband noise was used with different durations and loudness levels, and the so-called images of the signals were found to expand as the duration or loudness level was increased. In other words, the perceived width was increased as either one of these attributes was increased. A similar effect of duration was earlier found with sinusoids (Perrott et al., 1980). Importantly but not surprisingly, uncorrelated noise into two ears was found to produce a wider perception than correlated noise into two ears or monaural noise.

The effect of interaural cross-correlation (IACC) has also been studied using headphones. The results of such an experiment are presented in Figure 5.1. When IACC had the value of 1, i.e., when two equal signals were played to both ears, a single sound event was perceived in the center of the head, and when IACC had the value of 0, one sound event was perceived in each ear (Chernyak and Dubrovsky, 1968), (Blauert and Lindemann, 1986). These findings were obtained with pink noise. When IACC was between these cases, both splitting and widening of the perception occurred (Blauert and Lindemann, 1986), as can be seen in the middle rows of the figure.

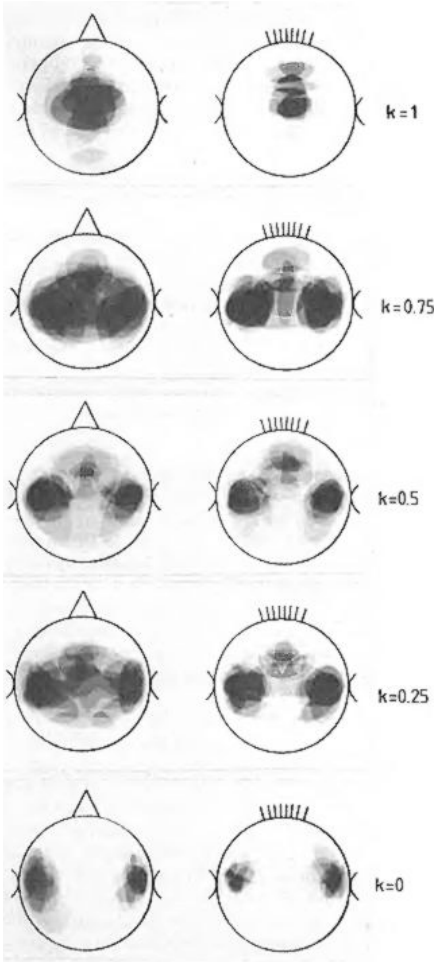

Figure 5.1: The effect of interaural cross-correlation on spatial perception with headphones. The value of IACC is denoted with k. The black areas represent the averaged sound event perceptions of the subjects. Horizontal-plane results are on the left column and frontal-vertical plane results are on the right column. (Adopted from Blauert and Lindemann (1986))

The effect of frequency has been studied alongside with IACC. It was found that as the center frequency of bandpass noise was increased, the perceived width decreased (Blauert and Lindemann, 1986). This effect was the same with various IACC values. However, in a more recent research (Mason et al., 2005) the effect of frequency was slightly different. With IACC of 1 the perceived width was found to decrease as center frequency was increased up to 1600 Hz, but above this, the perceived width increased. Nevertheless, with low frequencies the perceived width was widest.

Briefly put, the perceived width increases as signal loudness level or duration is increased, and decreases as frequency is increased. Also, the correlation of the signals presented to the ears also affects width perception.

5.3 Perceived spatial width and distribution using loudspeakers

Loudspeakers have not been used as extensively in the studies on perceived spatial width and distribution. However, several such experiments exist. Many results of those studies are in connection with headphone experiments.

The effects of signal loudness level and frequency were studied using a vertical loudspeaker array consisting of five loudspeakers (Cabrera and Tilley, 2003). Loudness level was found to have a similar but smaller effect on perceived spatial width when using loudspeakers instead of headphones. The perceived width was larger with low frequency stimuli than with high frequency stimuli which is also consistent with headphone studies.

There are many real sound events that are perceived as being wide or spatially distributed, e.g. waves of the ocean, crowd, thunder and a flock of birds. It was shown that when such sounds were presented using virtual wide sound sources, they were perceived as being more natural than when the sounds were presented using point sources (Potard and Burnett, 2003). In another study, different loudspeaker configurations were preferred with different sound environments – a 3-D loudspeaker configuration was preferred when listening to indoor environments and a two-dimensional surrounding arrangement was preferred with outdoor environments (Guastavino and Katz, 2004).

Apparent source width has also been studied using virtual sound sources that were constructed using 16 loudspeakers (Potard and Burnett, 2004). Different virtual sound sources with various spatial extents, locations and geometry were tested using uncorrelated white noise. It was found that with some of the wide sound sources the perceived spatial width

was narrower than the intended width. This finding was most prominent with sound sources that were 60° wide. They suspect that this was due to source density being too high, as, according to them, it creates a narrower source. It is left undiscussed that the effect could be caused by the human's perception accuracy on wide sound sources. With 180° wide sound sources the perceptions were more varying – some were perceived as being 180° wide, some were perceived narrower, and some were even perceived as being somewhat surrounding. In the case of a single loudspeaker, the perception was not point-like but had some spatial width. In addition to horizontal width, also vertical extent was tested and the results imply that the perception of vertical sound sources was more inaccurate than that of horizontal sound sources. This result is in correlation with the knowledge of localization in general.

In the next sections, four studies that were done with loudspeakers are discussed more in-depth.

5.3.1 Effect of frequency

In 1984, a series of experiments were conducted where the discrimination of the spatial distribution was studied (Perrott, 1984). This was done using a loudspeaker array with 10 loudspeakers that were separated by 3.125° except that there was a gap of 6.25° in the center. Sound was always emitted simultaneously from two loudspeakers. The test subjects were first presented with a reference sound in which the emitting loudspeakers formed an angle of 18.75°. The task of the subjects was to report whether the second sound event was emitted from sources further apart or closer together than the reference. A few different effects of frequency on the performance of the discrimination task were studied.

First, the frequency difference of the signals from the two loudspeakers was altered. One of the frequencies was always 1000 Hz while the other was selected from six values between 1000 - 1300 Hz. The results indicated that there were most correct answers when the frequencies were 1000 Hz and 1030 Hz, i.e., when the difference was 3%. When there was no difference, the task was reported to be very challenging if not impossible. With a difference of 10%, the task was performed almost as well as with 3%, but when the difference was further increased to 30% the performance worsened.

As the results indicated the best performance at frequency difference of 3%, it was further analyzed whether the base frequency of the signal affected the performance. Four different base frequencies were selected and the frequency difference between the two signals was

always 3%. The results indicated that the performance worsened as the base frequency was increased. With a base frequency of 1500 Hz and particularly 2000 Hz, the discrimination task was performed poorly. It is concluded that based on this, interaural time differences may well be essential for the performance of this task when the frequency difference is small. Also, it can be said that the perception of two spatially distributed sound sources was affected by the frequency and the frequency difference of those signals.

Quite recently, the effect of frequency bands on perceived spatial width and direction perception was studied (Hirvonen and Pulkki, 2006b). The test setup was constructed inside an anechoic chamber and is illustrated in Figure 5.2. Eleven loudspeakers were placed in front of the listener at the height of the ears on a line that was approximately 2 m long and 90° wide when looking from the listening position. As the distances between loudspeakers and the listener were not equal, delays were used to compensate this so that all the sounds arrived at the listening position simultaneously.

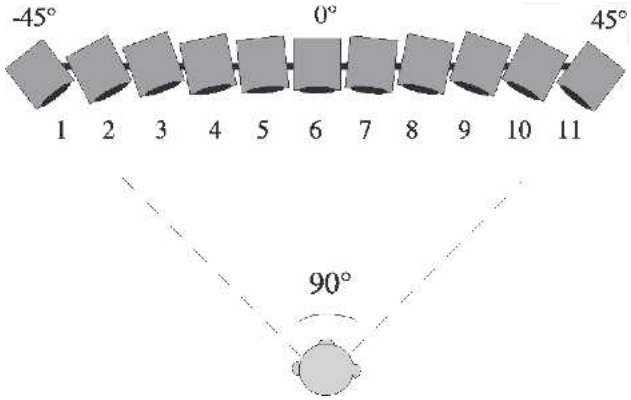

Figure 5.2: The loudspeaker setup of the listening test in Hirvonen and Pulkki (2006b). The setup was constructed inside an anechoic chamber. (Adopted from Hirvonen and Pulkki (2006b))

Gaussian noise and click train stimuli were used. The bandwidth of the noise signals was from 200 to 1179 Hz and they were divided into nonoverlapping frequency bands. The equivalent rectangular bandwidth (ERB, see Chapter 3.4) was used in the division and it resulted in 11 critical bands. The click train stimuli were 9 ms long and the interclick interval was 330 ms, whereas the noise signal was 60 seconds long continuous noise segment that had 100 ms fade-in and fade-out.

Test cases were created by routing the frequency bands into different loudspeakers. There were nine different cases, e.g. one where consecutive frequency bands were routed into

adjacent loudspeakers from left to right, other with all bands in the center loudspeaker, and third with a number of bands in the center loudspeaker and the rest of the bands in several loudspeakers on the sides. Using a keyboard, the subjects were asked to mark the loudspeakers that they perceived as emitting sound.

When examining the results, the main finding was that the high and low frequency bands affected the spatial perception more than the middle frequencies. In the cases where there was a large frequency change in adjacent loudspeakers the perceived width was slightly wider than in the other cases. Also, on average fewer than half of the loudspeakers were perceived as emitting sound even in the cases where all the loudspeakers emitted sound. This led to an implication that different ERB-bands were integrated together spatially. It can be said that a sound event that consists of many frequency bands was not perceived spatially as wide as it physically was – rather, the perception was narrower or discontinuous.

A preceding study to the above-mentioned one was made by the same authors (Hirvonen and Pulkki, 2006a). There, they used noise that was divided into frequency bands and emitted from nine loudspeakers that formed a 45° wide sound source. There were various frequency regions among the test cases – in the lowest case, the region was 100-640 Hz and in the highest case 2072-5858 Hz. The frequency regions were divided into nine ERB-bands so that there was one ERB-band in each loudspeaker. Also, cases with two and three ERB-bands in each loudspeaker were also tested. The order of the ERB-bands in the loudspeakers was varied from case to case. In the basic case, the ERB-bands were placed from the lowest to the highest from left to right, respectively. The other cases were formed from this by rotating the setup, i.e., shifting the ERB-bands one loudspeaker to the right. This way, adjacent frequency bands were in consecutive loudspeakers except for the discontinuity region, i.e., the place where the signals in two adjacent loudspeakers were wide apart in frequency because the highest and lowest ERB-bands were in them.

Most of the findings of the listening test in (Hirvonen and Pulkki, 2006a) were further analyzed in (Hirvonen and Pulkki, 2006b) and already discussed above, but the perceived center of the sound event was an interesting issue that was not investigated later in the same way. It was found that highest and lowest frequencies dominated the localization, as in the cases where there was one ERB-band in each loudspeaker the perceived center was very near to the discontinuity region. However, this did not happen with the highest frequency region – there, the perceived center was a little to the right from the actual center. In the cases with two and three ERB-bands the perceived center was close to the center of the loudspeaker group. Overall, the results of this experiment led to the same conclusion as the results of the above-mentioned experiment (Hirvonen and Pulkki, 2006b) – they indicated

that the lowest and highest frequencies dominated the perception.

5.3.2 Effect of signal length

Another recent research concentrated on the effects of signal length on perceived spatial width (Hirvonen and Pulkki, 2008). The article consists of two different experiments that are closely related to each other.

The test setup was similar in both experiments. In an anechoic chamber, nine loudspeakers were placed in the shape of a curve. The test subjects were placed in front of them so that all the loudspeakers were equidistant from the listener. This way all the signals arrived to the listening position simultaneously and thus the precedence effect could be excluded. The loudspeakers were visible to the listener, and the complete arrangement was in a 120° angle. The subjects were asked to mark those loudspeakers that they perceived as emitting sound and at the same time also to indicate the relative amounts of perceived sound from different directions. This was done using a touch-screen where each loudspeaker was represented by a slider.

In the first experiment, the test signals were 2.5 ms white noise bursts that were independently generated from Gaussian distribution. For each of the nine loudspeakers, there was a different signal of equal loudness. They were then simultaneously played from all the loudspeakers, and this resulted in a spatially wide sound source. In total, there were ten different cases.

The purpose of the first experiment was to find out the perceived spatial width of these short signals and whether or not there would be differences between the cases. The results indicated that there were notable differences. Some of the cases were perceived as point-like around the frontal direction, whereas some others were perceived wider. There were also differences between the perceptions of individual test subjects, usually such that the direction of the signal was perceived differently. An important finding was that the subjects were able to perceive varying localization cues even though the signals were very short. However, when analyzing the auditory width, the results showed that the overall perceived signal width was not very different between cases but rather they were generally perceived equally wide. The first experiment gave useful information for the second experiment.

The goal of the second experiment was to examine how length of the signal affects perceived spatial width. There were signals with lengths from 5 to 640 ms. Two cases from experiment one were selected to form the basis of the signals. The selected cases were the

two most different, one being perceived as spatially wide and the other as more point-like at frontal direction. Additionally, the perception of the former case was most inconsistent among test subjects and the perception of the latter was most consistent. The signals from these two cases were used as the beginning of all the longer signals. There were signals of eight different lengths, each one being double the length of the previous one, and because these signals were made from two different short signals, there were 16 test cases in total. As known, loudness level of the signal could have an effect on perception and therefore all the signals were adjusted to be equally loud. In the experiment, there were seven subjects that had also participated in the first experiment.

The results indicated that as signal length increased, the perceived spatial width clearly increased. Figure 5.3 shows the results for the signals that had the point-like 2.5 ms noise signal as the basis. The 5 ms case was perceived as being clearly more point-like than the rest of the cases. The perceived width increases significantly up to 40-80 ms and after that the distribution is close to a situation where all speakers are perceived to emit sound. It can be said that the differences between two sequential cases were not that significant, but when all sample lengths are considered, the relation between perceived width and sample length is obvious.

It can be seen from the results that the task was quite difficult. The subjects did not always answer consistently. Also, the differences between subjects were notable.

To conclude, the signal length affected the width perception. When the signal length was under 10 ms, the perception was mostly point-like. As the signal length was increased, the perceived spatial width also increased, but above 80 ms there were no more significant differences.

5.4 The perceived similarity of different spatial distributions

In this section, an article that reports a set of three listening tests is reviewed in detail. In the article, the minimum number of loudspeakers that are required to produce the spatial impression of a diffuse sound field was studied with a large loudspeaker setup (Hiyama et al., 2002). The goal was also to find an optimal arrangement of loudspeakers for that purpose. This was done so that they actually also tested how similarly sound events are perceived when either a large number of loudspeakers or different loudspeaker setups with less loudspeakers are used to produce them. In the field of this thesis, the last-mentioned is the most interesting issue.

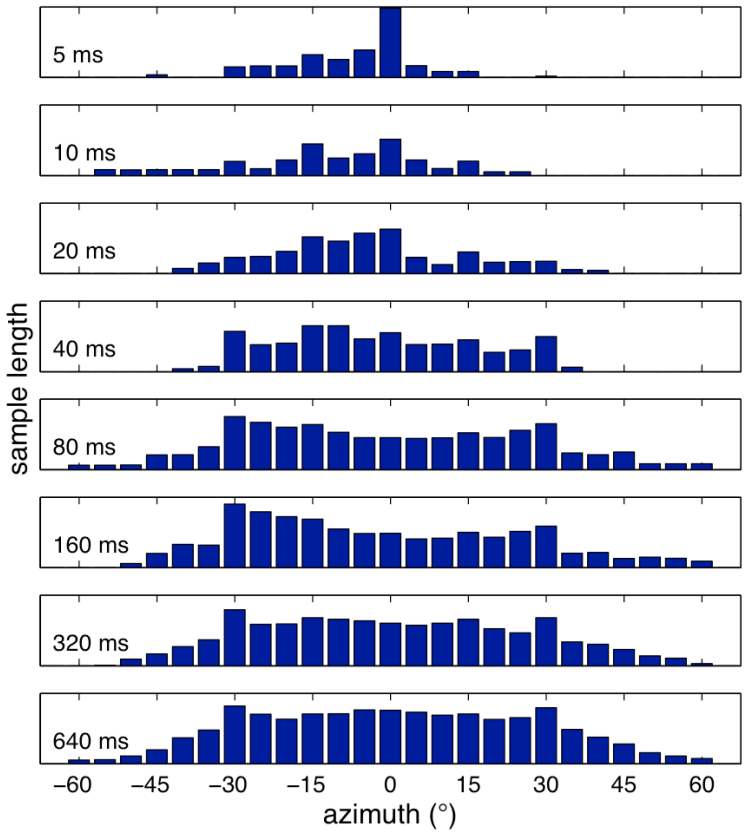

Figure 5.3: The results of an experiment where the effect of signal length on perceived spatial width was studied. There were eight samples with different lengths. The distributions of all responses are summed up to form the shown cumulative distributions. (Adopted from Hirvonen and Pulkki (2008))

The listening tests were conducted in an anechoic chamber where there were 24 loudspeakers placed in a circle around the listener. The loudspeakers were 15° apart from each other. White noise, band pass noise and musical sounds were used as the stimuli in three different experiments.

In the first experiment, uncorrelated white noise at equal loudness level was emitted from a number of loudspeakers. In the reference, 24 loudspeakers emitted sound, and this was compared to two stimuli – a hidden reference and the target stimulus that was emitted using 12, 8, 6, 4 or 3 loudspeakers that were placed at even intervals. The loudspeaker setups are presented in Figure 5.4. The black triangles indicate the loudspeakers that were emitting sound in each case. The subjects sat at the center of the setup facing towards the loudspeaker that is at twelve o'clock in the illustration. As can be seen in the figure, some of the cases

were arranged so that the loudspeaker directly at the front of the listener was used and some were not.

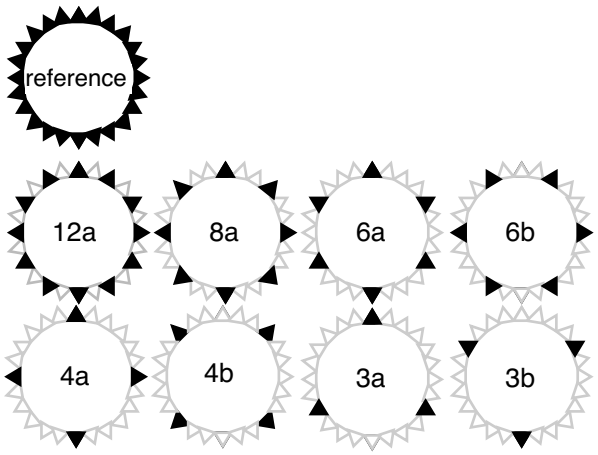

Figure 5.4: The loudspeaker setups that were used in the first experiment in (Hiyama et al., 2002). The loudspeakers that were emitting sound in each case are marked with a black triangle. The listener was placed at the center of the setup. (Adopted from Hiyama et al. (2002))

The subjects were asked to rate the difference of the reference compared to each of the two stimuli. The rating was done on a scale from 1.0 to 5.0 where 5.0 stands for the perception that the sound events were the same and 1.0 that the were very different. Then, the difference between the grades of the hidden reference and the target stimulus was calculated and thus a difference grade was obtained.

The results are presented in Figure 5.5. They indicated that the loudspeaker setups with 12, 8 and 6 loudspeakers were rated as being very similar with the reference, receiving difference grades between 0 and -0.7, whereas the setups with 4 and 3 loudspeakers were rated as being different from the reference, receiving difference grades between -2.2 and -3.2. These two groups were significantly different from each other and there was a substantial drop in the grade between the case with six loudspeakers and four loudspeakers.

Bandpass noise with three different bandwidths were used in the second experiment – namely, 0.1 - 1.8 kHz, 1.8 - 7 kHz and 7 - 20 kHz. There were a lot more loudspeaker setups than in the first experiment, many of them with uneven intervals between the loudspeakers.

Similarly as in the first experiment, the grades of the loudspeaker setups with 12, 8 and 6 loudspeakers were very high. With four loudspeakers, the placing of the loudspeakers

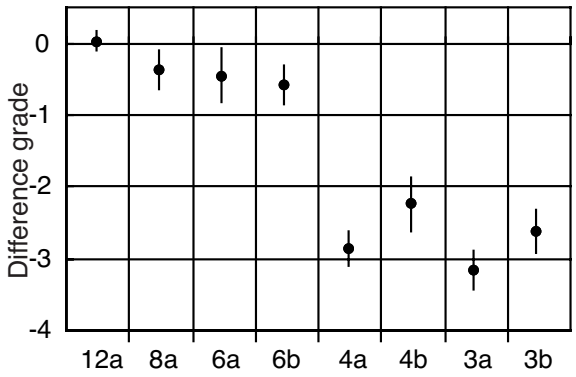

Figure 5.5: The results of the first listening test in (Hiyama et al., 2002). The loudspeaker setups in the x-axis are indicated by the same numbers as in Figure 5.4. (Adopted from Hiyama et al. (2002))

played an important role. There, the grades of the lowest bandwidth were as good as with eight or more loudspeakers. However, the grades were worse for the higher bandwidths whenever the loudspeakers were not placed so that there was a pair both at the front and at the back. There were two setups with four loudspeakers that received very good grades at all bandwidths. Those were setups where there were loudspeakers at 30° and at either 90° or 120° to the left and right from the front. There were also loudspeaker setups with three or two loudspeakers only. Overall, they received low grades.

Third experiment was similar to the second experiment except that musical instrument samples were used as the stimuli. It was tested whether the loudspeaker setups that performed well in the second experiment would perform equally well when the test stimuli consisted of direct sound, early reflections and reverberation. The results indicated that the loudspeaker setups that received highest grades were indeed the same in both tests.

The listening tests of the article resulted in a few conclusions. Uncorrelated white noise that was emitted from 24 loudspeakers was perceived as being very similar when it was emitted from only 12, 8 or 6 loudspeakers. In other words, not that many loudspeakers are needed in order to produce the same perception. When even less loudspeakers were used, the arrangement played a significant role. The conclusion regarding the minimum number of loudspeakers and their optimal arrangement was that four loudspeakers is optimal – a pair at the front 30° to the left and right and another pair at the rear with angles 120° to the left and right.

Chapter 6

The conducted listening tests

The main goal in this thesis is to study spatial hearing by conducting listening tests. In this chapter, the listening tests that were conducted are described in detail. The used methods and obtained results are presented and discussed.

Knowledge that is needed when conducting listening tests was presented in Chapter 4. This information was used in the planning of the listening tests so that problems would be avoided as much as possible and the whole process would be effective. Also, the work was done following the suitable standards and recommendations and therefore, the results should be reliable and generalizable.

Chapter 5 reviewed research that has been done on the topic of this thesis - studies that have concentrated on spatial sound, and particularly, spatial distribution perception. As was pointed out, not that many studies on the topic have been done using loudspeakers. Also, there are no studies – to the author's knowledge – on the resolution of perception of sound sources that are spatially distributed in complex ways.

There is always a desire to understand human hearing in a more detailed way. Nowadays, this is needed for example because the multichannel reproduction systems are developing and becoming increasingly more common e.g. in home theater use. More knowledge is needed on how wide sound sources and complex spatial distributions are perceived in order to reproduce them with multichannel systems so that humans perceive them as intended. The goal is that the results of the listening tests in this thesis could be used to improve the reproduction methods. Additionally, computer models may be developed and tuned based on the results so that they could predict how humans perceive certain sound events.

6.1 Resolution of spatial distribution perception

The first listening test that was conducted concentrated on spatial distribution perception. Particularly interesting was the resolution of this perception. Sound source width was a key variable in the test. First in this section, the relation of the experiment to the previous research is explained. Then, the used methods are described in detail, and finally, the results of the listening test are analyzed and discussed.

This experiment has close connections to a few preceding studies. In the experiments described in Chapter 5.3.1, width and distribution perception was studied concentrating on the effect of frequency (Hirvonen and Pulkki, 2006a), (Hirvonen and Pulkki, 2006b). This was done using noise that was divided into frequency bands that were emitted from various loudspeaker combinations. The principle in the studies is quite similar, but different from those experiments, the desire in this experiment was to study the perception using similar but uncorrelated sound in all the loudspeakers. This kind of approach was also used in the first two experiments of Hiyama et al. (2002) as described in Chapter 5.4. However, they had a loudspeaker setup that surrounded the listener, i.e., covered the horizontal plane of 360° and their aim was to find out the minimum number of loudspeakers that are needed to produce a perception of a surrounding sound. Thus, they used loudspeaker combinations in which the sound-emitting loudspeakers were not close to each other but distributed all around. Now, wide sound sources were especially of interest and therefore, loudspeaker combinations in which there were many sound-emitting loudspeakers very near each other were selected. Another fundamental difference was that there were loudspeakers only in the frontal half of the horizontal plane.

6.1.1 Experimental setup

The test was conducted in an anechoic chamber equipped with a multichannel reproduction system. The test setup is illustrated in Figure 6.1. There were 15 loudspeakers surrounding the subject in the horizontal plane so that they all were equidistant from the subject. When looking from the listening position, the loudspeakers were 15° apart from each other, thus covering the azimuth sector from -105° to 105° symmetrically at the height of the listener's ears. All loudspeakers were Genelec model 8030A active monitors.

Even though there were 15 loudspeakers, only 13 of them were actually used in the test to produce sound. The farthest ones on both sides were inactive. They were present in order to make it possible to register perceptions equally on any side of the actual sound source.

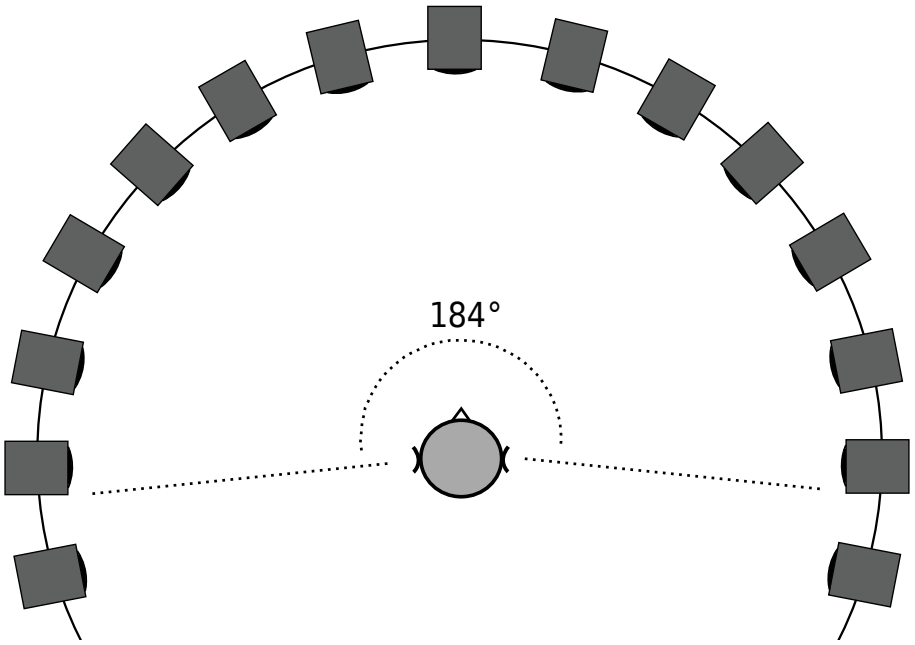

Figure 6.1: The loudspeaker setup used in the listening test. 15 loudspeakers were surrounding the subject equidistant in the horizontal plane, but only 13 of them emitted sound in the experiment. They formed a $184°$ wide sound source.

When all 13 loudspeakers were emitting sound they formed a physical sound source that was $184°$ in width. This is because the angle between the centers of the farthest loudspeakers was $180°$, and two halves of a loudspeaker added the $4°$. Each loudspeaker in this study is interpreted as a $15°$ wide element of a spatially distributed sound source, i.e., a gap of one loudspeaker corresponds to a gap of $15°$ – not $30°$.

6.1.2 Stimuli

The stimuli were pink noise that was randomly generated throughout the test instead of being similar in all cases. Each loudspeaker was driven with independent noise signals so that they were uncorrelated.

The total stimulus length was 1000 ms. There was a 100 ms linear fade-in and fade-out in the beginning and the end of the signal. In the middle, the signal remained at the same loudness level for 800 ms. This stimulus was constantly repeated every 2 seconds, i.e., 1 second stimulus and 1 second silence followed one another until the subject gave the answer.

In each test case, a selected combination of loudspeakers emitted sound at equal loudness level. All the test cases had equal loudness level as well, i.e., regardless of the number of loudspeakers emitting sound the overall loudness level was always the same at listener position.

6.1.3 Test design and research questions

A total of 21 different test cases, i.e., loudspeaker combinations were tested. They are presented in Figure 6.2. The combinations can be divided into four groups that focus on slightly different details. Such cases were selected in order to test the accuracy of perceiving fine details in wide sound sources. The test hypothesis was that the resolution of spatial distribution perception is not adequately high to accurately perceive these wide sounds sources and gaps in them.

The first group was a straightforward auditory width test, where the loudspeakers that emitted sound were symmetrically around the centre and the width was from 1 to 13 loudspeakers. These cases show how accurate width perception is and also if it is possible to perceive the sound source as a continuous wide sound event.

The second group tested the resolution of perception of a hole or a gap in a sound source, and additionally, the resolution of perceiving two separate wide sound sources. In each of these cases, the width of the sound source was 180°, but a number of loudspeakers in the centre were not active. Therefore, the cases could be thought either as having a hole in the centre or as two separate sound sources with altering width. With these cases it was possible to investigate how wide a hole has to be in order for it to be perceived, and also the accuracy of perceiving the width of a hole in a sound source.

The third group consisted of so-called chessboard-type combinations, where every other loudspeaker was emitting sound and every other not emitting. These combinations tested the perception of a complex distributed sound source. Particularly interesting was whether the perceptions of these cases would be significantly different or not. This, among with some of the other cases, indicates if the resolution of perception is adequately high in order to perceive such small spatial details in the sound source. Also included was a third chessboard-type combination in which there were sound sources and gaps that were 30° (two loudspeakers) wide.

The last group consisted of cases where there were two holes. Alternatively, the cases could be seen as three separate loudspeaker arrays. As in the case of the second group, the total

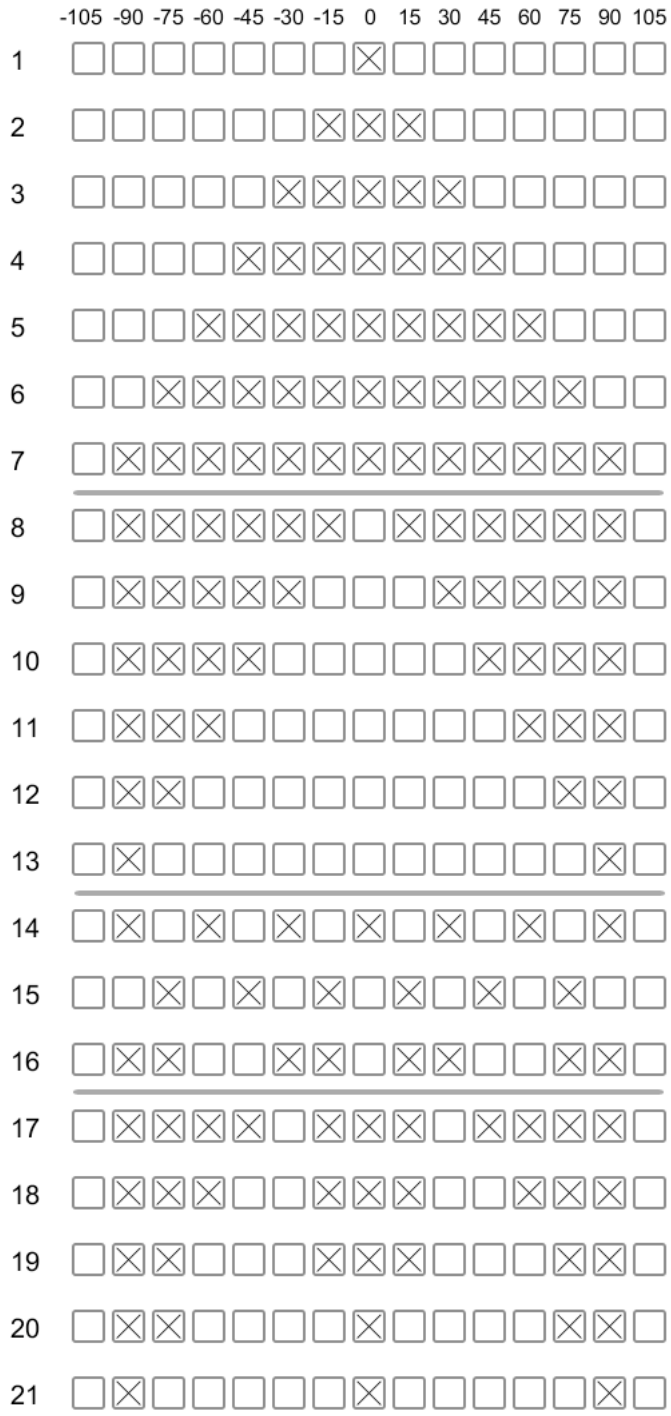

Figure 6.2: The 21 test cases of the listening test. The boxes marked with an 'X' represent the loudspeakers that were emitting sound in each case. The numbers on the top of the figure indicate the angles in which the loudspeakers were. The test cases can be divided into four groups as described in the text.

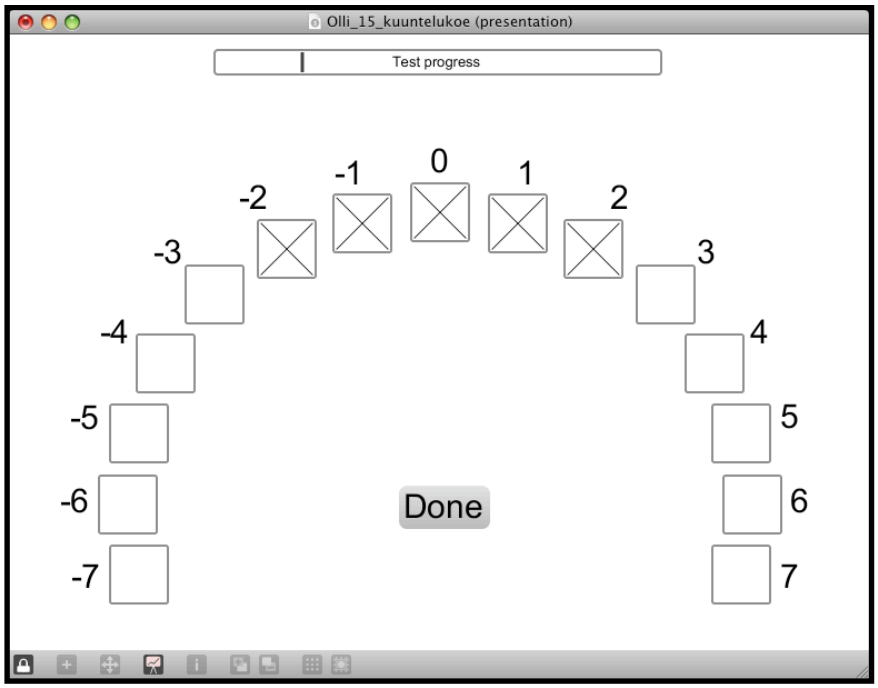

Figure 6.3: The graphical user interface of the listening test. Here, the subject has marked loudspeakers -2, -1, 0, 1 and 2 as emitting sound.

width in these cases was $180°$. These cases were included in order to test the perception of more complex sound sources.

6.1.4 Procedure

The subjects were told that there are 15 loudspeakers, and in each test case, any of them may emit sound so that any loudspeaker combination is possible. The subjects were instructed that they are allowed to rotate their heads freely but that they must not change their position or move otherwise. Moving was not allowed because it could have caused differences in distances between the listener and loudspeakers.

The task of the subjects was to identify which loudspeakers emit sound in each case, according to the subjects' own perception. Answers were registered with a touch-screen. The graphical user interface is presented in Figure 6.3. The boxes represent the loudspeakers, and the numbers from -7 to 7 that can be seen on the screen were also clearly visible on the loudspeakers. When answering, the subject checked all the boxes that corresponded to the loudspeakers that he perceived as emitting sound.

As mentioned, the test included 21 cases with different loudspeaker combinations. This was repeated twice for each test subject, and there was a short break between the test runs. In each run of the experiment, the presentation order of the stimuli was randomized to avoid biases caused by order effects. On average, the subjects completed the test in a little less than 30 minutes. The second run was typically completed slightly faster than the first run.

6.1.5 Test subjects

Ten voluntary subjects participated in the test. The author of this thesis did not participate in the test. All the subjects were staff or students in the Department of Signal Processing and Acoustics of Helsinki University of Technology. None reported any hearing defects.

6.1.6 Results

A histogram of the results is presented in Figure 6.4. All the 21 different test cases are presented in the same order as previously in Figure 6.2. The loudspeakers that were emitting sound are marked with a black box, and the gray bars represent the number of times the subjects marked the loudspeakers in question as emitting sound. The y-axis ranges up to 20 as there were ten subjects who each answered twice, thus making a total of 20 responses possible.

Statistical analysis

First, the results were analyzed using a statistical analysis method. This was done in order to see which cases have statistically different distributions and which have statistically similar distributions. It was also checked whether the distributions differed from a uniform distribution which would mean that the subjects had marked all loudspeakers with equal probability.

The used analysis method was Kolmogorov-Smirnov (K-S) goodness-of-fit test (Jr., 1951). In the test, two case distributions at a time are compared. The null hypothesis is that the two compared distributions are from the same continuous distribution. The test gives a p-value for the null hypothesis, and if this value is below a given limit, the null hypothesis is rejected. Usually, the limit is $p \leq 0.05$, and it is also used in this analysis. The higher the p-value, the more similar the two compared distributions are.

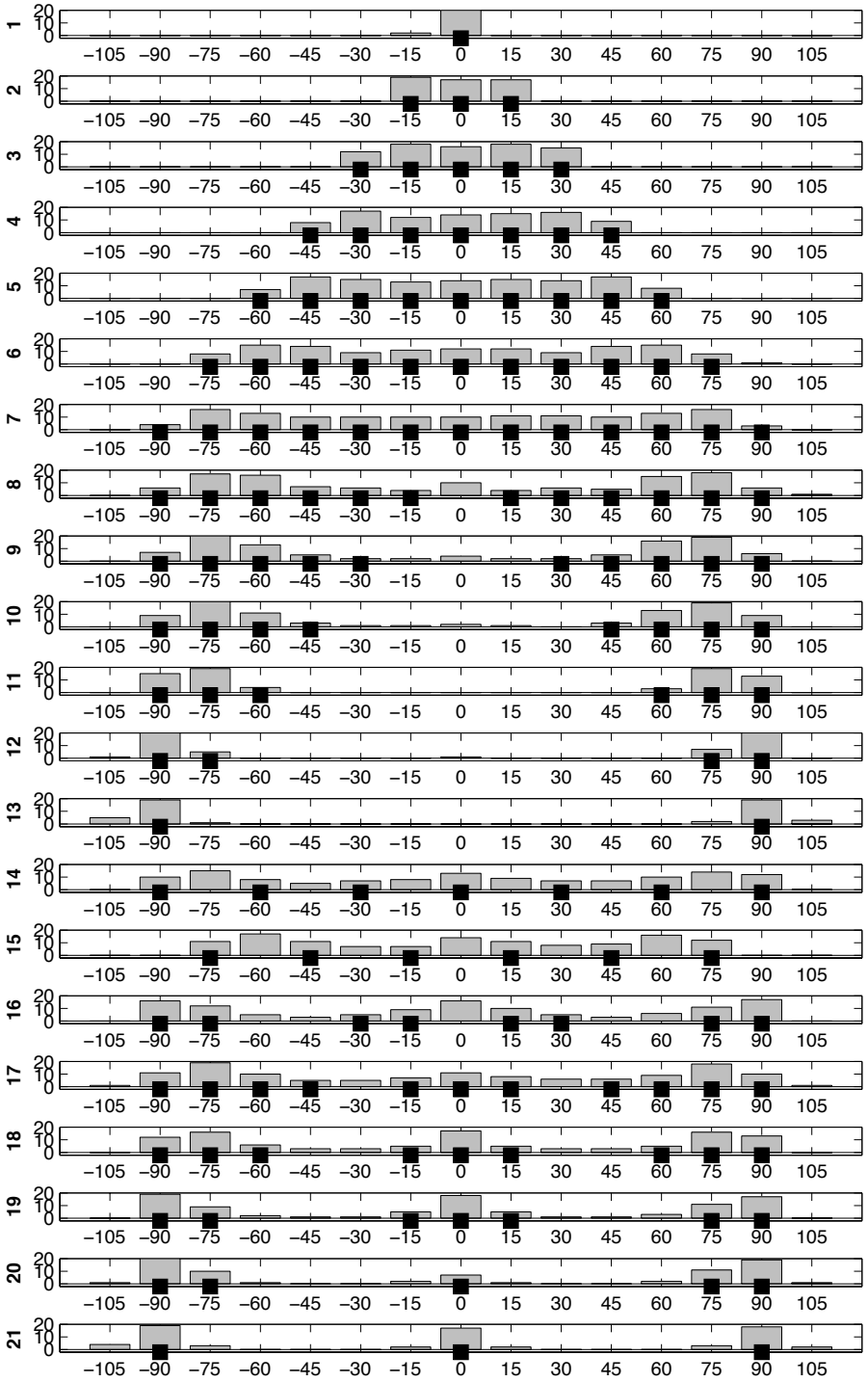

Figure 6.4: Results for all 21 test cases of the listening test presented in a histogram. The numbers from -105 to 105 indicate the angles in which the loudspeakers were. Black boxes indicate the loudspeakers that were emitting sound in each case. The height of the grey bars indicates how many times the particular loudspeaker was marked as emitting sound, 20 being the maximum.

The weighted histogram distributions of all the cases were compared to each other with the K-S test, making a total of 210 comparisons. The weighting was done so that each marked loudspeaker received a weighting coefficient depending on the total number of marked loudspeakers in that particular answer. This was done to all the subjects' answers in all cases.

The results of the K-S test are presented in Table 6.1. Upper right triangle shows the null hypothesis h-values - here, 1 means that the null hypothesis was rejected, i.e., the two distributions of the cases are not similar and 0 means that there is at least some similarity between the two distributions. This similarity is analyzed more accurately with the p-value which is presented in the lower left triangle of Table 6.1. As said, high p-value indicates high similarity.

It can be seen from upper right triangle in Table 6.1 that most case distributions are significantly different from each other, indicated by null hypothesis h-value of 1. This means that the subjects were able to discriminate most cases as being different from each other even though their perceptions were not similar to the actual sound source distributions.

The cases in which the K-S test indicates a lot of similarity are of particular interest. Highest p-values are between cases 6&15, 7&15, 9&10, 10&17, 12&13, 14&17, and 16&18. All these case similarities, particularly that between cases 6&15 and 7&15, strongly indicate that the resolution of distribution perception is not adequately high for the subjects to be able to perceive gaps of 15° in the sound source. It should be noted that the K-S test did not indicate that cases 14&15 were perceived similarly, but this is likely caused by the fact that the widths of those cases were different and not because the gaps in those cases would have been perceived.

The K-S test was also used to analyze if any distribution of a test case was close to a uniform distribution, i.e., a situation where subjects marked loudspeakers with equal probability. This hypothesis was rejected for all 21 test cases. This analysis result proves that subjects perceived something particular in every case and did not simply mark loudspeakers randomly.

Discussion

As mentioned earlier, the test cases can be divided into four groups. The first group - cases from 1 to 7 in Figure 6.4 - tested the auditory width perception. The cases with 1, 3 and 5 loudspeakers were perceived quite accurately, but already with 7 loudspeakers the furthermost loudspeakers were clearly less often marked. This phenomenon can also be

Table 6.1: The results of Kolmogorov-Smirnov (K-S) test for significance of differences. h-values are in the upper right triangle and p-values in the lower left triangle.

Case	1	2	3	4	5	6	7	8	9	10	11
1		1	1	1	1	1	1	1	1	1	1
2	0.00		1	1	1	1	1	1	1	1	1
3	0.00	0.01		1	1	1	1	1	1	1	1
4	0.00	0.00	0.00		1	1	1	1	1	1	1
5	0.00	0.00	0.00	0.00		1	1	1	1	1	1
6	0.00	0.00	0.00	0.00	0.01		0	1	1	1	1
7	0.00	0.00	0.00	0.00	0.00	0.06		1	1	1	1
8	0.00	0.00	0.00	0.00	0.00	0.00	0.00		0	1	1
9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.09		0	1
10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.84		1
11	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	
12	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
13	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
14	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.11	0.00	0.00	0.00
15	0.00	0.00	0.00	0.00	0.00	0.14	0.50	0.02	0.00	0.00	0.00
16	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
17	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.11	0.14	0.28	0.00
18	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
19	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
20	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
21	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Case	12	13	14	15	16	17	18	19	20	21
1	1	1	1	1	1	1	1	1	1	1
2	1	1	1	1	1	1	1	1	1	1
3	1	1	1	1	1	1	1	1	1	1
4	1	1	1	1	1	1	1	1	1	1
5	1	1	1	1	1	1	1	1	1	1
6	1	1	1	0	1	1	1	1	1	1
7	1	1	1	0	1	1	1	1	1	1
8	1	1	0	1	1	0	1	1	1	1
9	1	1	1	1	1	0	1	1	1	1
10	1	1	1	1	1	0	1	1	1	1
11	1	1	1	1	1	1	1	1	1	1
12		0	1	1	1	1	1	1	0	1
13	0.28		1	1	1	1	1	1	1	1
14	0.00	0.00		1	1	0	1	1	1	1
15	0.00	0.00	0.03		1	1	1	1	1	1
16	0.00	0.00	0.05	0.00		1	0	1	1	1
17	0.00	0.00	0.20	0.00	0.00		1	1	1	1
18	0.00	0.00	0.00	0.00	0.57	0.00		0	1	1
19	0.00	0.00	0.00	0.00	0.05	0.00	0.06		1	1
20	0.11	0.00	0.00	0.00	0.00	0.00	0.00	0.00		1
21	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	

seen with 9, 11 and 13 loudspeakers. A wide sound source was thus perceived as a little narrower than it was. Also of importance is that the middle loudspeakers were not marked as often as the loudspeakers in the edges on the far left and right. This indicates that it was challenging to tell whether all the loudspeakers between the perceived edges were emitting sound or not.

The second group - cases 8-13 - tested the resolution of perception of a gap in the sound source. Overall, it was difficult to accurately perceive a hole or its edges in the sound source. In case 8 when there was only one loudspeaker that was not emitting sound the results clearly show that the hole was not perceived - rather, the silent loudspeaker was marked more often than the ones next to it. On the other hand, the K-S test shows that cases 7 and 8 were statistically different from each other and additionally, the total number of marked loudspeakers was smaller in case 8, so the hole did have an effect on the perception.

Already in case 9 when the hole was $45°$ (three loudspeakers), the tendency was to perceive an even wider hole than it actually was. When the situation is thought to be a setup with two wide sound sources it can be said that a $75°$ wide sound source in case 9 was perceived a lot narrower as typically two or three loudspeakers were marked on both sides, corresponding to $30° - 45°$ in width. This is a different perception compared to case 3 where there was an equally wide sound source. The perception was different also in cases 2 and 11, where the width of the sound sources was $45°$. Thus, it can be said that the perception of a wide sound source is different when there are two of them presented simultaneously.

The third group - cases 14-16 - consisted of chessboard-type combinations. Here, the interesting part was to find out whether the subjects could perceive such complex sound source setups accurately or not, and also whether there were some differences compared to cases 6 and 7 where the total widths of the sound sources were similar to these cases.

When looking at cases 14 and 15, it is clear that the subjects' perception did not match the sound source composition. Rather, the subjects often marked the loudspeakers between the actual sound-emitting loudspeakers which indicates that such narrow gaps in a wide sound source could not be perceived. In case 16 the two $30°$ sound sources on the right and left were perceived quite well but the center area was inaccurate, as the center loudspeaker was often marked even though it was not emitting sound.

As noted before, the statistical analysis indicates similarity between the perceptions of cases 7&15 and 6&15. Even though only every other loudspeaker was emitting sound in case 15 and every loudspeaker in cases 6 and 7, the subjects perceived them quite similarly. It is clear that a gap of $15°$ was not perceivable.

Cases 17-21 tested the resolution of perception of three sound source groups - or alternatively, the perception of two holes. In all of these cases the three groups were perceived but the perceived width of the groups was mostly narrower than the actual width. Conversely, the holes were perceived wider than they actually were. Additionally, it should be noted that in case 20 it was difficult to perceive the sound source in the center. Statistical analysis shows that the perceptions of cases 16&18 were quite similar which again suggests that a gap or a change of 15° in the sound source could not be perceived.

These results lead to the conclusion that resolution of spatial distribution perception is not adequately high to be able to distinguish smallest details in these test cases.

Many subjects reported that they found it quite easy to find the edges of the sound event, i.e., the furthermost sound-emitting loudspeakers on the left and right. On the other hand, they reported that it was very difficult to discriminate whether the center loudspeakers were emitting sound or not. Some also reported that when they perceived sound in the center area, they most likely marked the loudspeaker 0 as emitting sound because they could not perceive the sound event more accurately. These comments are in correlation with the results except that the perceived and actual edges of a wide sound source were not necessarily in the same direction.

In some of the cases, the perceptions between subjects were quite different. This was perhaps most notable and most interesting with cases 7 and 14. The answers of all ten subjects for those cases are presented in Figure 6.5. As each subject answered twice, there are a total of 20 answers for both cases. On the left are answers for case 7 and on the right for case 14. Two consecutive answers are from the same subject, as are the answers on the same row on the left and right – e.g., the first and second answers in both cases are from one subject.

As can be seen in Figure 6.5, some of the subjects marked most of the loudspeakers as emitting sound whereas some marked only few loudspeakers. Same kind of behavior can be seen on both cases, usually with the same subjects. This suggests that in these cases, some subjects typically perceived a wide sound event whereas others perceived multiple distributed sound events.

This closer inspection of the cases 7 and 14 shows that they were indeed perceived as being quite similar. Even though there were different perceptions, each individual subject usually perceived both cases very similarly.

The average number of marked loudspeakers was, in most cases, smaller than the actual number of loudspeakers emitting sound. This data is presented in Table 6.2. The highest

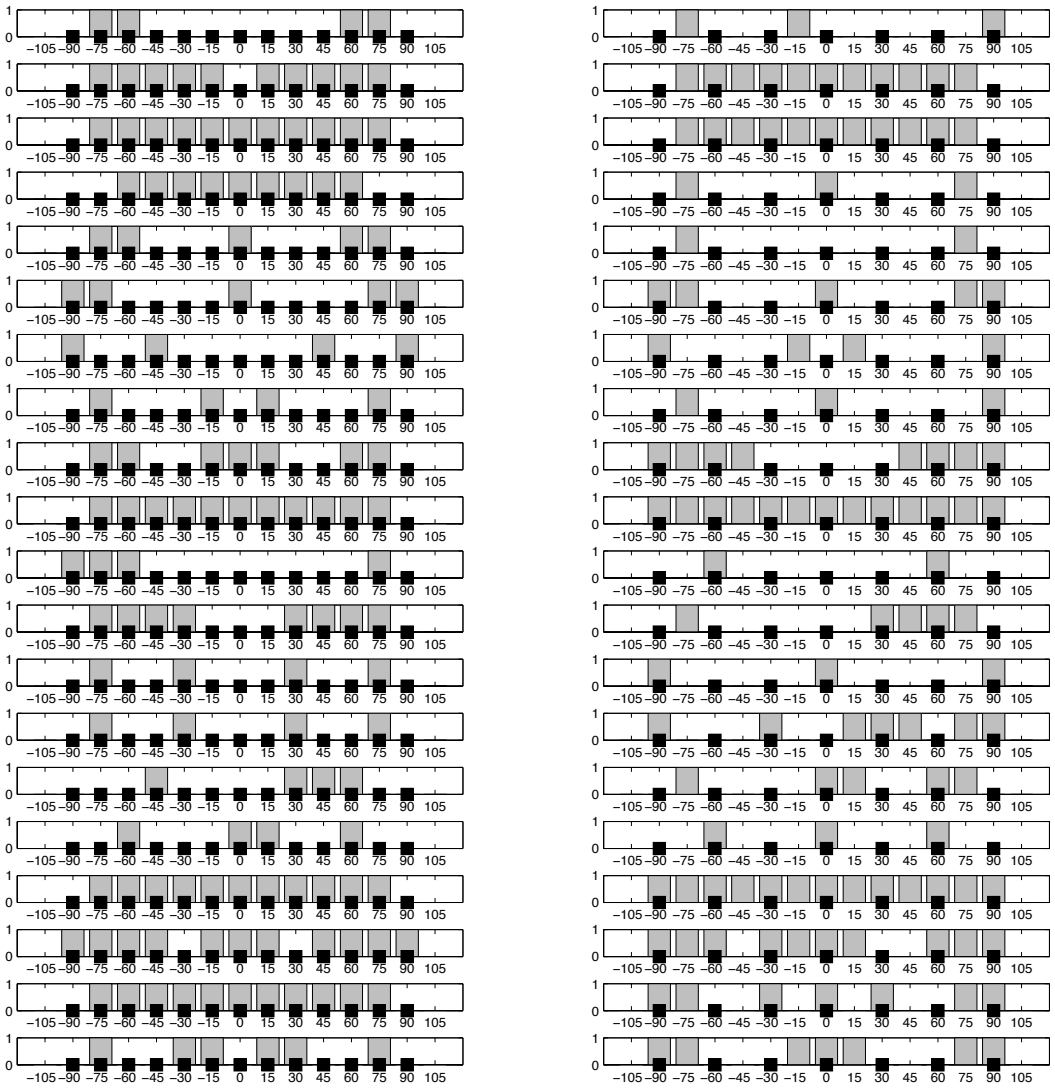

Figure 6.5: Results of cases 7 (left) and 14 (right). All 20 answers are plotted. The numbers from -105 to 105 indicate the angles in which the loudspeakers were. Black boxes indicate the loudspeaker that were emitting sound. The grey bars indicate that the particular loudspeaker was marked as emitting sound.

average number of marked loudspeakers was 6.85 in case 7 where the actual number was 13. In cases 2-7 there was one sound source with varying width, and there the average number was clearly smaller than the actual. There was a similar effect in most of the other cases as well. The opposite effect happened with cases 1, 13, 15 and 21 where there were single loudspeakers emitting sound simultaneously. In these cases the average number was more than the actual which was due to the fact that some subjects perceived the sound

events wider than just one loudspeaker. Deviation in the number of marked loudspeakers was greater in the cases where there were many loudspeakers emitting sound. However, it should be noted that the numbers in Table 6.2 do not directly indicate that the perceived width was smaller than the actual width. Rather, they indicate that not all loudspeakers that were emitting sound were perceived by the subjects.

Table 6.2: The actual number of loudspeakers, average number of marked loudspeakers, percentage, and standard deviation (SD) of marked loudspeakers for all 21 test cases.

Case	Actual	Marked	[%]	SD
1	1	1.10	110.0	0.31
2	3	2.65	88.3	0.49
3	5	3.95	79.0	1.10
4	7	4.55	65.0	1.61
5	9	6.00	66.7	2.27
6	11	6.40	58.2	2.78
7	13	6.85	52.7	3.03
8	12	6.05	50.4	3.19
9	10	5.15	51.5	2.96
10	8	4.60	57.5	2.35
11	6	3.65	60.8	1.35
12	4	2.70	67.5	0.92
13	2	2.45	122.5	0.83
14	7	6.25	89.3	3.65
15	6	6.15	102.5	2.48
16	8	5.90	73.8	3.32
17	11	6.35	57.7	4.03
18	9	5.35	59.4	3.21
19	7	4.65	66.4	2.35
20	5	3.75	75.0	1.48
21	3	3.50	116.7	1.32

6.1.7 Conclusions

The experiment resulted in a number of conclusions. First, a gap of 15° (one loudspeaker) in the sound source could not be accurately perceived when the sound source was wide. Even though the subjects' perception changed when there is a gap, the actual location of the gap was not perceived. Second, larger gaps were perceived wider than they were. Conversely, multiple wide sound sources were perceived slightly narrower than they were. Third, very wide sound sources were also typically perceived a little narrower than they actually were.

Finally, the average number of loudspeakers that subjects perceive as emitting sound was smaller than the actual number. All in all, based on these results it can be concluded that the finest detail that a human can perceive is of order of $15^\circ - 30^\circ$ when the sound source is wide.

6.2 Discrimination of spatially distributed sound sources

The purpose of the second listening test was to find out more about the resolution of perception of a spatially distributed sound source. The first listening test indicated that the resolution of perception was not adequately high for perceiving the smallest details of the sound source distributions in the test. Consequently, it was desired to further investigate the phenomenon and find out what kind of spatial distributions of a sound source could be perceived, and moreover, the point at which the spatially distributed sound source could no longer be perceived accurately.

In addition, the effect of bandwidth was also studied. This was done in order to find out whether wide-band noise was easier to perceive than narrow-band noise.

6.2.1 Experimental setup

The experimental setup in the test was the same as in the first listening test (illustrated in Figure 6.1) with the exception that this time there were only 13 loudspeakers surrounding the listener. However, as the loudspeakers that were farthest on the left and right were not actually emitting sound in the first test, the setup can be considered to be similar in both tests.

6.2.2 Stimuli

A total of 13 different samples were used. They can be divided into three groups: bandpass filtered noise, pink noise and sine wave samples. All samples were 1000 ms in length - a 100 ms fade-in, 800 ms constant loudness and a 100 ms fade-out.

The main focus in the test was on the bandpass filtered noise cases. Two center frequencies, 500 Hz and 4000 Hz, and five filter bandwidths, 1, 1/3, 1/8, 1/12 and 1/24 octaves in width, were selected as the parameters. This resulted in 10 different samples. Each sample had 13

channels which were independently generated from Gaussian distribution, thus resulting in uncorrelated channels. These signals were then bandpass filtered, the fade-in and fade-out were applied, and the energies were normalized to be similar in all cases. The normalization resulted in aligned loudness levels between the samples.

Pink noise was also used in the first listening test, and therefore, it can be considered to be a link between the first and second listening test. Like the bandpass filtered noise cases, also pink noise had 13 uncorrelated channels.

There were two different sine wave cases where sine waves with seven different frequencies were selected so that the frequency range was similar to the octave-band noise cases, i.e., from 353.5 to 707.1 Hz and from 2828.4 to 5656.9 Hz. Inside these ranges, the frequencies of the sines were calculated with an equation

$$f_x = f_l * 2^{x/6} \tag{6.1}$$

where f_l is the lowest frequency, i.e., 353.5 Hz or 2828.4 Hz, and integer x runs from 0 to 6. Thus, the sines were separated with equally large intervals.

Five loudspeaker setups were selected, with 2, 3, 4, 5 and 7 loudspeakers emitting sound. All setups included the leftmost and rightmost loudspeakers, and the others were selected so that they were evenly divided, i.e., each loudspeaker was equally far from one another in the pattern. It can be said that as the number of loudspeakers increased, the density of the loudspeaker setup increased. Four different pairs were formed from the five loudspeaker setups: 2&3, 3&4, 4&5 and 5&7. These pairs and their loudspeaker arrangements are presented in Figure 6.6.

As there always had to be seven sine waves in the sine wave cases, they were divided to the loudspeakers depending on the setup. For example, with three loudspeakers the sines were divided into those so that there were two, three and two sines in the loudspeakers.

6.2.3 Procedure

In the test, the subjects were first presented with illustrations of two different loudspeaker setups. Then, after pressing 'play', they heard a sample as played back with those two setups. The task of the subjects was to discriminate which of the two loudspeaker setups was used in the latter sound event, i.e., the task was a two alternative forced choice proce-

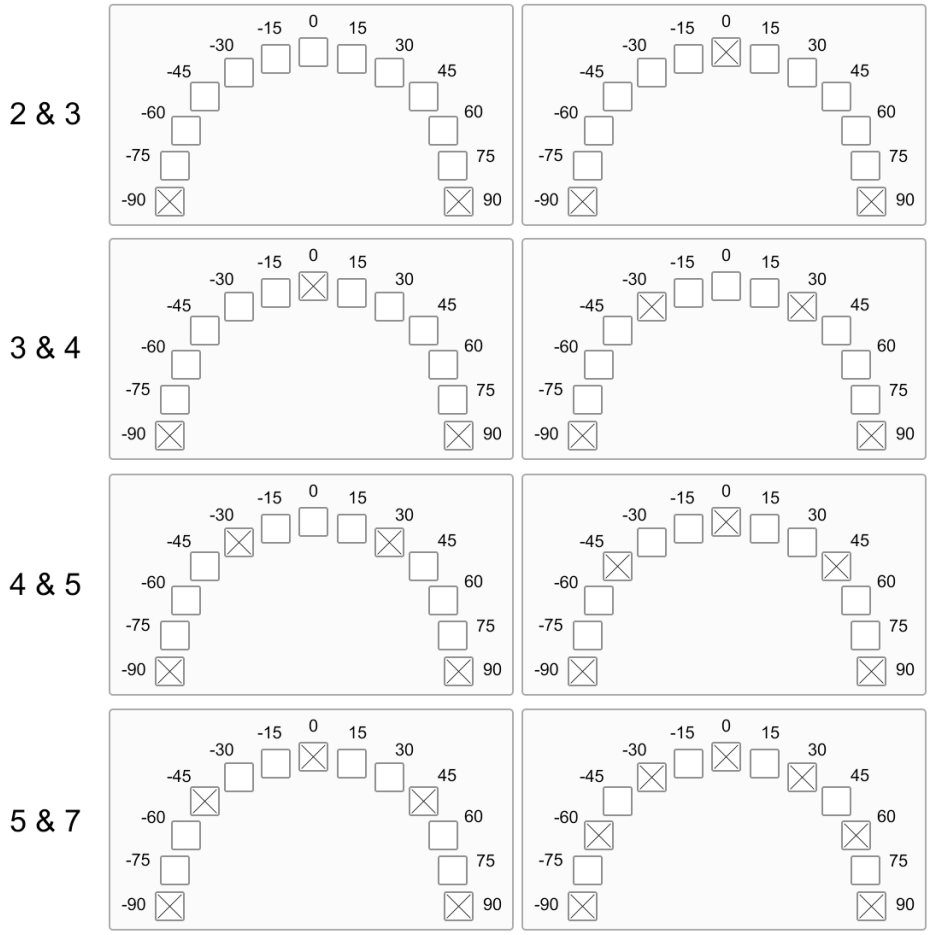

Figure 6.6: The four loudspeaker setup pairs of the second listening test. The boxes marked with an 'X' represent the loudspeakers that were emitting sound. The numbers by the boxes indicate the angles in which the loudspeakers were.

dure. The two sound events were played back twice with a short pause between them – as illustrated in Figure 6.7 – and the subjects were forced to listen to this whole chain before answering. It was not possible to listen to the chain again. The subjects were instructed to look at the center loudspeaker (clearly marked with number 0) and remain still when the sounds were playing.

The graphical user interface is presented in Figure 6.8. There, the illustrations of the loudspeaker setups can be seen in the middle. The numbers in the illustrations are labels of the loudspeakers, and they were also visible on the actual loudspeakers. The large buttons correspond to the loudspeaker setups above them and were used for the answering.

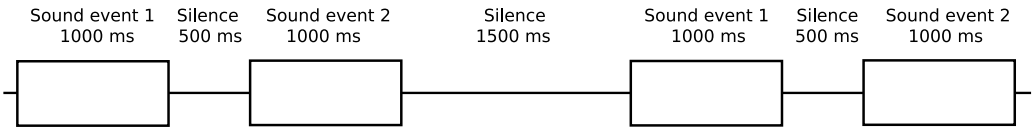

Figure 6.7: Timeline presentation of one test case. The durations of sound events and silences are in milliseconds.

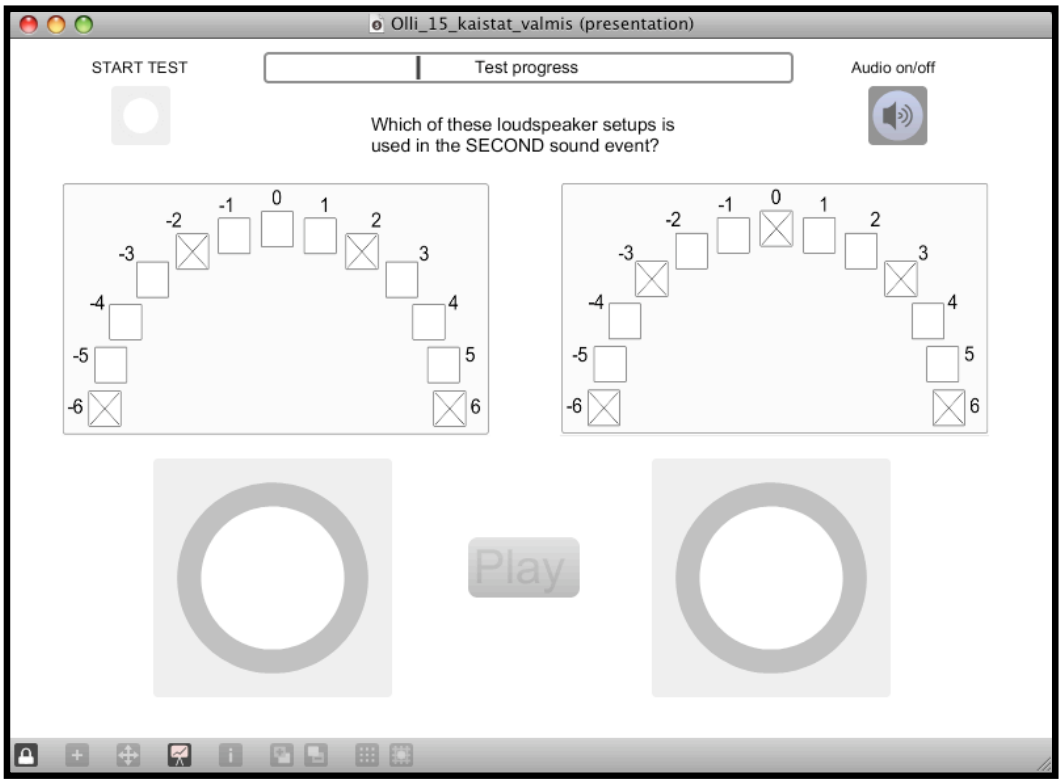

Figure 6.8: The graphical user interface of the second listening test. Here, the loudspeaker setup pair that is being presented to the test subject is four versus five loudspeakers.

The total number of test cases was 52 as there were 13 different stimuli and four loudspeaker setup pairs. Each subject heard each test case four times so the whole test had 208 test cases. There was a short training session prior to the actual test in order to make the user interface and the test procedure familiar to the subjects. The actual test was divided into four sections and between them was a short break. It took approximately 75 minutes for the subjects to complete the whole test.

6.2.4 Test hypotheses

Two main hypotheses are stated. First, the perception of spatially distributed sound sources becomes more challenging as the density of the sound source group is increased. Second, the bandwidth affects in such a way that the perception becomes more challenging as the bandwidth becomes narrower.

The cases with a center frequency of 500 Hz are in the frequency region where interaural time difference (ITD) dominates localization, whereas the cases with a center frequency of 4000 Hz are in the frequency region where interaural level difference (ILD) dominates localization. This way, the effects of different localization cues can be studied.

6.2.5 Test subjects

A total of 19 subjects participated in the test. Again, they all were staff or students in the Department of Signal Processing and Acoustics of Helsinki University of Technology, and the author of this thesis did not participate in the test. Some of the subjects had also participated in the first listening test. Almost half of the subjects had never before participated in any listening test, so they can be said to be naive listeners. None reported any hearing defects.

6.2.6 Results and analysis

The combined results and their 95% confidence intervals for all the subjects are presented in Figure 6.9. There, the cases are divided into three subfigures: topmost and in the middle are the cases with center frequency of 500 Hz and 4000 Hz, respectively, and at the bottom are pink noise and sine wave cases. The bandwidth of the bandpass filtered noise signals is one octave on the left and reduces when going to the right. The different loudspeaker setup pairs are presented with individual lines in all subfigures. However, with the pink noise and sine wave cases the connections are drawn only with dashed lines because there are no values between them and only the real data points and their confidence intervals should be examined. Also, it should be noted that the points of each sample are horizontally misaligned only for easier visual inspection – bandwidth was always the same inside each group.

As the task was a two alternative forced choice procedure, the probability of answering

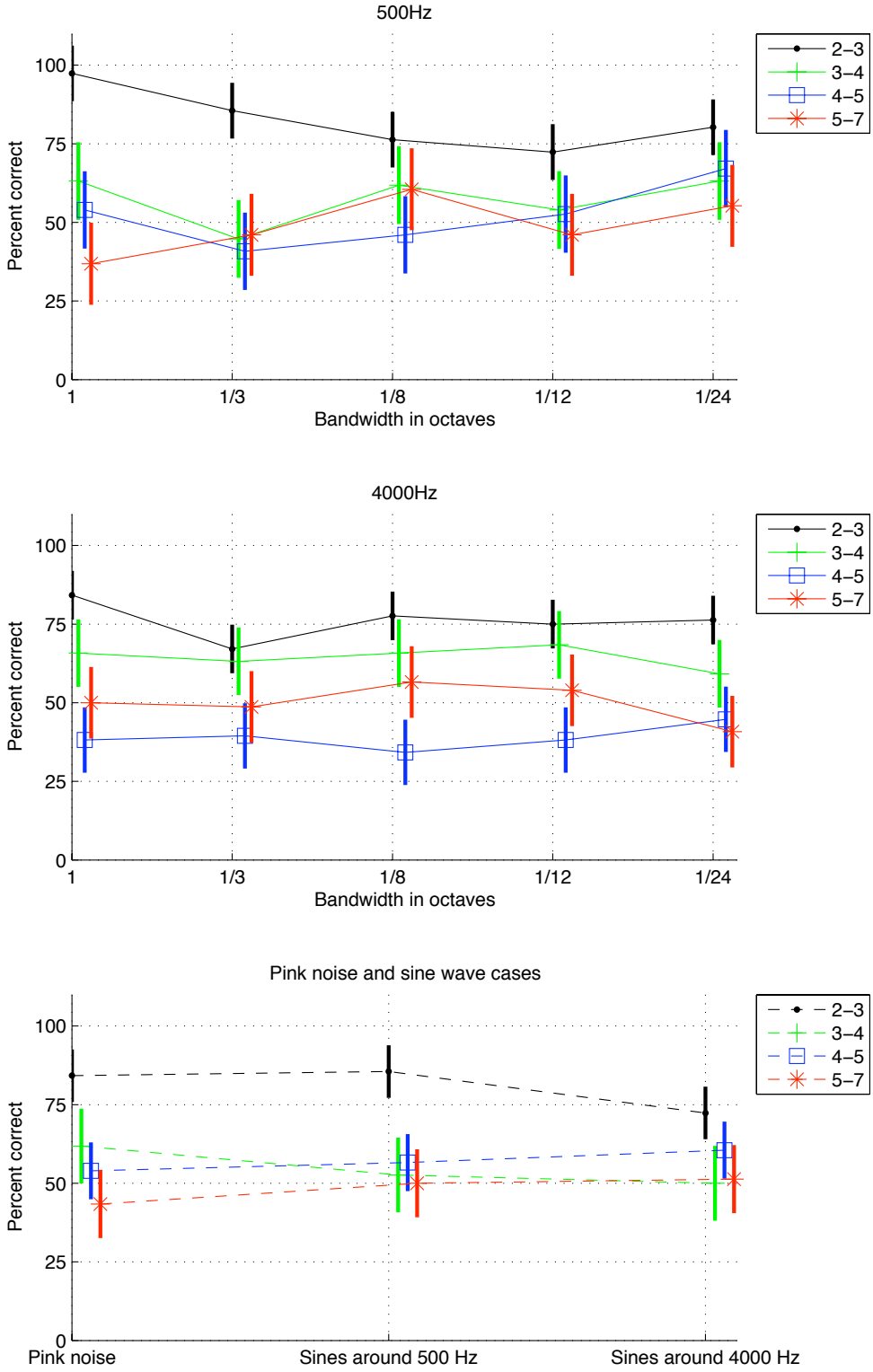

Figure 6.9: Results of the second listening test. The percentage of correct responses in each case is presented in the y-axis. The different loudspeaker setups are presented with individual lines.

correctly by pure chance is 50%. Commonly, the threshold for determining whether the answers were correct by knowledge or chance is 75%. In other words, any case where the percentage of correct answers is less than 75% cannot be said with confidence to be perceived correctly.

The results were statistically analyzed using a three-way analysis of variance (ANOVA) (Bech and Zacharov, 2006). It was used to test which of the test variables – sample, loudspeaker setup pair and test subjects – had a statistically significant effect on the results. Also, the effect of interaction between the variables was tested. The results of the ANOVA are presented in Table 6.3. Typically, the variable is said to have a significant effect on the results if $p \leq 0.05$, and that is also used in this analysis. In addition, a large F-value indicates a significant effect. It can be seen from the table that loudspeaker setup pair clearly had the most statistically significant effect. Also test subject had a statistically significant effect, but the sample did not. The effect of interaction between the variables was significant only with the combination of sample and loudspeaker setup pair.

Table 6.3: Results of the three-way analysis of variance (ANOVA). 'Subject' stands for test subject and 'loudspeaker' stands for loudspeaker setup pair. The p-value indicates the significance of the variable.

Variable	**Type III sum of squares**	df	F**-value**	p**-value**
Subject	1.859	18	1.659	0.042
Sample	1.116	12	1.493	0.122
Loudspeaker	15.711	3	84.105	0
Subject x Sample	12.42	216	0.923	0.756
Subject x Loudspeaker	4.261	54	1.267	0.101
Sample x Loudspeaker	4.712	36	2.102	0
Error	40.348	648		
Total	425.625	988		

The effect of loudspeaker setup pair

As can be seen in Figure 6.9, the cases with 2&3 loudspeakers were perceived most accurately with all 13 samples. In most of these cases, the proportion of correct answers was above the threshold of 75%. In contrast, the perception accuracy of all the other loudspeaker cases was always below the threshold of 75%. However, the confidence intervals of the cases with 2&3 loudspeakers stay above 75% only with five of the samples. This means that only those five cases were certainly perceived correctly. Nevertheless, it can be said

that the results suggest that the first hypothesis is true – increasing the density of the sound source group made the perception more difficult. This is supported by the ANOVA results for they indicate that loudspeaker setup pair had a significant effect on the results.

According to Tukey's Honest significant difference test, the loudspeaker setup pairs can be divided into three groups that are significantly different from each other: 2&3 loudspeakers in the first, 3&4 loudspeakers in the second, and 4&5 and 5&7 in the third group. The mean values and 95% confidence intervals for the loudspeaker setup pairs are presented in Table 6.4. Only the pair with 2&3 loudspeakers has the mean and confidence interval above the threshold of 75%.

Table 6.4: The mean values and 95% confidence intervals for the loudspeaker setup pairs.

Loudspeaker setup pair	**Mean**	**Std. error**	**95% confidence interval Lower bound**	**95% confidence interval Upper bound**
2&3	0.796	0.016	0.764	0.827
3&4	0.595	0.016	0.564	0.626
4&5	0.482	0.016	0.451	0.513
5&7	0.492	0.016	0.461	0.523

The effect of bandwidth

The effect of bandwidth was clearest in the region where ITD dominates localization, i.e., with the center frequency of 500 Hz. There, the results of the cases with 2&3 loudspeakers show that as bandwidth became narrower, the perception accuracy decreased. Only with the two widest bandwidths the confidence intervals were above the threshold of 75%. At the narrowest bandwidth, 1/24 octave band, the accuracy was better than with 1/8 or 1/12 octave bands, but with all these three cases the confidence intervals were under 75%.

In the ILD region, i.e., with the center frequency of 4000 Hz, the effect of bandwidth was not clear. Only the case with 2&3 loudspeakers was correctly perceived with octave band noise around 4000 Hz. The results of all the other 19 cases with the center frequency of 4000 Hz were under the threshold of 75%, and therefore, it is deemed irrelevant to examine the effect of bandwidth in those cases.

When the inspection is bounded only to the cases with the center frequency of 500 Hz, the ANOVA results are slightly different than when all cases were included. They are presented in Table 6.5. When these ANOVA results are compared to those for the whole data, it can be seen that now, besides the loudspeaker setup pair, also the sample (i.e., bandwidth) had

a significant effect on the results.

Table 6.5: Results of the three-way analysis of variance (ANOVA) for the cases with the center frequency of 500 Hz.

Variable	**Type III sum of squares**	df	F**-value**	p**-value**
Subject	1.214	18	0.985	0.479
Sample	0.742	4	2.707	0.031
Loudspeaker	6.57	3	31.967	0
Subject x Sample	4.483	72	0.909	0.677
Subject x Loudspeaker	3.577	54	0.967	0.545
Sample x Loudspeaker	1.852	12	2.252	0.011
Error	14.798	216		
Total	170.938	380		

The effect of bandwidth was separately analyzed using the cases with the center frequency of 500 Hz and loudspeaker setup pair with 2&3 loudspeakers. This is justified due to the ANOVA results that indicated that the loudspeaker setup pair in question was significantly different from the others and that bandwidth had an effect on the results with the center frequency of 500 Hz.

Two-way ANOVA was performed to these cases and the results are presented in Table 6.6. Now, both the sample (i.e., bandwidth) and the test subject had a significant effect on the results. This indicates that when the percentage of correct answers was above the threshold of hearing, the bandwidth of the noise signal did have a statistically significant effect on the perception.

Figure 6.10 presents the multiple comparison results for the cases with the center frequency of 500 Hz and loudspeaker setup pair with 2&3 loudspeakers. According to Tukey's Honest significant difference test, the case with one octave bandwidth is significantly different from 1/8 and 1/12 octave bands. On the other hand, the 1/3 octave band is not significantly different from any other case. It can be said that the bandwidth did have some effect on the results. However, it is left unexplained at this point why the 1/24 octave band noise was perceived more accurately than two of the wider bandwidths. This phenomenon will require further studies to be understood better.

Table 6.6: Results of the two-way analysis of variance (ANOVA) for the cases with the center frequency of 500 Hz and loudspeaker setup pair with 2&3 loudspeakers.

Variable	**Type III sum of squares**	df	F**-value**	p**-value**
Sample	0.71447	4	4.7	0.002
Test subject	1.28421	18	1.88	0.0318
Error	2.73553	72		
Total	4.73421	94		

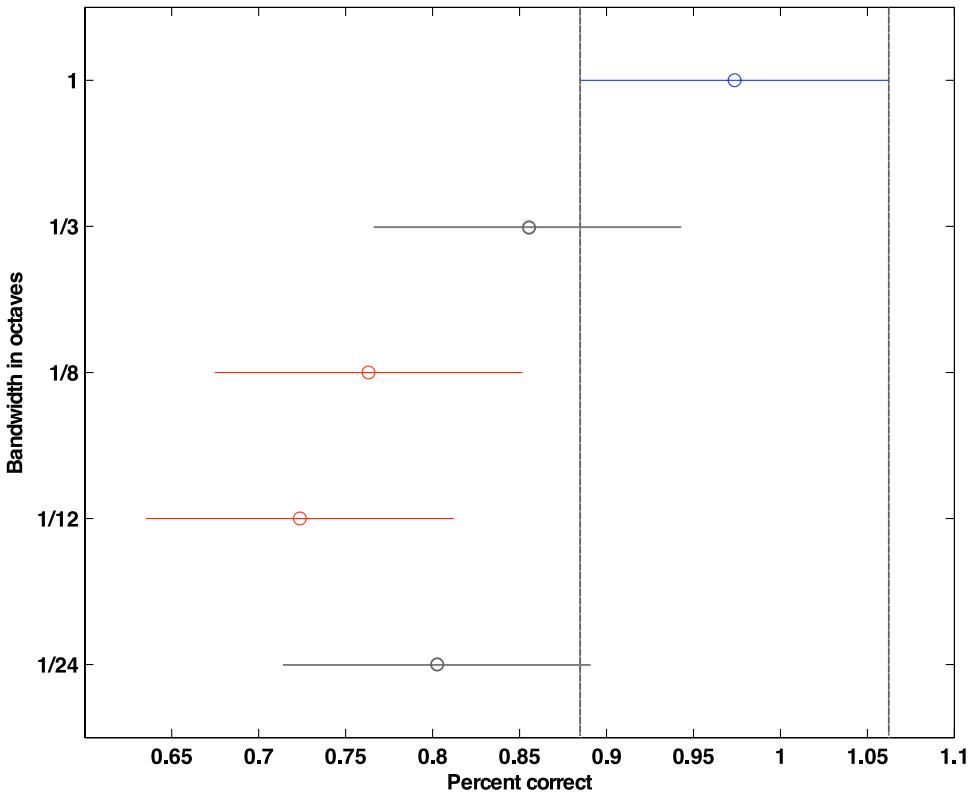

Figure 6.10: The perception accuracy and confidence intervals only for the cases with the center frequency of 500 Hz and loudspeaker setup pair with 2&3 loudspeakers. Tukey's Honest significant difference test indicates that one octave band case (topmost, denoted with a blue line) is significantly different from two of the other cases (denoted with red lines).

6.2.7 Discussion

Overall, it can be said that this discrimination task presented to the subjects was challenging. The results were less informative as was desired – the effect of bandwidth was not as

clear as was hypothesized and the perception accuracy for the cases with 3&4, 4&5 and 5&7 loudspeakers was below the threshold. However, the latter result indicates that the resolution of spatial distribution perception was even worse in this listening test than in the first listening test.

The sine wave cases provide an interesting additional aspect to the results. As the frequency range in the sine wave cases was the same as with the octave band noise cases, the results of those can be compared to each other. The results in the ITD region, i.e., the cases with the center frequency of 500 Hz, are rather similar – the case with 2&3 loudspeakers was perceived both with the octave band noise and with the sine wave case whereas the other loudspeaker pairs were not. In the ILD region, i.e., the cases with the center frequency of 4000 Hz, the results look different. Even though the case with 2&3 loudspeakers was perceived with octave band noise it was not perceived with the sine wave case. This suggests that the frequency band is not the only attribute that affects the perception of these samples.

The bandpass signals can each be thought to be equivalent to a sine wave that is modulated with a noise signal with a certain spectrum. It is assumed that the modulation causes differences in the signal so that when several loudspeakers emit uncorrelated signals, different loudspeakers become more audible, i.e., stand out at different times. With the octave band noise this effect is more clear, thus making the perception easier. When the bandwidth becomes narrower the effect is diminished and the perception becomes more difficult. The situation is similar with the sine wave cases – there is no such modulation aid with pure sine waves that are not close to each other in frequency and the perception is more difficult than with the octave band noise.

With pink noise the results were close to the octave band noise cases – again, the case with 2&3 loudspeakers was perceived but the others were not. The percentage of correct answers in the case with 2&3 loudspeakers was slightly better with the octave band around 500 Hz than with pink noise.

Remarks of the test subjects

The subjects typically reported that the cases with 2&3 loudspeakers were the easiest to discriminate whereas the others were more difficult. A few subjects noted that they found the case with 4&5 loudspeakers more difficult than 5&7. These subjects reported that in the latter case, one sound event was more surrounding and full than the other sound event, and they linked the alternative with seven loudspeakers with the surrounding sound event.

These comments are in correlation with the results.

As for the different samples, subjects reported that the cases with more noise-like sounds, i.e., the octave-band noises and pink noise, were easier to perceive than others. In contrast, the narrow-band samples were reported to be very challenging. In the results this effect can be seen only in the cases with 2&3 loudspeakers and center frequency of 500 Hz.

Some subjects noted that higher frequency samples were generally more challenging. This could partially explain the low ratings of the samples with 4000 Hz.

Many subjects also reported that even though they perceived differences in the two sound events they found it difficult to decide which loudspeaker setup was used in the second sound event. If there had been a learning session prior to the test this problem could have been reduced to some degree because the subjects could have recognized sound events that they had heard. Such session was not added because more intuitive and neutral perceptions were desired, as they give more information about natural situations in which the perception is spontaneous.

One subject also noted that some of the differences were so insignificant that if the loudspeaker setups were presented individually, i.e., not in pairs with just a short time gap between them, the differences would have gone unnoticed and the perceptions would have been similar.

6.2.8 Conclusions

The perception of spatially distributed sound sources became more difficult as the density of the sound source group was increased. Already the case with 3&4 loudspeakers was hard to perceive correctly. This effect was independent of stimulus type and was supported by statistical analysis. Thus, the resolution of spatial distribution perception was rather poor. When examining the cases that were perceived the most accurately, the bandwidth of the noise signals did have some effect on the results since the perception accuracy was decreased when bandwidth was narrowed. It is assumed that the modulation in the noise signals of wider bandwidths made different loudspeakers stand out at certain moments, thus making the perception easier than with narrower bandwidths. This effect is supported by the fact that the cases with sine waves in the ILD region in which there are no such modulation aids were also more inaccurately perceived.

Chapter 7

Conclusions and Future Work

The aim of this thesis was to study the perception of spatial sound, and the main emphasis was on the conducted listening tests. First, a brief introduction on the theme of the thesis was given, and thereafter theory on sound and psychoacoustical phenomena was discussed.

When it comes to listening tests, extensive planning and preparation is essential for success. Therefore, some traditional and commonly accepted methods for listening test measurements were presented. A few general recommendations that should be followed were also introduced. Then, research on the perception of spatially distributed sound sources and wide sound sources was discussed. Many previous studies have been made using headphones in which case the perceptions tend to be localized inside the head. Despite that, the results from those studies are in correlation with studies that have been made using loudspeakers.

The two listening tests were conducted in order to obtain knowledge about the spatial distribution perception. In the first listening test, the resolution of spatial perception was studied using various sound source distributions. The test setup consisted of 15 loudspeakers that were placed in an anechoic chamber. Different loudspeaker combinations were used to produce the test cases such as one wide sound source or multiple simultaneous sound sources with varying widths. The latter can also be thought as being a wide sound source with gaps in the distribution. The subjects sat at the center of the room so that all the loudspeakers were equidistant from them. Using a touch-screen, they were asked to mark the loudspeakers that they perceived as emitting sound.

The results of the first listening test indicated that the resolution of spatial distribution perception is not high enough to be able to distinguish smallest details in the presented test

cases. The resolution for fine spatial details can be said to be of order of 15° - 30°. When there was one wide sound source, it was typically perceived as being slightly narrower than it was. Also, multiple simultaneous wide sound sources were perceived narrower than they were – conversely, it can be said that a gap in the sound source distribution was perceived as being wider than it actually was. Overall, when spatially distributed sound sources are presented to a subject, the overall perceptions and the resolution of perception are different from a situation where a single sound source at a time is presented.

In the second listening test, the ability to discriminate spatially distributed sound sources from one another was studied using noise with various bandwidths as well as pink noise and sine waves. These were presented to the subjects using the same test setup as in the first listening test. Loudspeaker combinations with different densities were used. The subjects' task was to discriminate which of the two shown loudspeaker combinations was used in producing the latter of two sound events.

The task proved out to be quite challenging and therefore, the results of the second listening test were not as conclusive as was desired. Nevertheless, they indicated that the discrimination task became more difficult as the the number of loudspeakers that were emitting sound was increased. The resolution of spatial distribution perception was rather poor. Also the bandwidth of the noise signals did have some effect on the results, but further studies are needed to find out more about this effect.

7.1 Future work

After performing the second listening test, some test subjects reported that sometimes they perceived differences in the two sound events but could not tell which loudspeaker combination corresponded to which sound event. Although training prior to the test could have reduced this problem, it was not used since intuitive and neutral perceptions were desired.

This led to a thought that the most interesting aspect is not whether the subjects can associate one of the two sound events with the correct loudspeaker combination. Rather, more interesting would be to find out whether they perceive differences in the sound events – and how large those differences are. So, instead of a discrimination task, a listening test where the subjects would have to rate the difference of two samples could yield interesting results. There, loudspeaker combinations with different number of loudspeakers emitting sound would be compared to a reference where all loudspeakers emit sound. The key variable which the subjects would have to pay attention to would be perceived spatial difference.

If the samples from the second listening test would be used, this further study could be a logical continuation to that test. It could confirm the indicative conclusions of the results of the second listening test.

Another issue that would need further studying are the sine wave cases of the second listening test. Those were such cases where multiple simultaneous sine waves were presented to the subjects. The results of those cases were compared to some of the noise cases and, interestingly, the comparison suggested that the frequency band was not the only attribute that affected the perception of the test cases. As there were only two different sine wave cases and only dissonant sets of sine waves were used, no certain conclusions can be drawn from them. A further study could address to this phenomenon more closely. A larger selection of frequency ranges and harmonic sets in addition to dissonant ones could be used.

Expanding the idea of sine waves to a more general level, a listening test with other sounds than noise could also be conducted. Spatial distribution and width perception is likely to be different when the test stimuli are, for example, different musical samples or birds singing.

Bibliography

S. Bech and N. Zacharov. *Perceptual Audio Evaluation - Theory, Method and Application*. John Wiley and Sons, Ltd., 2006.

L. Beranek. *Concert and Opera Halls: How They Sound*. Published for the Acoustical Society of America through the American Institute of Physics, 1996.

J. Blauert. *Spatial Hearing*. The MIT Press, Cambridge, MA, USA, revised edition, 1997.

J. Blauert and W. Lindemann. Spatial mapping of intracranial auditory events for various degrees of interaural coherence. *J. Acoust. Soc. Am*, 79:806–813, 1986.

E. G. Boring. Auditory theory with special reference to intensity, volume, and localization. *American Journal of Psychology*, 37(2):157–188, 1926.

D. Cabrera and S. Tilley. Parameters for auditory display of height and size. In *Proceedings of the 2003 International Conference on Auditory Display*, Boston, MA, USA, July 6-9, 2003.

D. Cabrera, A. Nguyen, and Y. J. Choi. Auditory versus visual spatial impression: A study of two auditoria. In *Proceedings of ICAD 04-Tenth Meeting of the International Conference on Auditory Display*, Sydney, Australia, July 6-9, 2004.

R. I. Chernyak and N. A. Dubrovsky. Pattern of the noise images and the binaural summation of loudness for the different interaural correlation of noise. In *Proceedings of the 6th International Congress on Acoustics*, Tokyo, Japan, 1968.

E. C. Cherry. Some experiments on the recognition of speech, with one and with two ears. *J. Acoust. Soc. Am*, 25(5):975–979, 1953.

H. Fletcher and W. A. Munson. Loudness, its definition, measurement and calculation. *J. Acoust. Soc. Am*, 5, issue 2:82–108, 1933.

M. B. Gardner. Distance estimation of 0° or apparent 0°-oriented speech signals in anechoic space. *J. Acoust. Soc. Am*, 45, issue 1:47–53, 1969.

B. R. Glasberg and B. C. J. Moore. Derivation of auditory filter shapes from notched-noise data. *Hearing Research*, 47:103–138, 1990.

E. B. Goldstein. *Sensation and Perception*. Wadsworth, sixth edition, 2002.

C. Guastavino and B. F. Katz. Perceptual evaluation of multi-dimensional spatial audio reproduction. *J. Acoust. Soc. Am*, 116:1105–1115, 2004.

M. L. Hawley, R. Y. Litovsky, and J. F. Culling. The benefit of binaural hearing in a cocktail party: Effect of location and type of masker. *J. Acoust. Soc. Am*, 115:833–843, 2004.

T. Hirvonen. *Perceptual and modeling studies on spatial sound*. Report 83, Helsinki University of Technology, Laboratory of Acoustics and Audio Signal Processing, Espoo, Finland, 2007.

T. Hirvonen and V. Pulkki. Center and spatial extent of auditory events as caused by multiple sound sources in frequency-dependent directions. *Acta Acustica united with Acustica*, 92(2):320–330, 2006a.

T. Hirvonen and V. Pulkki. Perceived spatial distribution and width of horizontal ensemble of independent noise signals as function of waveform and sample length. In *The 124th AES Convention*, Amsterdam, The Netherlands, May 17-20, 2008.

T. Hirvonen and V. Pulkki. Perception and analysis of selected auditory events with frequency-dependent directions. *J. Audio Eng. Soc.*, 9(54):803–814, 2006b.

K. Hiyama, S. Komiyama, and K. Hamasaki. The minimum number of loudspeakers and its arrangement for reproducing the spatial impression of diffuse sound field. In *The 113th AES Convention*, Los Angeles, California, USA, October 5-8, 2002.

F. J. Massey Jr. The kolmogorov-smirnov test for goodness of fit. *Journal of the American Statistical Association*, 46(253):68–78, 1951.

M. Karjalainen. *Kommunikaatioakustiikka*. Otamedia Oy, 1999.

R. G. Klumpp and H. R. Eady. Some measurements of interaural time difference thresholds. *J. Acoust. Soc. Am*, 28:859–860, 1956.

R. Y. Litovsky, H. S. Colburn, W. A. Yost, and S. J. Guzman. The precedence effect. *J. Acoust. Soc. Am*, 106, issue 4:1633–1654, 1999.

Lord Rayleigh (a.k.a. J. W. Strutt 3rd Baron of Rayleigh). On our perception of sound direction. *Phil. Mag.*, 13:214–232, 1907.

R. Mason, T. Brookes, and F. Rumsey. Frequency dependency of the relationship between perceived auditory source width and the interaural cross-correlation coefficient for time-invariant stimuli. *J. Acoust. Soc. Am*, 117(3, pt. 1):1337–1350, 2005.

A. W. Mills. Lateralization of high-frequency tones. *J. Acoust. Soc. Am*, 32:132–134, 1960.

H. Møller, M. F. Sørensen, D. Hammarshøi, and C. B. Jensen. Head-related transfer functions of human subjects. *J. Audio Eng. Soc.*, 43:300–321, 1995.

D. Perrott. Discrimination of the spatial distribution of concurrently active sound sources: Some experiments with stereophonic arrays. *J. Acoust. Soc. Am*, 76(6):1704–1712, 1984.

D. Perrott and T. Buell. Judgments of sound volume: Effects of signal duration, level, and interaural characteristics on the perceived extensity of broadband noise. *J. Acoust. Soc. Am*, 72(5):1413–1417, 1981.

D. R. Perrott, A. Musicant, and B. Schwethelm. The expanding image effect: The concept of tonal volume revisited. *Journal of Auditory Research*, 20:43–55, 1980.

G. Potard and I. Burnett. A study on sound source apparent shape and wideness. In *Proceedings of the 2003 International Conference on Auditory Display*, Boston, MA, USA, July 6-9, 2003.

G. Potard and I. Burnett. Decorrelation techniques for the rendering of apparent sound source width in 3d audio displays. In *Proceedings of the 7th Int. Conference on Digital Audio Effects*, Naples, Italy, October 5-8, 2004.

V. Pulkki. *Spatial Sound Generation and Perception by Amplitude Panning Techniques*. Report 62, Helsinki University of Technology, Laboratory of Acoustics and Audio Signal Processing, Espoo, Finland, 2001.

V. Pulkki. Spatial sound reproduction with directional audio coding. *J. Audio Eng. Soc.*, 55 (6):503–516, June 2007.

V. Pulkki and T. Hirvonen. Computational count-comparison models for binaural cue decoding. *Unpublished manuscript*, 2009.

D. W. Robinson and R. S. Dadson. A re-determination of the equal-loudness relations for pure tones. *British Journal of Applied Physics*, 7:166–181, 1956.

T. D. Rossing, F. R. Moore, and P. A. Wheeler. *The Science of Sound*. Addison Wesley, third edition, 2002.

H. Wallach, E. B. Newman, and M. R. Rosenzweig. The precedence effect in sound localization. *American Journal of Psychology*, 57:315–336, 1949.

R. M. Warren. *Auditory perception: a new analysis and synthesis*, pages 189–196. Cambridge University Press, 1999.

F. L. Wightman and D. J. Kistler. The dominant role of low-frequency interaural time differences in sound localization. *J. Acoust. Soc. Am*, 91(3):1648–1661, 1992.

W. A. Yost. *Fundamentals of Hearing - An Introduction*. Academic Press, third edition, 1994.

W. A. Yost and G. Gourevitch. *Directional Hearing*. Springer-Verlag New York Inc., 1987.

E. Zwicker, E. G. Flottorp, and S. S. Stevens. Critical band width in loudness summation. *J. Acoust. Soc. Am*, 29:548–557, 1957.