Eindhoven University of Technology

MASTER

Localisation of surround sound signals

Groeneveld, R.J.H.

Award date:
1993

Link to publication

Disclaimer
This document contains a student thesis (bachelor's or master's), as authored by a student at Eindhoven University of Technology. Student theses are made available in the TU/e repository upon obtaining the required degree. The grade received is not published on the document as presented in the repository. The required complexity or quality of research of student theses may vary by program, and the required minimum study period may vary in duration.

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain

411354

Technische Universiteit Eindhoven
Instituut voor Perceptie Onderzoek

Localisation of Surround Sound Signals

R.J.H. Groeneveld

Verslag van een afstudeeropdracht verricht in de periode december 1992 t/m oktober 1993 bij de groep Physical Acoustics van het Philips Natuurkundig Laboratorium te Eindhoven.

Begeleider: Dr.Ir. W.R.T. ten Kate (Philips Research)

Afstudeerhoogleraar: Prof. Dr. Ir. J.A.J. Roufs (I.P.O. / T.U.E.)
Prof. Dr. A.J.M. Houtsma (I.P.O. / T.U.E.)

Abstract

The human localisation precision for sounds reproduced by the surround loudspeakers in a multichannel audio set-up has been measured as a function of the level of interfering sound reproduced by the front loudspeakers.

The experimental method comprised the generation of click-triplets which were reproduced as phantom sources using surround loudspeakers. Pink noise of varying levels was reproduced using two front loudspeakers. The subjects were asked to perform a so-called AAB/ABA comparison, by indicating which of the 3 clicks was localised differently compared to the two other clicks. Four pink noise levels were employed, one of which had a value of $-\infty$ dB. The clicks were positioned by panning at three different locations behind the subject. At each such location, three illusory spatial seperations of the A and B clicks were used.

It was found that the results of this test neither showed a marked dependence nor fully precluded any influence of noise level on localisation precision.

Contents

Terms used in acoustics

The following list is intended for those who are not so familiar with acoustics and auditive perception. It is not designed as an extensive explanation of acoustical terminology, but rather to give an introductory insight into certain subjects sometimes regarded as being of common knowledge by those who work in acoustics.

- Minimal Audible Angle: A measure of the accuracy with which human listeners can localise a sound source. It is defined as the angle between two sound sources, as perceived from the listener's position, when those sources can just be separately localised by the listener.

- Pink noise: Noise for which equal fractions of an octave contain the same power; the spectrum level decreases uniformly with increasing frequency, with a slope of −3 dB/octave. An octave is a frequency interval characterised by a doubling in frequency.

- Sound Pressure Level (SPL): Terminology for expressing pressure using a logarithmic scale. SPL is defined as:

$$\mathrm{SPL} = 20 \log \left(\frac{p}{p_{\mathrm{ref}}} \right) \qquad [\mathrm{dB}] \tag{0.1}$$

  where $p_{\mathrm{ref}} = 20\ \mu$ Pa is the effective pressure (in air at STP) of a 1000 Hz sine wave on the threshold of audibility. A widely used measure of environmental noise is the A-weighted SPL [dB(A)]. A-weighting assigns to each frequency a "weight" that is related to the sensitivity of the ear at that frequency.

- Digital Recording: Technique for recording sounds using analogue-to-digital and digital-to-analogue converters [hereinafter ADC and DAC respectively] . The electrical signal of the sound to be recorded is fed through an ADC that will convert it into a number. This number is recorded. In playback, the number is converted to an electrical signal by the DAC. The quality of the recording is determined by the range of numerical values to which the ADC can convert the electrical signal, as well as the frequency at which it operates, called the sampling rate. Most modern digital recorders have 16 bit resolution, meaning that the electrical signal is converted to a 16-bit number. This implies a value in the range $-2^{15} = -32768$ to $2^{15} - 1 = 32767$. The sampling rate is usually 44.1 kHz (employed for instance in CD recording), though 32 kHz and 48 kHz are also widely used.

It also possible to create sounds on a computer that are then reproduced using a DAC.

- Time code: Coded signal that can be recorded and played back like an audio signal, containing information about elapsed time. Once recorded on a tape, it can be used to refer to a position on the tape in terms of time.

- Phantom Source: Apparent position of a sound source, i.e. the position at which a sound source is perceived.

- Panning: Technique for positioning a phantom source using multi-channel sound reproduction, by changing the level of the channels. In stereo sound reproduction, a sound source can be perceptually moved from side to side by increasing the level of one channel, whilst decreasing the level of the other channel. It is important to keep the total power output of the stereo channels constant, to prevent the loudness of the phantom sound source from changing. For instance in the case of a sound source panned half way between left and right references, the power levels in the left and right channels will be equal, with a value of S, say. If this sound source is completely panned to the left, the output of the right channel will be $0 \times S$ and the output of the left channel will be $\sqrt{2}S$. The combined power output of the two channels remains unchanged: $S^2 + S^2 = 2S^2 = 0^2 + (\sqrt{2}S)^2$.

- Live end, Dead end: Terms used for features common in the design of recording studios and professional listening rooms. The former specifies the immediate vicinity of a wall which reflects sound (live end) whereas the latter specifies the locality of a wall which absorbs sound (dead end), the walls being oppositely located in an enclosed space.

Chapter 1

Introduction

1.1 Multi-channel sound

To take television into the 21st century, manufacturers of television equipment have developed High Definition Television (HDTV). This is a television system that encompasses all aspects of television broadcasting and reception. The differences with today's television system encompass the size and quality of the television image, as well as the quality and number of channels of the sound.

HDTV provides a picture which has an aspect ratio of 16:9, whereas that of today's television is 4:3. The 16:9 ratio is similar to the aspect ratio of a cinema screen, and has a better consistency with the human field of vision. Furthermore, the number of pixels comprising the HDTV television picture is four times larger than that employed in today's television standard.

To complement this improvement in visual quality, HDTV will involve digital multi-channel sound. The number of channels has not yet been decided upon, but five is the most likely. These channels are Left, Right, and Centre, in front of the viewer, and Surround Left and Surround Right behind the viewer [hereinafter FL, FR, FC, SL, SR, respectively]. Fig. 1.1 shows how the five loudspeakers are ideally placed in a living room environment.

The FL and FR loudspeakers will be further apart than the conventional stereo loudspeakers. The added FC channel stabilises the stereo image. The main purpose of sounds being reproduced by the two surround loudspeakers is to recreate some of the acoustic environment of the scene shown on the television screen.

This improvement in the sound reproduction system of television is also thought to be applicable to sound-only reproduction systems. The five-channel sound-only reproduction setup will be similar to the system shown in Fig. 1.1.

Surround Sound is nothing new and is perhaps best known from cinemas. In such existing applications, one channel is used to create the Surround Sound signal which is then presented via a number of loudspeakers placed at various positions throughout the room, thereby creating a spatial effect.

The Japanese HDTV standard uses four sound channels viz. three front channels and one for surround sound. This surround channel is reproduced using two loudspeakers placed behind the viewer. The loudspeaker setup is therefore identical

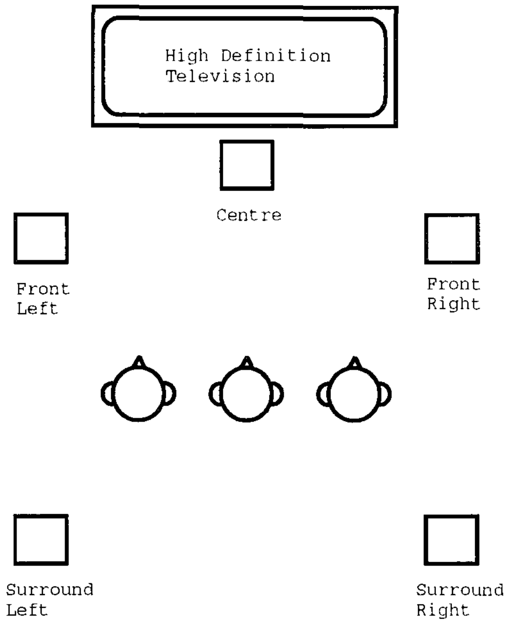

Figure 1.1: *HDTV setup of television and loudspeakers.*

to that shown in Fig 1.1.

The Surround Sound systems having one channel for Surround Sound are labeled Mono Surround Sound. Systems using two channels for Surround Sound and two loudspeakers to reproduce them are labelled Stereo Surround Sound systems. When, in a stereo Surround Sound system, both surround channels carry the same signal, then this is referred to as mono surround.

1.2 Sound information reduction

Using stereo Surround Sound instead of mono Surround Sound may seem the way to better spatial sound reproduction, but it also means an increase of a quarter of the sound information. To find out if implementing stereo surround rather than mono surround is worth the extra effort, listening tests were held to see if people preferred stereo over mono surround. The results published in [1] show that in just a few cases listeners preferred stereo surround over mono surround.

This led to the aim of reducing the amount of sound information of stereo surround whilst retaining the advantages of stereo surround over mono surround.

Over the past ten years, Philips Compact Disc has, through its growth in popularity, set a standard for digital sound encoding. A drawback of digital recording in CD quality is the amount of information that it produces. Certain sound information reduction techniques do exist, such as those used in Digital Compact Cassette (DCC) recorders or in the Hidden Channel Technique. The latter is a technique used to hide information in an audio signal. Such information can comprise, for example, another audio signal. Both techniques make use of properties of human perception. To be able to use techniques like these to reduce surround sound information, it is necessary to investigate the nature of perception of sounds being reproduced via

Surround Sound loudspeakers, i.e. sound events that occur behind the vertical plane running through both ears of the listener.

1.3 Ambience and localisation

Sound events in Surround Sound can be divided into two categories:

- Sounds that add to the experience of ambience. These are sounds of a reverberant nature, for example applause from an audience during concerts or sporting events. The precise location of each sound event is not required, but rather the general impression thus created.
- Sound-events that are localised in a certain position. These can include many kinds of special effects or localised ambient sounds.

This division into two categories is made judging on the correlation of left and right channels. It is generally assumed that the main characteristic of sounds in the first category is that there is very little or no correlation between the channels. For sounds in the second category, there is a well-defined correlation between the channels defining the location of the sound events behind the viewer. The fact that these two categories have different characteristics as far as correlation of left and right channels is concerned makes information reduction difficult. More knowledge relating to the correlation between left and right surround channels is therefore needed. The research described in this report concentrates on the ability of humans to localise sound events behind them as well as factors that influence this localisation.

1.4 Aim of this localisation test

A listening test was designed to investigate the localisation precision for Surround Sound signals, as a function of level of pink noise being reproduced in front of a subject. The listening test was set up in a room with a short reverberation time, so as to simulate a living room environment. The FL, FR, SL and SR of a five-channel-sound loudspeaker setup were used to reproduce the test signals. The front loudspeakers were used to reproduce noise on several levels, and the surround loudspeakers were used to reproduce clicks as phantom sources located at several positions behind the subject. With this setup, the localisation precision of phantom clicks as a function of the level of noise being reproduced in front of the listener, was investigated.

In what follows, Chapter 2 introduces the background of the statistics used in processing the data from the listening test. Chapter 3 then gives an outline of the listening test itself. The results of the listening test are subsequently discussed in Chapter 4, and Chapter 5 concludes the report by summarising the main findings.

Chapter 2

Statistical background

2.1 Introduction

The difficulties associated with testing the localisation precision of a subject, for sounds that are reproduced behind the subject, are twofold:

- First there is the problem of indicating the sound source. When a sound is reproduced behind a subject, it is difficult for the subject to point out where the sound is located. If the subject were to point out the position of the sound source directly, he would have to turn around each time, thereby introducing an uncertainty in locating the source. Using a sound pointer aligned with the phantom source involves a lot of time and requires an elaborate setup. Having the subject indicate the direction of two sound events might cause confusion in directivity.

- When a subject has to perform on the threshold of his perceptual capacities it is not possible for him to distinguish between locations from memory. In this situation, the only way to find out if a subject is still able to distinguish between locations is by direct comparison (i.e. without an intermediate memory-step).

A method which requires no prior knowledge of the subject's individual characteristics and involves an easily understood well-defined task to be performed by the subject is the AAB/ABA test. To explain this test in more detail, an introductory perceptual model of localisation will now be given.

2.2 A perceptual model for localisation

A physical event that is probed by our senses is turned into a sensation. This translation of an event in the physical world outside a human into a sensation in a human's psyche takes place in what is called the psychophysical space. This is an abstract space, a model of the cerebral and endocrine processes which turn physical reality into sensations. In this psychophysical space, we depict localisation by a decision axis. When a stimulus is applied a localisation sensation is invoked along that axis. The most likely sensation is called μ. The probability of registering a different sensation is given by a Gaussian probability distribution around μ with a standard deviation σ, as shown in Fig. 2.1. This distribution is given by:

$$P(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right) \tag{2.1}$$

When two different stimuli A and B are presented, they will invoke the two asso-

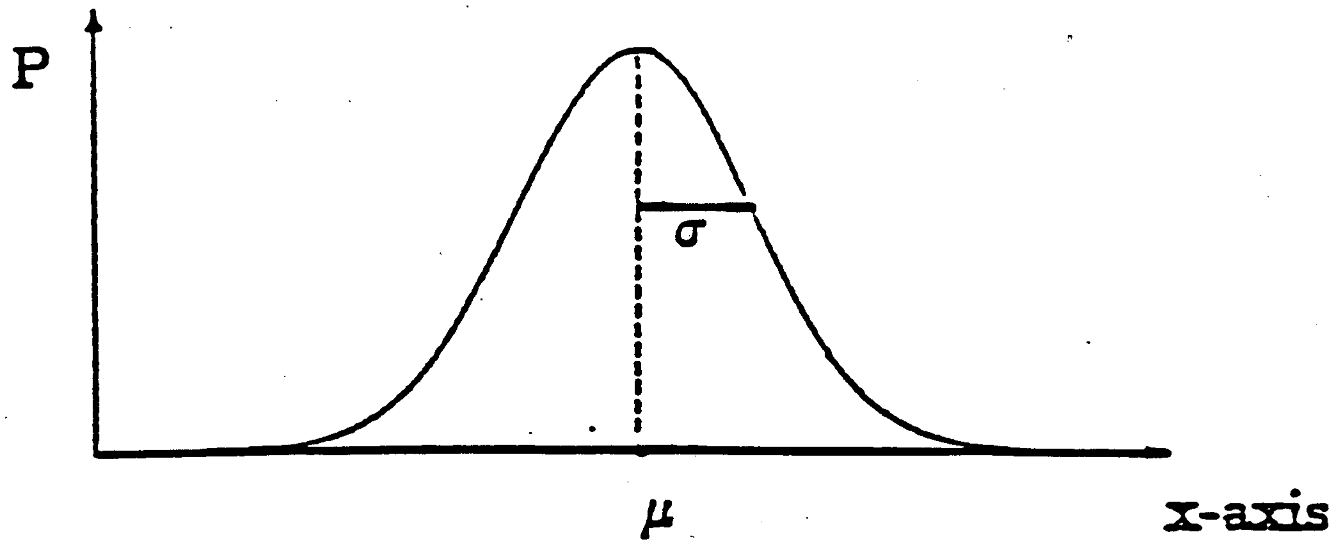

Figure 2.1: *Gaussian probability distribution along the localisation decision axis in psychophysical space.*

ciated most probable sensations, μ_A and μ_B respectively. If the two stimuli A and B do not differ too much, the probability distributions around μ_A and μ_B will have the same σ and can be depicted as in Fig. 2.2. Although the sensation μ_A is the most likely to occur when stimulus A is presented, though another sensation may also occur, perhaps even μ_B. In this latter case, the subject will think stimulus B was presented. In deciding whether stimulus A or B has occurred, a subject will use a simple criterion:

- $x < C$, then stimulus A has occurred;
- $x > C$, then stimulus B has occurred.

Here, x is the decision variable, and C is a point on the decision axis. It usually lies at the co-ordinate of the intersection of the Gaussian distribution curves of stimuli A and B see 2.2. We now introduce a sensitivity index, see [2]:

$$d' = \frac{\mu_B - \mu_A}{\sigma} \tag{2.2}$$

This is a measure of the perceptibility of the stimuli A and B. The sensitivity index $d' = 1$ is commonly defined as the discrimination threshold: if $d' < 1$ then the stimuli A and B are said to be not significantly different in a perceptual sense.

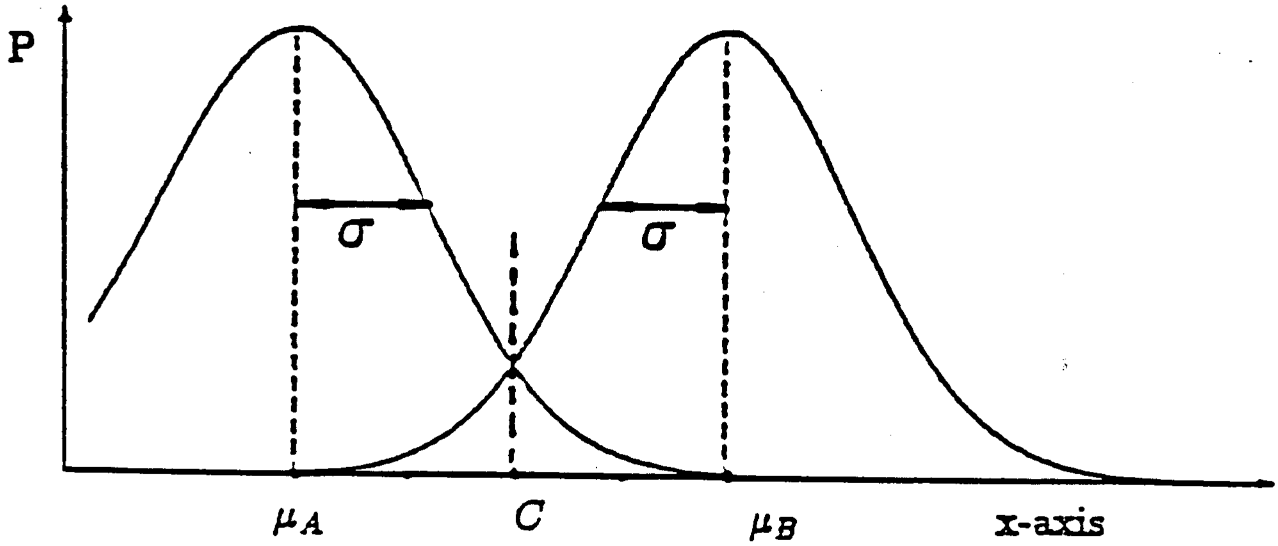

Figure 2.2: *Probability distribution of two stimuli A and B on the localisation decision axis.*

It is known that the localisation precision varies with the position of the sound source in the horizontal plane [3]. This means that the value of σ varies as a function of the decision variable. In this test it is assumed that the value of σ will be constant for those stimuli which are used in the same localisation region. Variations within such a region was needed to measure localisation precision att all.

2.3 The AAB/ABA test

In investigating the limits of perceptual abilities, the following situation occurs upon reaching the threshold. A subject is able to distinguish between two stimuli but he is not able to tell them apart when just one is presented. This situation will occur in this localisation test. Therefore the method used to test the localisation precision in this test must be based on comparison rather than recognition. The AAB/ABA test, which is just such a test, was chosen.

In an AAB/ABA test, three stimuli are presented to the subject, two of which are the same. It is the task of the subject to indicate whether the second or third stimulus differs from the first.

The three sensations x_1, x_2 and x_3 can be depicted on a decision axis in psychophysical space as in Fig. 2.3.

Making a decision on whether the second stimulus, resulting in sensation x_2, or the third stimulus, resulting in sensation x_3, was the same as the first stimulus, resulting in sensation x_1, a similar criterion to that presented in the previous section is applied. The subject decides which of x_2 or x_3 is closest to x_1, as follows:

- if

  $$|x_1 - x_2| > |x_1 - x_3| \tag{2.3}$$

  then the subject will choose ABA;

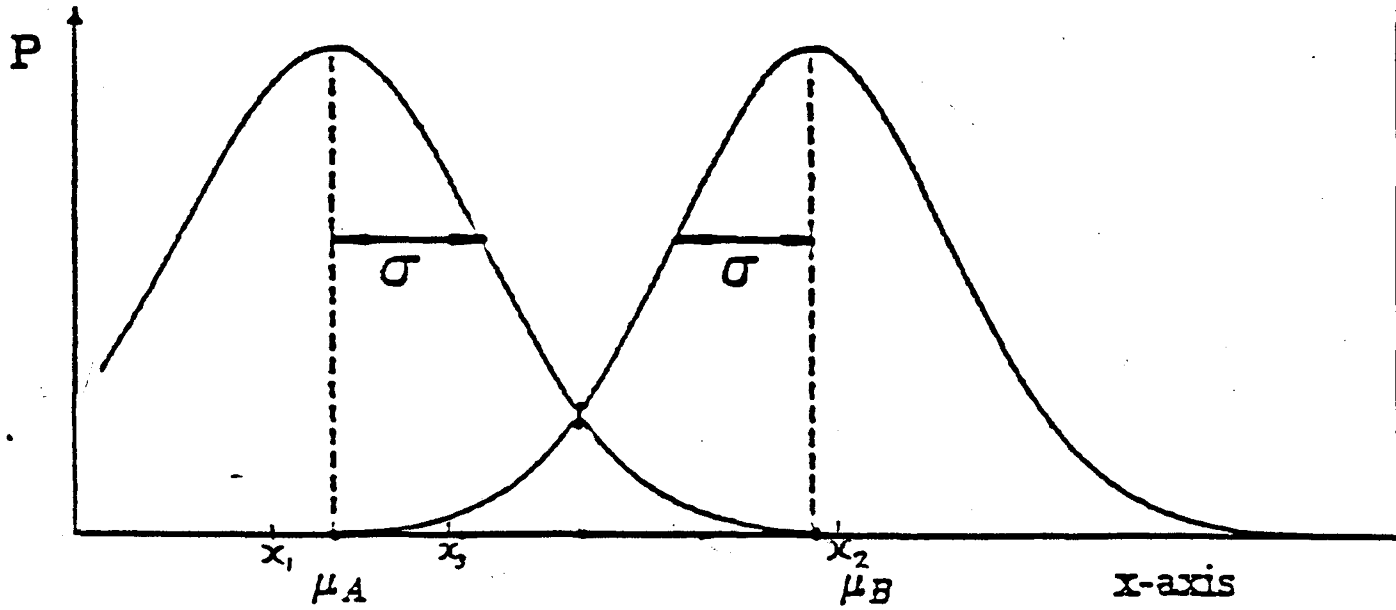

Figure 2.3: *Three possible sensations x_1, x_2 and x_3 on the localisation decision axis, resulting from an ABA click triplet.*

- if

  $$|x_1 - x_2| < |x_1 - x_3| \tag{2.4}$$

  then the subject will choose AAB.

All stimuli are assumed to be grouped close enough together on the decision axis so that the value of σ will be the same for the three stimuli. It is assumed that there is no bias.

The possible click-triplets are AAB, ABA, BBA and BAB. If the subject's answers are denoted by "2" or "3" when the second or third click is, in his opinion, the odd-one-out, then the probability of a correct answer is:

$$\begin{aligned} P_{\text{correct}} &= P(\text{“2”}|\text{ABA})P(\text{ABA}) + P(\text{“2”}|\text{BAB})P(\text{BAB}) + \\ &\quad P(\text{“3”}|\text{AAB})P(\text{AAB}) + P(\text{“3”}|\text{BBA})P(\text{BBA}) \\ &= P(\text{“2”}|\text{ABA}) = P(\text{“3”}|\text{AAB}) \end{aligned} \tag{2.5}$$

It has hereby been assumed that there is no bias, i.e. $P(\text{“2”}|\text{ABA}) = P(\text{“3”}|\text{AAB})$ and that the design of the test is balanced i.e. $P(\text{ABA}) = P(\text{AAB})$.

When AAB is presented, the probability of a subject thinking the third click to be the odd-one-out is:

$$P(\text{“3”}|\text{AAB}) = P(\text{“AAB”}|\text{AAB}) + P(\text{“BBA”}|\text{AAB}) \tag{2.6}$$

Note that the subject may answer correctly even when the sensation he experiences is totally opposite to the one that is actually presented.

If we consider x_1, x_2 and x_3 as coordinates in a three dimensional decision space, then the criteria (2.3) and (2.4) are equal to surfaces in that space. The surfaces are defined as:

$$\begin{aligned} x_1 - x_2 = x_1 - x_3 \quad &\Rightarrow s_1: \quad x_2 - x_3 = 0 \\ x_1 - x_2 = -(x_1 - x_3) \quad &\Rightarrow s_2: \quad 2x_1 - x_2 - x_3 = 0 \end{aligned} \tag{2.7}$$

The surfaces are perpendicular and divide the decision space into four parts . In each of these four parts, one of the expected sensations is located. The expected sensations are: μ_{AAB}, μ_{ABA}, μ_{BBA} and μ_{BAB}. The corresponding parts of the decision space are called: V_{AAB}, V_{ABA}, V_{BBA} and V_{BAB}. Eq. (2.6) can now be written as:

$$P(\text{“3”}|\mathrm{AAB}) = \int_{V_{\mathrm{AAB}}} P_{\mathrm{AAB}}(\vec{x})\,\mathrm{d}V + \int_{V_{\mathrm{BBA}}} P_{\mathrm{AAB}}(\vec{x})\,\mathrm{d}V \tag{2.8}$$

with:

$$P_{\mathrm{AAB}}(\vec{x}) = \frac{1}{\sigma^3 (2\pi)^{\frac{3}{2}}} \exp\left(-\frac{1}{2}\left(\frac{x_1-\mu_{\mathrm{A}}}{\sigma}\right)^2 - \frac{1}{2}\left(\frac{x_2-\mu_{\mathrm{A}}}{\sigma}\right)^2 - \frac{1}{2}\left(\frac{x_3-\mu_{\mathrm{B}}}{\sigma}\right)^2\right) \tag{2.9}$$

In order to calculate these integrals, the decision space is transformed by an orthonormal transformation. After rewriting the integrals (for details refer to [4]) they finally led to:

$$P(\text{“AAB”}|\mathrm{AAB}) = \int_{V_{\mathrm{AAB}}} \mathrm{d}V P_{\mathrm{AAB}} = \Phi\left(\frac{d'}{\sqrt{2}}\right)\Phi\left(\frac{d'}{\sqrt{6}}\right) \tag{2.10}$$

where d' is defined previously in Eq. (2.2) and Φ is the error function. And:

$$P(\text{“BBA”}|\mathrm{AAB}) = \int_{V_{\mathrm{BBA}}} P_{\mathrm{AAB}}\mathrm{d}V = \left(1-\Phi\left(\frac{d'}{\sqrt{2}}\right)\right)\left(1-\Phi\left(\frac{d'}{\sqrt{6}}\right)\right) \tag{2.11}$$

Combining Eq. (2.5), (2.6), (2.10) and (2.11) we now find for P_{correct}:

$$P_{\mathrm{correct}} = 1 - \Phi\left(\frac{d'}{\sqrt{2}}\right) - \Phi\left(\frac{d'}{\sqrt{6}}\right) + 2\Phi\left(\frac{d'}{\sqrt{2}}\right)\Phi\left(\frac{d'}{\sqrt{6}}\right) \tag{2.12}$$

The results from the AAB/ABA tests will give values of percentage correct Q_{correct}. These values will be considered as an estimate for the value of P_{correct}. From this, d' can be subsequently estimated by inverting:

$$Q_{\mathrm{correct}} = 1 - \Phi\left(\frac{d'}{\sqrt{2}}\right) - \Phi\left(\frac{d'}{\sqrt{6}}\right) + 2\Phi\left(\frac{d'}{\sqrt{2}}\right)\Phi\left(\frac{d'}{\sqrt{6}}\right) \tag{2.13}$$

An estimate for the value of the standard deviation σ of the probability curve in our perception model σ then follows as:

$$\sigma = \frac{\mu_{\mathrm{B}} - \mu_{\mathrm{A}}}{d'} \tag{2.14}$$

Using Eq. (2.13) a value for d' can be estimated. However there is no precise knowledge of the values μ_{A} and μ_{B}. For evaluation purposes, the value of Δ, expressed in dB, shall be used in place of $\mu_{\mathrm{B}} - \mu_{\mathrm{A}}$. This is an arbitrary choice. However, as the results will show, this enables results to be compared for various values of Δ.

Chapter 3

Localisation test design

3.1 General setup

The aim of this localisation test is to establish the human precision in localising sound events being reproduced at several locations behind the subject as a function of the level of noise being reproduced in front of that person. The test was done in a listening room with a short reverberation time, using a 5-channel audio loudspeaker setup. The test sounds being reproduced behind the subject were clicks. They were created by panning between the SL and SR loudspeakers. The interfering noise being reproduced in front of the subject was stereo pink noise. This was reproduced using the FL and FR loudspeaker. The FC loudspeaker was not used in this test. The experimental setup is shown in Fig. 3.1.

During the listening test, the subjects were asked to indicate if, in a sequence of three clicks (a so called triad) the second or third click was located in a different position to the other two clicks.

The stereo pink noise was reproduced from the SQAM CD [1] [5]. Pink noise has a broad frequency range and a stable power spectrum, which approximates to most common audio signals. Furthermore, in this test it was important that the subject should not be distracted by the nature of the sound being played through the front loudspeakers. The pink noise was digitally recorded into a computer. With this computer a four-track test sound tape was made. The sampling frequency used was 44.1 kHz, and the sampling resolution was 16 bits. It was decided that the noise levels should upon reproduction be $-\infty$ dB(A), 40 dB(A), 50 dB(A) and 60 dB(A). The latter three noise levels were chosen to simulate sound levels of normal television viewing. The average value of the noise samples was 122 RMS (16 bit). The power spectrum of the pink noise used in the experiment is shown in Fig. 3.2.

Clicks were chosen as the sounds to be localised by the subjects, as clicks are relatively easy to localise. In addition, clicks have a broad power spectrum, making them comparable with full-range signals. Furthermore, using a click makes the results of this experiment easy to reproduce. The clicks used were generated using a computer. They were simply a pulse with a width of one sample (= 1/44100 sec.)

[1]Sound Quality Assessment Material. CD with recordings of a various nature, used in listening tests

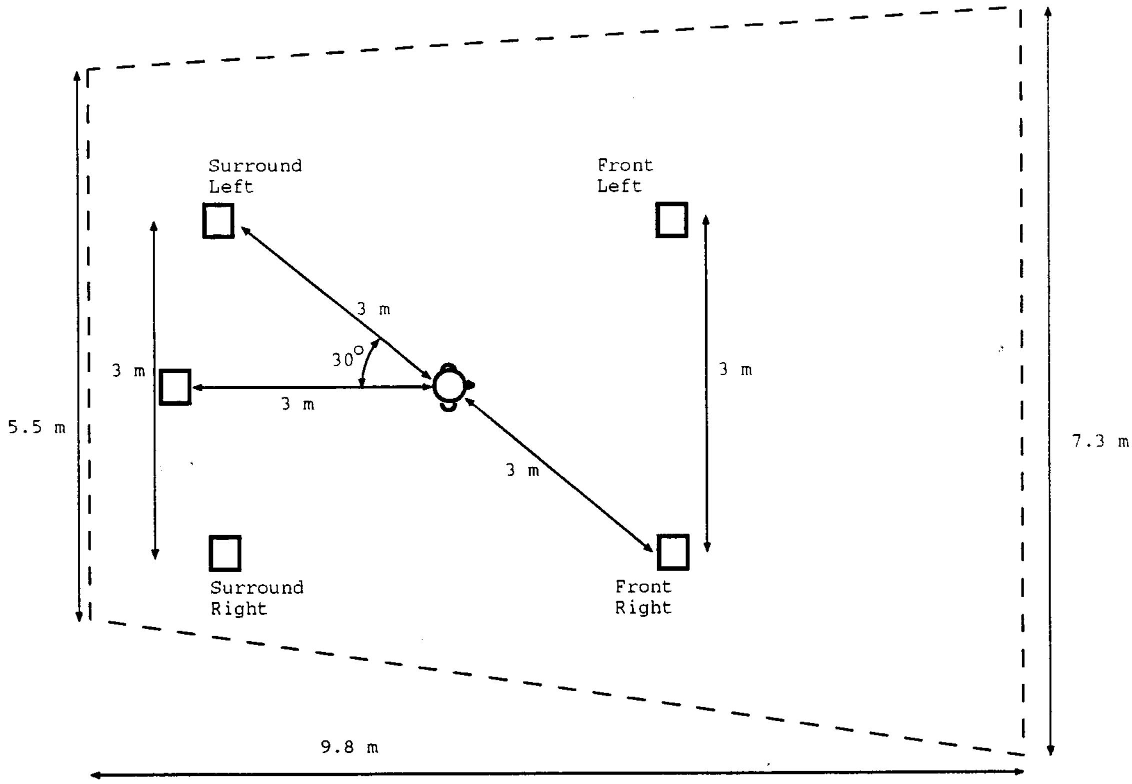

Figure 3.1: *Loudspeaker setup in the listening room.*

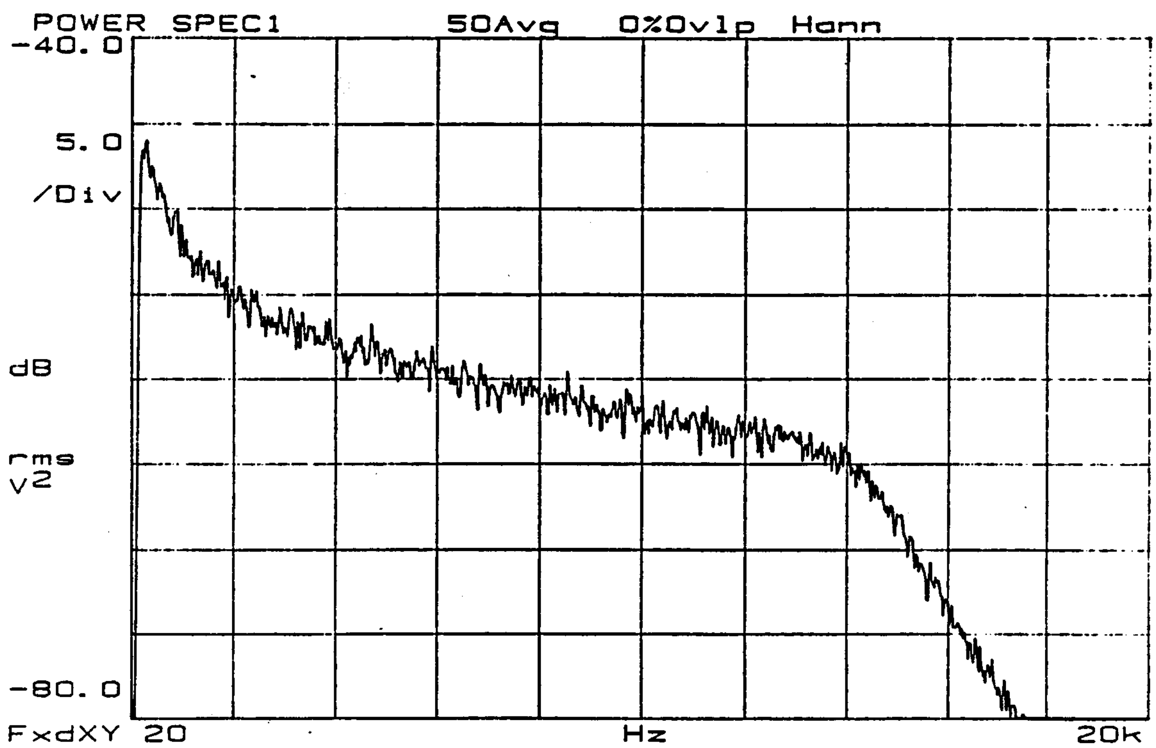

Figure 3.2: *Power spectrum of pink noise used in the localisation test.*

and an amplitude of 23210 ($\approx 2^{15}/\sqrt{2}$) (16 bit). The clicks were panned between SL and SR loudspeakers. The power of the clicks was kept constant during panning, and their power spectrum is shown in Fig. 3.3.

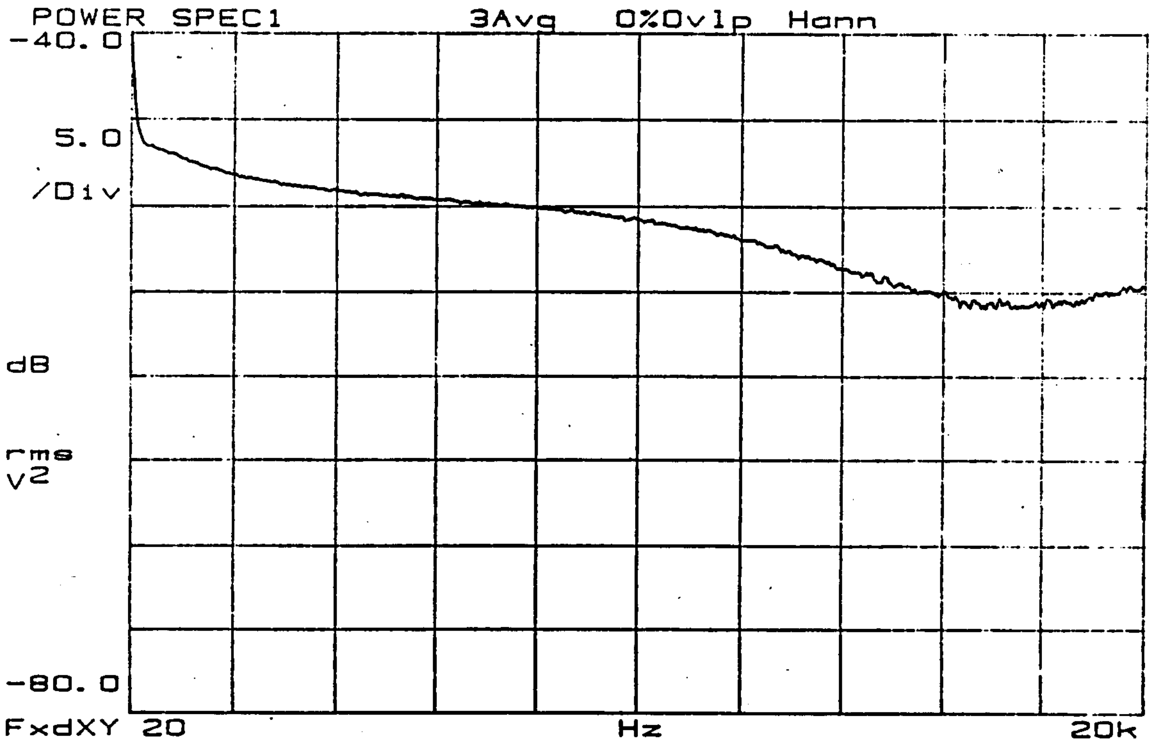

Figure 3.3: *Power spectrum of the clicks used in the localisation test.*

The localisation of sounds being reproduced through the surround loudspeakers depends, amongst other things, on the intensity-difference between the surround channels. The dependency between localisation of sound events and the intensity-difference between two channels has previously been a subject, of research e.g. [6]. This research concentrated on localisation in front of the subject, for which the dependency is depicted in Fig. 3.4.

Assuming these results also to be valid for sounds being panned behind the subject, the phantom clicks can be positioned between $-30°$ and $30°$ with respect to the central axis behind the subject see Fig. 3.5. Because of the number of parameters that had to be varied in this test, and the limited time for which one can perform a listening test, it was decided that three positions behind the subject were to be used, whereby symmetry between perception on the left and right side of a subject was assumed. The three chosen positions used in this test were 5°, 15° and 25° as depicted in Fig. 3.5. These positions correspond to intensity-differences of 2, 6 and 10 dB between the left and right surround channels. Hereinafter the positions will be referred to in terms of level difference, in dB, between SL and SR.

In order to investigate the localisation precision for sounds localised at the three chosen positions behind the subject, it is necessary to vary the position of the clicks around those three set positions. The employed variations around the three set

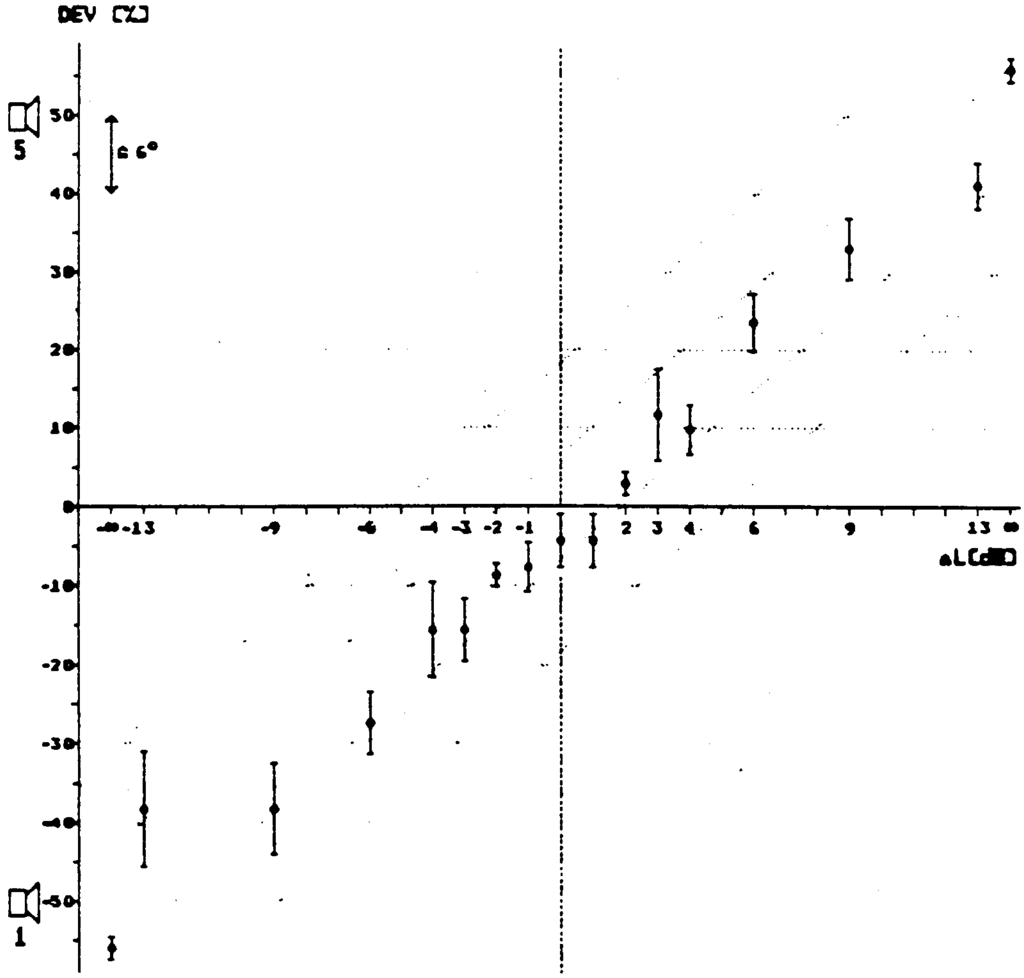

Figure 3.4: *Location of phantom sources as a function of intensity-differences between two loudspeakers. Taken from [6].*

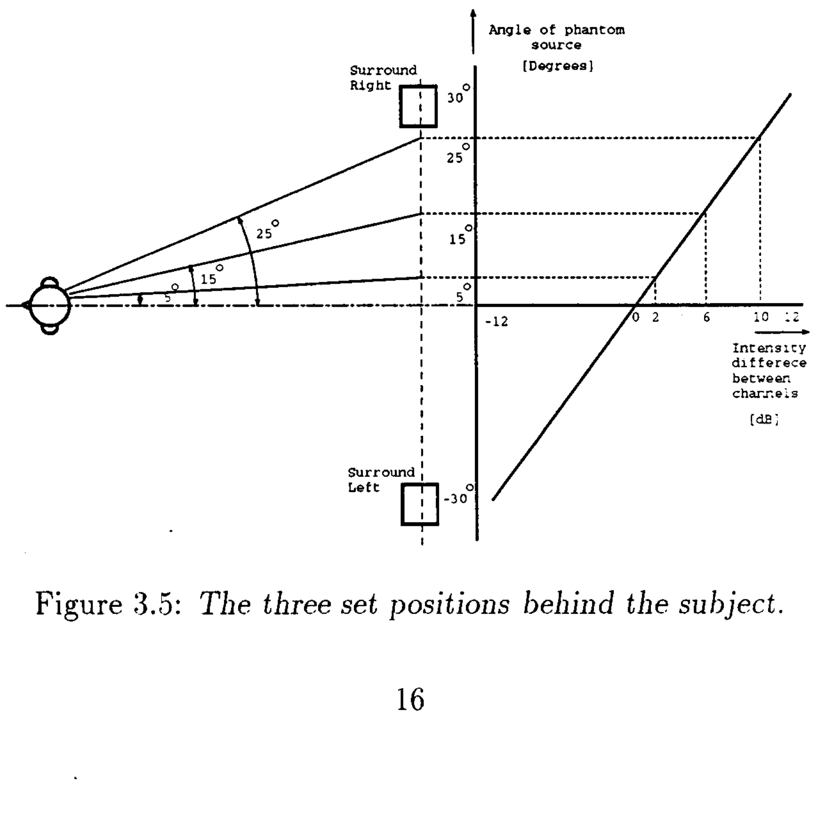

Figure 3.5: *The three set positions behind the subject.*

positions behind the subject are depicted in Fig. 3.6. The value of a given variation in the position of a click will be referred to as delta (Δ) expressed in dB terms. As tests have shown, the localisation precision of a person depends on the position of the monitored sound event in the subject's horizontal plane, as demonstrated by Blauert [3], amongst others. Therefore, the value of Δ was not equal for the three positions. This is also depicted in Fig. 3.6.

For every position, three values of Δ were used in the test. The values of Δ were assessed in preliminary tests, to provide optimal results, the chosen values being listed in table 3.1. The clicks were positioned symmetrically around the three positions behind the subject, to ensure that results found with these values of Δ would comply with the method of measuring the Minimum Audible Angle as described in [7].

Position (dB)	2	6	10
Δ (dB)	1, 2, 3,	1, 2, 3	4, 5, 6

Table 3.1: *Values of Δ used at different positions behind the subject.*

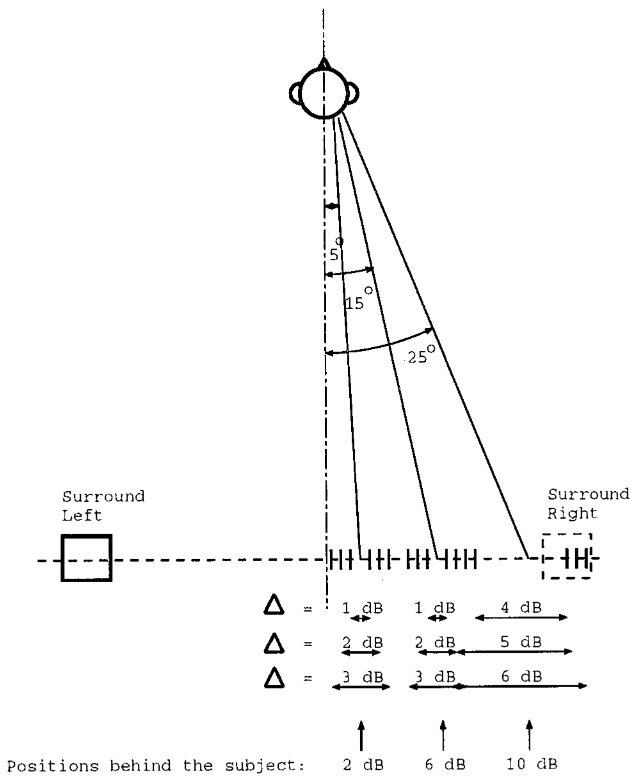

Figure 3.6: *Differences in localisation precision at different locations.*

The interval between the clicks was half a second. They had to be at least 2 tenths of a second apart so as to avoid localisation problems for the subject, caused by lack of speed in the localisational ability of the brain. Research into this subject is published in [8].

The localisation of the clicks has to conform with certain rules. The triads have to be symmetrical, so that each AAB triad must be played exactly as many times as each ABA triad. Furthermore, in a given AAB or ABA triad, either the A or B click can be localised closer to the centre axis through the subject's head. The click closest to this axis has to be the A click exactly as many times as it is the B click. So for each position behind the subject and each value of Δ, four AAB/ABA triads need to be performed to calculate the results. It was decided to do a total of 9 sets of measurements for each position behind the subject and for each value of Δ, implying a total of 36 AAB/ABA triads.

The clicks and the noise being reproduced during the experiment were recorded beforehand on a 24-track recorder. The attenuation of the noise needed to obtain the various noise levels was done using the computer so that everything could be recorded without changing the settings on the mixing desk whilst recording. When playing back the tape during the experiment, two separate amplifiers were used for the front and surround speakers. The amplifier connected to the front loudspeakers was set so that the loudest noise level was 60 dB(A). The amplifier connected to the surround loudspeakers was set to a level that assured easy localisation of the clicks. When the loudest noise was played through the surround loudspeakers, the sound pressure level at the listener's position was 65.5 dB(A). A time code was recorded on the tape, thus allowing a PC to control the experiment by reading in this time code, and, via a response box, recording the answers given by the subjects. This setup is depicted in Fig. 3.7.

The loudspeakers were set up in the listening room at Philips Research in Eindhoven. This listening room has a live and a dead end (LEDE). It was found that the reflections that occurred in the live end of the room made localisation of phantom sources from the rear very difficult, and so curtains were used to damp the live end. Furthermore, the setup was turned 180° so that the subject was now facing the dead end of the room. The reverberation time of the listening room is shown in Fig. 3.8, before and after the curtains were closed. The reverberation time during the experiment was 0.3 ± 0.1 seconds

The loudspeakers used were DSS 930 digital loudspeakers for the Surround Sound reproduction and FB 825 loudspeakers for FL and FR channels. An extra loudspeaker was set up straight behind the subject to check the subject's position.

3.2 Test method

To be able to perform 36 AAB/ABA triads for all combinations of the positions behind the subject, all 3 values off Δ, and the four noise levels (which sums up to 1296 AAB/ABA triads) the subject would have to do a listening test of about 2 hours duration. This is obviously not practical. Therefore, the test was divided into three equal parts of about 45 minutes. The subjects did these three parts on

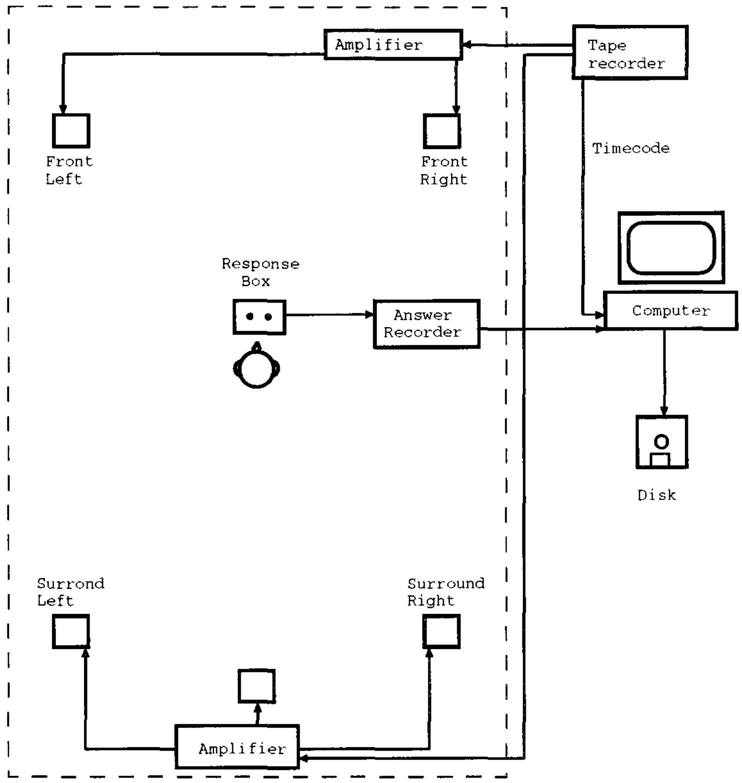

Figure 3.7: *Experimental set-up showing PC and response box setup.*

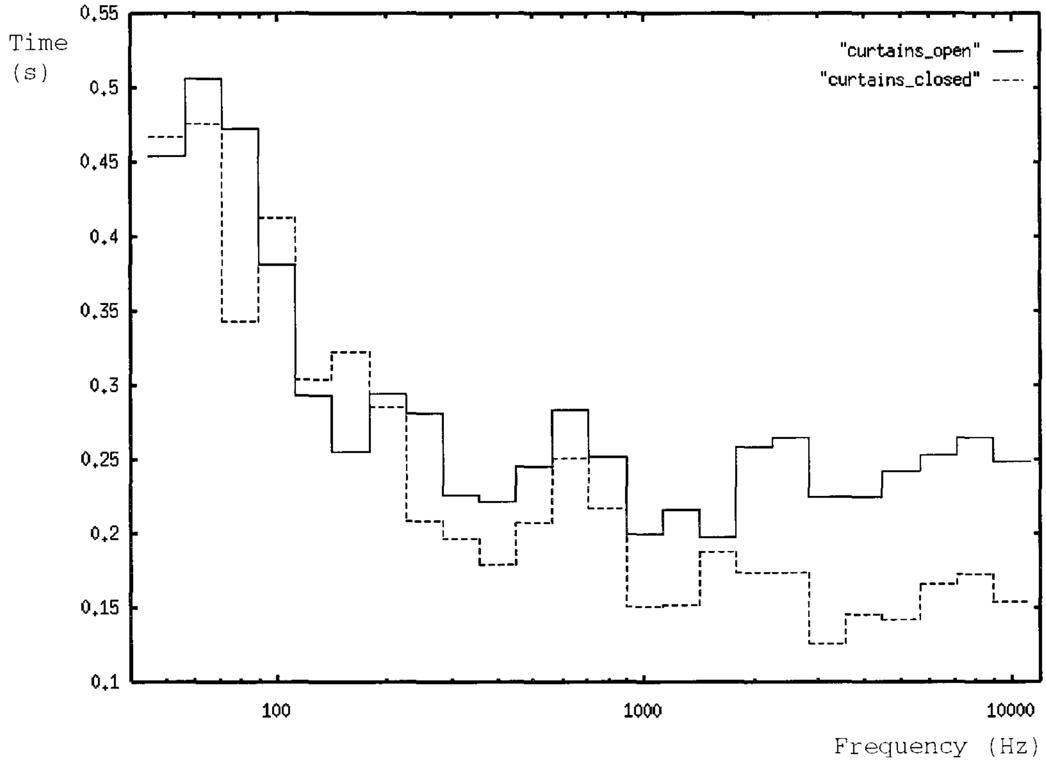

Figure 3.8: *Reverberation time measurement of the listening room used in the localisation test.*

different days to ensure maximum concentration. Each of these parts was divided into four blocks of 108 AAB/ABA triads, lasting for about 10 minutes. The triads were presented in a random order. In between the first and second block of the test, a pause of about one and a half minutes was held. After the second block, the test was interrupted so that the subjects could relax for a few minutes. During this break the 24 track tape had to be rewound, and four different channels had to be connected up. After that, blocks three and four were played with a break of one and a half minutes in between. Each test was performed by one subject at a time.

Each AAB/ABA triad lasted 6 seconds. In that period, the three clicks were reproduced through the surround loudspeakers, the subject had to make his choice on what he thought to be the the right answer, replied, and was shown the correct answer. This is shown in Fig. 3.9

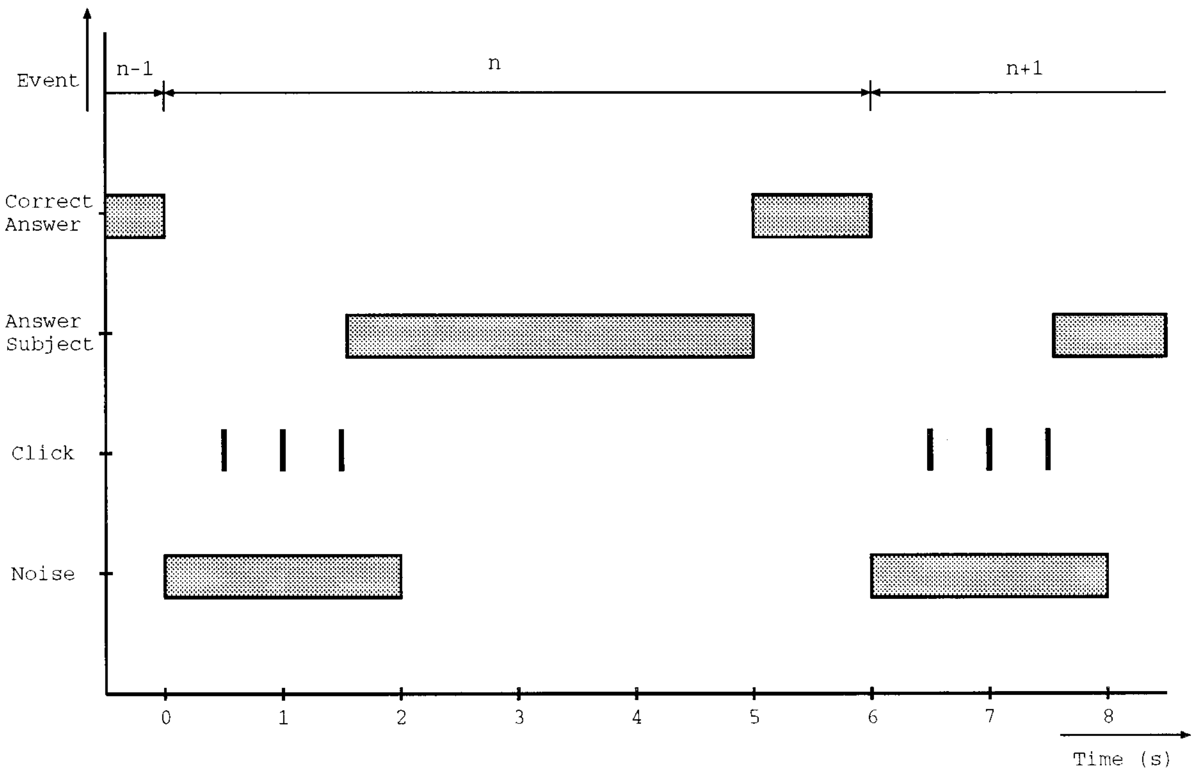

Figure 3.9: *Time schedule of one AAB/ABA triad.*

First the pink noise was played through the FL and FR, for a duration of two seconds. After half a second, the three clicks were played through the surround loudspeakers at intervals of half a second. When the third click had been played the subject was allowed to answer within a period of 3.5 seconds. Two buttons on the response box were marked AAB and ABA. As soon as the subject was permitted to enter his answer, an LED over those buttons would go on. As soon as a button was pushed, the LED over the other button would go off, to indicate the button pushed. The subject could, in the answering time, change the answer by pushing the other button. The answer last given in the answering time was recorded. After expiry of the answering time, the correct answer appeared on the display of the response box.

The answer remained visible for a second.

Before the start of the first test, the subjects were explained their task. They were allowed to do a test-run in order to get acquainted with the test. This test run was performed until the subject felt familiar with the test procedure.

Before each of the three parts of the localisation test, the subject had to listen to clicks being reproduced using either the two surround loudspeakers or a single loudspeaker placed straight behind the listener position. He was asked to change his position in the chair, if necessary, in order to perceive the clicks coming from the real source behind him in the same location as the phantom clicks being reproduced through the surround loudspeakers. After that the test itself began.

Chapter 4

Results and discussion

4.1 Introduction

The localisation test was performed by 10 subjects (all male) varying between 20 and 35 years of age. All had previous experience in performing listening tests. They each did a full test, meaning 3 parts of 4 blocks consisting of 108 triads. During each of the 3 parts of the complete test, each of the subjects answered 3 sets of 4 triads (AAB ABA BBA BAB) for each of the 36 combinations of 4 noise levels, 3 positions behind the subject, and 3 values of Δ. A value of the percentage correct Q_{correct} was calculated for each of these 3 sets of 4 triads, so that for each subject, 3 values of Q_{correct} were available for each of the 36 combinations of noise level, positions, and values of Δ. Both Q_{correct} and the standard deviation in the perceptual Gaussian probability distribution σ (calculated from Q_{correct}) shall be presented.

4.2 Results for percentage correct Q_{correct}

4.2.1 First results

The localisation performance of each different subject has to be seen as a random pick from a large group of possible performances. To provide information on the localisation performance of a population, since the results from this test will be used on a wide scale, it was necessary to average over the subjects. Investigation of the results indicated that the results were not time-dependent, so that it is therefore assumed that a learning effect did not occur. Therefore, the values of Q_{correct} were averaged over the 10 subjects and the 3 parts of the test. The values of Q_{correct} are presented in Fig. 4.1 as a function of noise level, position behind the subject, and value of Δ.

The graphs show that there is no overall dependence of the value of Q_{correct} on the noise level. Nevertheless almost all values of Q_{correct} decrease when the noise level increases from $-\infty$ dB(A) to 40 dB(A), and increase when the noise level is further increased to 50 dB(A). The amplitudes of such tendencies in the graphs are similar in magnitude to the size of the error bars in the values for Q_{correct}. The

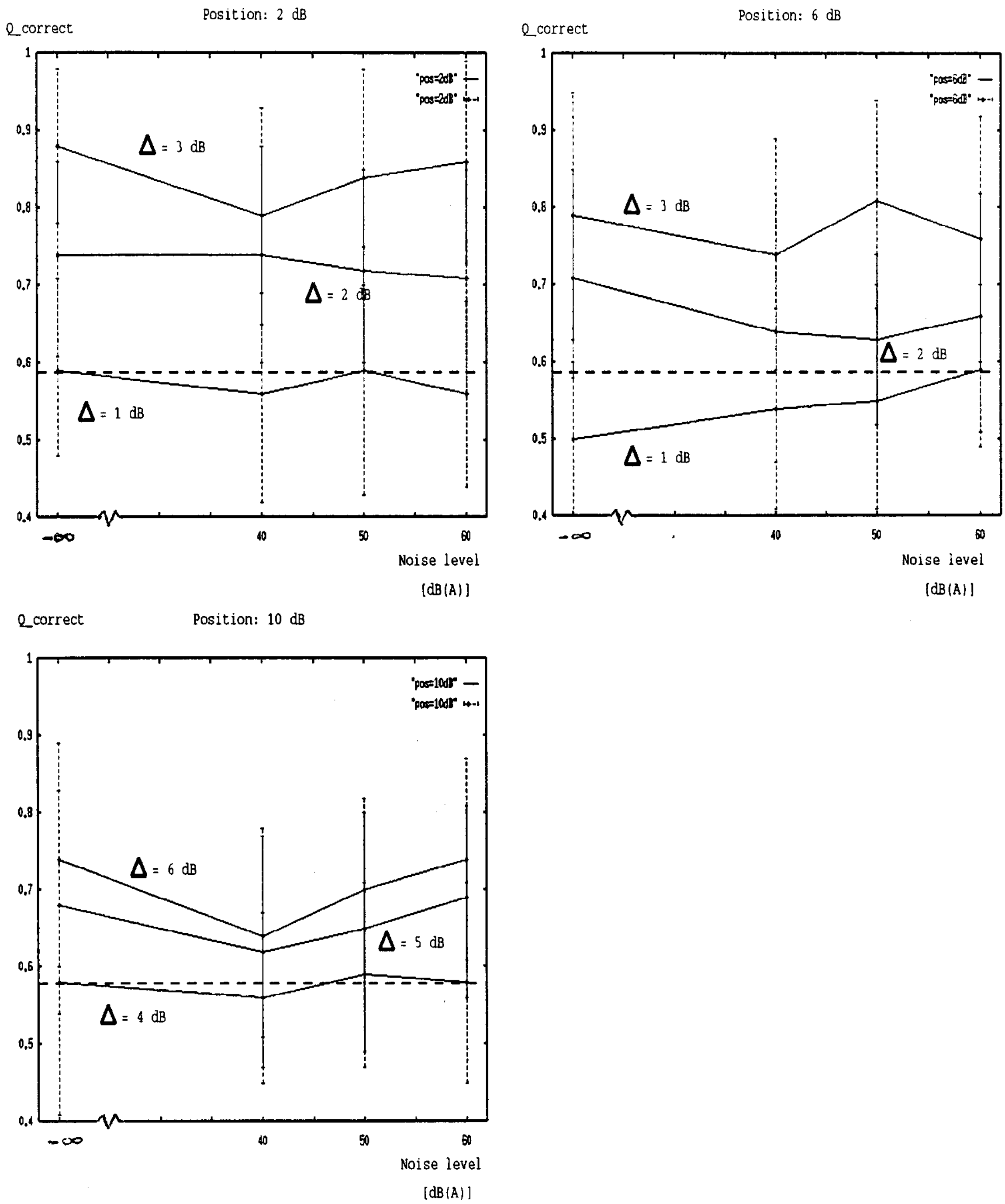

Figure 4.1: *Values of Q_{correct} as a function of noise level and Δ.*

horizontal line in the graph at $Q_{\text{correct}} = 58\%$ indicates the line at which $d' = 1$. This is the perceptual division between localisation and chance.

These tendencies suggest two counteracting processes taking place. First, there is the negative effect the noise level has on the localisation precision. The effect that counteracts this is a masking of reverberation of the clicks by the noise. Higher noise levels mask the reverberation more effectively. The localisation precision is negatively influenced by these reverberations, so that masking them improves precision. As was mentioned in Chapter 3, the setup in the listening room was rotated, and extra damping was applied to decrease the negative influence of reverberations. This negative effect of reverberation was also found in [9].

Fig. 4.1 shows the dependency of the value of Q_{correct} on the value of Δ very clearly. As expected, the value of Q_{correct} increases as the value of Δ increases. The dependency of the value of Q_{correct} on the position behind the subject is also clear. The localisation precision decreases as the position behind the subject becomes further from the central axis. This manifests itself as a decrease of Q_{correct} for positions further from the central axis, or as an increase in the value of Δ necessary to find comparable values of Q_{correct}.

4.2.2 Further investigation of the results

To estimate the influence of the noise level on the value of Q_{correct} more precisely, and to determine the influence of position behind the subject and of the value of Δ in more detail, the statistical method of Analysis of Variance (Anova) was used.

Anova is a statistical method that allows an investigation of the influence of more than one parameter on data. If data are gathered as a function of a number of parameters, it is possible with Anova to find not only the influence of each parameter on the observations, but also the influence of parameters combined. In this case, the parameters of our observations Q_{correct} are noise level (α_i^N), position behind the subject (α_j^P), value of Δ (α_k^D), number of the part of the complete test (α_l^T) and subject (subscript n). Anova allows estimation of the influence of each parameter and combination of parameters on the observations Q_{correct} with the following linear model:

$$\begin{aligned}(Q_{\text{correct}})_{ijkln} &= \mu + \alpha_i^N + \alpha_j^P + \alpha_k^D + \alpha_l^T + \alpha_{ij}^{NP} + \alpha_{ik(j)}^{ND(P)} + \alpha_{il}^{NT} \\ &+ \alpha_{jl}^{PT} + \alpha_{k(j)l}^{D(P)T} + \alpha_{ijl}^{NPT} + \alpha_{ik(j)l}^{ND(P)T} + \epsilon_{(ijkl)n} \qquad (4.1)\end{aligned}$$

Each value of Q_{correct} is thought to be made up of a linear combination of influences of single and combined parameters described with the symbol α, with indices corresponding to the parameters that may influence the values of Q_{correct}. The fact that the parameter Δ is represented as D(P) is a result of the fact that the value of Δ is dependent upon the position behind the subject. This is called nesting. A full explanation of the theory behind Anova and a precise explanation of the meaning of the brackets used in the indices can be found in [10]. The last term in the equation is an error term. This term is assumed to be a random, normally distributed value.

The value of the terms in Eq. (4.1) is determined using Anova by calculating the value of the mean square of the observations that have a certain parameter or combination of parameters in common. Comparing this value with the value of the total mean square of the observations, allows deduction of an indication of the influence of a certain parameter or combination of parameters on the observations can be deduced.

To find out if the values of the terms in Eq. (4.1) describe all variations in the values of Q_{correct}, $\epsilon_{(ijkl)n}$ can be calculated by subtracting the estimated value of Q_{correct} (using the values of the α terms, calculated with Anova) from the observed values of Q_{correct}. A plot of $\epsilon_{(ijkl)n}$ against Q_{correct} should depict a cloud of randomly placed dots.

Anova was performed using STATA, a program specially designed for statistical investigation of data sets. The results are listed in table 4.1, and the plot of the residue $\epsilon_{(ijkl)n}$ against Q_{correct} is shown in Fig. 4.2.

Parameter	Partial SS	df	Mean Square	F	Prob > F
test	0.1435	2	0.0717	3.68	0.0256
noise	0.2785	3	0.0929	4.76	0.0027
position	0.9131	2	0.4566	23.40	0.0000
position×Δ	8.4988	6	1.4165	72.59	0.0000
test×noise	0.0739	6	0.0123	0.63	0.7057
test×position	0.1212	4	0.0303	1.55	0.1848
test×position×Δ	0.1398	12	0.0116	0.60	0.8459
noise×position	0.0822	6	0.0137	0.70	0.6482
noise×position×Δ	0.4771	18	0.0265	1.36	0.1441
test×noise×position	0.1321	12	0.0110	0.56	0.8717
test×noise×position×Δ	0.4044	36	0.0112	0.58	0.9793
residual	18.9668	972	0.0195		
total	30.2315	1079	0.02801		

Table 4.1: *Results of Anova calculations using STATA on values of* Q_{correct}.

The value of F indicates the influence of the parameter, for which it was calculated, on the total variance of the observations. The higher the value of F, the more important the influence of that parameter or combination of parameters on the total variance. The significance level Prob > F indicates at what level this value of F is significant. It is defined as the probability of occurrence of a value of F larger than or equal to the value listed, assuming the parameter for which it is calculated to be, in reality, of no influence on the total variance of the observations. The value of F indicates that there is an interaction, but does not show its extent. The parameters, noise, and number of the part of the test have a high level of significance, although not as high as that of position and Δ.

The values of Q_{correct} suggest that the number of the part of the test only has a small influence.

Anova indicates that the influence of the noise level on Q_{correct} is significant.

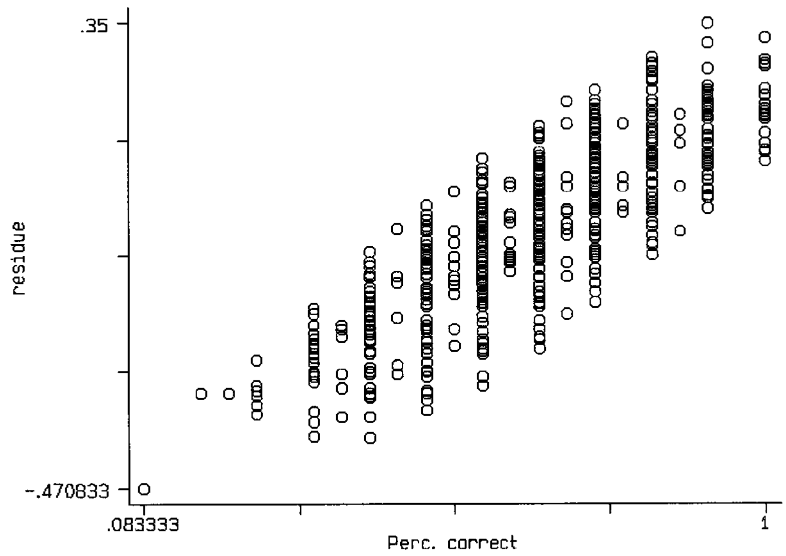

Figure 4.2: *Plot of the residue* $\epsilon_{(ijkl)n}$ *against the value of* Q_{correct} *for which it was calculated.*

However, Anova cannot differentiate between separate positive and negative effects caused by a single influence, but can only assess the overall effect. This suggests that the magnitudes of the separate effects are equal and opposite, therefore producing a total effect which results in a level of significance that is not as high as that of position and Δ.

Anova reveals the largest influences to be those resulting from Δ and position. The influence of position on localisation precision is already known [3, 11]. The influence of Δ can be explained as follows: the values of Δ were chosen as a result of preliminary tests, such that they lay at the threshold of perception. Due to this, one expects the value of Q_{correct} to vary greatly[1], since the range of values of Δ causes differences varying from imperceptible to fully perceptible, i.e. Q_{correct} varying from 50 % to 100 %.

The plot of the residue $\epsilon_{(ijkl)n}$ against the value of Q_{correct} is not shaped like a cloud of dots. An effect causing lower values of Q_{correct} to be over-estimated and higher values of Q_{correct} to be under-estimated remains. As Anova cannot describe all the influences on the set of observations, the values found can only be used as indicators. Therefore, too much weight should not be attached to the conclusions drawn from the Anova results.

The results found with Anova are in line with the conclusions drawn from the Q_{correct} graphs.

Although only a small number of the values for Q_{correct} reach a 100% level, and no value reaches 0%, it can be argued that a model which uses a normally distributed error term cannot be used on observations that consist of percentages. Near the extreme values, values for Q_{correct} will not be normally distributed. Percentages should therefore be treated with a special form of weighted Anova. Considering

[1]A similar analysis using Anova on the value of σ as defined by Eq. 2.14, indicated a the influence of Δ on σ was considerably smaller

the values of the residuals, ϵ, and the fact that this more complicated form of Anova will only correct for the most extreme values of Q_{correct} (as Q_{correct} results from perception, in which many stochastic processes usually add up to a normal distribution) it was decided not to calculate this form of Anova, but rather to use the results calculated with STATA as presented above.

4.3 Results for σ

The values of Q_{correct} can, using Eqs. (2.13), (2.14), be translated into values of σ, i.e. values of the standard deviation in our perceptual model of localisation.

The calculation of the values of σ had to be partially performed numerically, because analytical inversion of Eq. (2.13) is not possible. This introduces extra uncertainty in the values of σ, particularly for values of $d' < 1$. Therefore, the values of Q_{correct} were averaged over the 4 noise levels to increase the number of observations. This averaged $\overline{\sigma}$ will have an error that is a factor of two smaller than the error in the individual values of σ for each noise level (standard error of the mean).

The calculated values of $\overline{\sigma}$ are listed in table 4.2. The error bars in $\overline{\sigma}$ are calculated as half the error in the four separate values of σ. These latter errors were equal, within 1 % of each other. They were derived from the values of σ for [Q_{correct} $\pm$ the standard deviation in the value of Q_{correct}].

Position (dB)	Δ (dB)	$\overline{\sigma}_{\text{min}}$ (dB)	$\overline{\sigma}$ (dB)	$\overline{\sigma}_{\text{max}}$ (dB)
2	1	0.73	1.05	3.33
	2	0.87	1.07	1.36
	3	0.89	1.12	1.37
6	1	0.85	1.36	-
	2	1.03	1.34	1.92
	3	1.08	1.37	1.75
10	4	2.90	4.19	12.7
	5	2.60	3.37	4.85
	6	2.76	3.44	4.48

Table 4.2: *Values of σ calculated from values of Q_{correct}.*

According to the perception model employed, the value of σ is independent of the value of Δ. This is illustrated vividly by the values of σ, which are so similar, particularly for the first two positions behind the subject, that the values of the errors in σ seem greatly exaggerated.

The value of σ is higher for positions further from the central axis. This is in accordance with previous research [3]. When moving away from the central axis, the value for σ seems, at first, to increase slowly. However, the rate of increase grows as one moves toward the lateral position.

4.4 Comments

During breaks in between the second and third block of a test, and after each part of the test, the subjects were asked if they noticed anything while doing the test that might be of interest.

One of the subjects noted front/back reversal at noise levels of $-\infty$ and 40 dB. The clicks seemed to be sounding in front of him rather than behind him.

Another subject stated that the clicks contained a frequency cue. It seemed to him as if the second click had a higher pitch than the first and third.

When asked, none of the other subjects noted any of these effects.

None of the subjects considered the test to be easy. They all thought they performed better when no noise was reproduced, their performance gradually worsening as the level of the noise increased.

Chapter 5

Conclusions

A test has been designed to measure the ability of a subject to localise phantom clicks produced behind him, as a function of noise generated from the front. The results of this test neither showed a marked dependency nor did they fully rule out any influence. However, the errors in these results are too large to enable a full analysis of the extent of any influence. From the tendencies in these results, it was concluded that noise has a negative influence on the localisation precision. However, such negative influence is counteracted when noise masks reverberation of the clicks.

The parameter representing the localisation precision in the employed perception model is the standard deviation σ of the perceptual localisation Gaussian curve. The values found for σ increase as one moves away from the central axis. This is in accordance with previous research into this subject.

Although the subjects found localisation more difficult with higher noise levels, the statistical analysis of the results did not reveal such an influence.

To determine the exact relationship between localisation precision for clicks reproduced behind a subject and the noise reproduced in front of the subject, a test will have to be devised which will incorporate reverberation as a possible parameter.

The fact that the two test signals differ so greatly may cause perceptual streaming mechanisms to prohibit a large influence of the noise on the localisation of clicks. Repeating the experiment with more similar signals may be revealing.

Appendix A

Apparatus listing

Device	*Model*
Loudspeaker	Philips DSS 930
	Philips FB 825
Amplifier	Philips DSC 950
	Philips FA 890
Realtime Frequency Analyser	B&K 2123
Right Ear Simulator	B&K 4158
Left Ear Simulator	B&K 4159
Head and Torso Simulator	B&K 4128
Precision Sound Level Meter	B&K 2232
Dynamic Signal Analyzer	HP 3562 A
Computer System	HP 9000/300
24 track Mixing Desk	Studer
Multi Channel Recorder	Studer A820

Bibliography

[1] *Multi-channel sound for HDTV: first subjective test results*, Philips Internal report, EU95 PG11 WPS, 1992.

[2] D.M. Green, J.A. Swets, *Signal detection theory and psychophysics*, Wiley, 1966.

[3] J.Blauert, *Spatial hearing: The psychophysics of human sound localization*, page 41, The MIT Press, 1989.

[4] S.L.J.D.E. van de Par, *On* $2\frac{1}{2}$*-channel sound compared to three-channel sound*, Masters thesis, Eindhoven University of Technology, 1992.

[5] Sound Quality Assessment Matarial (recordings for subjective tests) CD. Cat. N° 422 204-2, 1988.

[6] R. Rebscher, *Untersugung zur Verbreitung der Hörzone durch Erhöhung der Lautsprecheranzahl innerhalb der Stereo-Wiedergabe.* Masters thesis, Fachhochschule Düsseldorf, 1989.

[7] W.M. Hartman, B. Rakerd, *On the minimum audible angle—A decision theory approach*, J. Acoust. Soc. Am. **85**(5) pp 2031-2041, 1989.

[8] J. Blauert, *Ein Beitrag zur Trägheit des Richtungshörens in der Horizontalebene*, Acustica, **20** pp 200-206, 1968.

[9] C. Giguère, S.M. Abel, *Sound localization: Effects of reverberation time, speaker array, stimulus frequency, and stimulus rise/decay*, J. Acoust. Soc. Am. **94** (2) pp 769-776, 1993.

[10] D.C. Montgomery, *Design and analysis of experiments*, Wiley, 1991.

[11] A.W.Mills, *On the minimum audible angle*, J. Acoust. Soc. Am. **30** pp 237-246, 1958.

Dankwoord

Met dit afstudeerprojekt eindigt een belangrijke periode in mijn leven. Vele mensen hebben mijn studententijd tot een zeer speciale gemaakt. Ik zou hen willen bedanken voor hun vriendschap, gezelligheid en hulp. In het bijzonder wil ik de volgende personen bedanken.

Mijn ouders en zus

Erna

Nigel, zonder wie ik het niet gered zou hebben

Ard Jan, Sonja

Eric, Gerd Jan, Charrel
Rob, Erwin, Will, Spock, Scotty
Philip, Tunes, Linda

Special thanks to my friends
Andrew, Russell, Oliver, Pierre, Carsten