The Influence of Phase Distortion on Sound Quality

Master Thesis by Group 1064
Section of Acoustics
Department of Electronic Systems
Aalborg University
2007

Department of Electronic Systems

Aalborg University

Section of Acoustics

THEME:
Acoustics

SUBJECT:
Master's Thesis

TITLE:
The Influence of Phase Distortion on Sound Quality

PROJECT PERIOD:

1. February 2007 -
07. June 2007

GROUP:
1064

GROUP MEMBERS:
Anton Günther
Yesenia Lacouture Parodi

SUPERVISOR:
Alex Karandreas

NUMBER OF DUPLICATES: 4.

NUMBER OF PAGES IN REPORT: 66.

NUMBER OF PAGES IN APPENDIX: 50.

TOTAL NUMBER OF PAGES: 116.

In the past years some studies had been directed towards the audibility thresholds of phase distortion and in general, electronically generated signals have been used. It has been proven that phase distortion is audible but little attention has been paid to its influence on sound quality. This thesis investigates the influence of phase distortion in the perception of sound quality. The work is focused on the perception of phase distortion in a normal listening environment with loudspeakers.

Two subjective listening test were implemented. For the first one, an adaptive algorithm was designed in order to find phase distortion audibility thresholds for a set of sixteen stimuli, among which were speech, instrumental music, electronic generated sounds and pop music. The purpose was to select those signals which showed to be more sensitive to phase distortion. The selected signals were used in the second listening test. This test consisted of a paired comparison experiment and its purpose was to assess the listener preferences regarding the phase distortion. A training session was implemented in order to teach the differences to the listener. This session was ran before the preference parts.

Results showed that phase distortion is not audible for most of the evaluated stimuli. Stimuli such as speech and electronic generated signals showed lower thresholds when played through headphones. Training showed to be a key factor on the phase distortion perception. From the preference test it was found that for speech some subtle distortion is rated similar to no distortion but when perceivable, distortion is not accepted at all. In the case of the electronic generated signal a monotonically decreasing scale was obtained, where the non-distorted sample was ranked highest and the most distorted lowest. An electronic music sample from a commercial record was evaluated as well. It was observed that distortion had a subtle effect through headphones, but there is no significant difference in the ranking.

In general, phase distortion showed to be not audible through loudspeakers in a normal listening environment for any of the selected signals. No significant differences were found in the results and subjects showed difficulties trying to discriminate the distorted from the original.

PREFACE

This thesis provides the documentation supporting the master work done by the 10th semester group 1064, Section of Acoustics at the Department of Electronic Systems, Aalborg University, 2007.

The main purpose of this thesis was to evaluate the influence of phase distortion in the perception of sound quality through loudspeakers in a normal listening environment. Two subjective listening experiment were performed to evaluated the audibility of phase distortion and its influence on sound quality.

The target audience of this document are the supervisor, censor, 10th semester students and people who have a general interest in sound quality. Acoustic, signal processing and statistic knowledge is advised. References are shown with a number from bibliography. When the page number is relevant this is included, e.g. [28, p. 22].

This document is divided into two main parts: the report and the appendices. The report consists of six chapters. The chapters provides an introduction, an analysis of related theories, an review of previous studies, measurements description and results, the design of threshold listening test and results, the design of the main test and results, a discussion and conclusions. The appendices includes the measurement reports, instruction sheets and questionnaires, audiograms, main test setup and detailed main test results.

A CD is enclosed containing:

- Matlab code for listening test interfaces and algorithms
- Audio samples
- The report in PDF format

Aalborg University, 1^{st} of June, 2007

Anton Günther   Yesenia Lacouture Parodi

Contents

CHAPTER 1

INTRODUCTION

In the past years effort has been made to enhance the sound quality of audio systems. Most of the research has been directed into minimizing the effects of the audio chain by reducing the distortion and noise introduced by them. An ideal sound system should not influence or modify the characteristics of the reproduced acoustic signals. However, that ideal case is not possible to achieved and the systems always modify the signals to some extend. When the wave shape or the spectral content of the signals is affected by the system, this is called distortion.

Some progress has been accomplish in the latest years regarding this. Nowadays, sound systems add very little distortion to the reproduced sound. Nevertheless, loudspeakers' characteristics are more complicated to approach and still introduce unwanted artifacts. Moreover, most of high quality loudspeakers are built with more than one driver in order to manage all frequencies. The frequency spectrum is usually split into three bands: low, midrange and treble. This is done by means of a crossover network which consists in a set of band-pass filters with flat frequency response. This crossover networks are not ideal and some distortion can be observed around the crossover frequencies. There are some techniques to counteract for distortion effects, but most of them are focused on achieving flat magnitude responses disregarding the phase.

Phase distortion is defined as a degradation of the waveform of a signal without altering its spectral content [34]. Few research has been carried out on phase distortion perception. In the beginning, it was stated that the ear was phase-deaf and that a flat magnitude response was enough to diminish the distortion introduced by the systems [29], [30], [15]. However, some studies have demonstrated that the phase distortion is clearly audible under certain conditions and certain kind of stimuli [24] [17] [21] [25].

Phase distortion has been the concern of some audio developers and researchers. Most of the studies had focused on whether or not the phase distortion is audible and which variables might influence this audibility. Up to now, results show that the phase distortion is audible under certain conditions and this suggests that there is a possible influence in the sound quality perception of the reproduced sound. These studies had used synthetic signals mainly and many of the subjective tests had been carried out with experienced listeners.

There are some who argue that a linear phase loudspeaker is needed in order to have a high reproduction quality [3], whereas others state that the phase distortion causes only a subtle effect in the signal and it can be disregard in the design of audio systems [24]. So far none of both argument has been conclusively demonstrated. In order to give a step forward on phase distortion research it would be desirable to study the perceptual effect the phase distortion has on the reproduction quality.

Project Goal

The goal of this thesis is to assess the effects of phase distortion introduced by loudspeakers in the perception of the quality of the sound. The research will not be focused on whether the distortions are audible or not, but rather on whether these distortions actually affects the perception of the reproduction quality. An attempt to measure the amount of phase distortion necessary to degrade the quality perception of signal, keeping the magnitude constant will be done. In order to give a more realistic approach the research will be directed to “normal” environments. This is, the assessment will be focused to typical listening situations. The objective of this thesis can be summarized in one question:

- Does phase distortion introduced by loudspeakers affects the perception of the reproduction quality?

Project Scope

To get a good insight into the phase distortion properties and its effects a number of measurements and comparisons will be carried out. Subjective tests will be designed to assess preference, measure the perceived characteristics and audibility thresholds. These tests will be run through headphones as well as loudspeakers. The purpose of the former is to have a better control over the variables and the later to recreate a typical listening situation. Additionally, an objective analysis of the phase properties will be done in order to arrive to an appropriate evaluation of the variables. The next items describe some general steps that will be followed along the project development.

- Measurement of loudspeakers’ and headphones’ impulse responses.
- Research of different kind of filters typically used in crossover networks.
- Design of a test to assess which kind of common acoustical signals are more sensitive to phase distortion. An attempt to find audibility thresholds will be carried out.
- Design of a subjective test to measure the tolerance of audible phase distortion and preference.

This report is divided into six chapters. The second chapter contains an overview over the theory behind phase distortion, phase perception, crossover networks and a review of previous researches is presented. Finally, a problem statement is formulated. The third chapter describes the measurements carried out and the criteria applied to select the loudspeaker and headphones, which were used in the listening test. The fourth chapter depicts the listening tests designed to select the stimuli and to assess the preferences. The fifth chapter shows the results obtained in the preference test and describes the probabilistic choice model used. At last, a general discussion and the conclusions of this thesis are presented.

CHAPTER 2

THEORY

This chapter explains the theoretical background of phase distortion in terms of mathematical definitions and filter theory. Further, the process of human sound perception is outlined to understand how these distortions could be processed in the auditory system. Additionally, a review of previous studies on the audibility of phase distortions will be given. Finally, this chapter leads to the problem statement of this thesis.

2.1 Human Hearing

This section describes the human ear and the basic hearing process and shows how phase differences could be perceived by humans. Healthy ears of young people can hear sound in a frequency range from approximately 20 Hz to 20 kHz, but the upper limit of this range decreases with age. The human ear is most sensitive in the frequency range of human speech, approximately from 200 Hz to 8 kHz. The frequency range of common music is slightly wider.

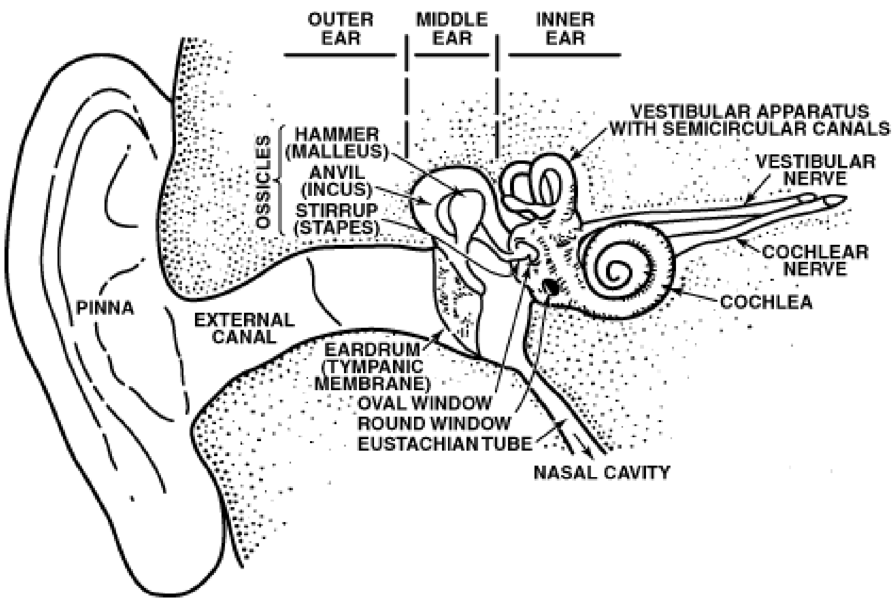

Figure 2.1: *The auditory system. [10].*

The human ear consists of three main parts, the outer ear, the middle ear and the inner ear (see figure 2.1). The pinna and the ear canal form the outer ear. The pinna reflects and attenuates the sound waves depending on the angle of incidence and their frequency and leads them into the 6-8 mm long ear canal. This modification of the incoming sound helps to localize the sound source.

At the end of the ear canal is the tympanic membrane or eardrum located, which vibrates when sound waves hit it. This vibrations are transmitted from the membrane to the ossicles, which are three small bones in the middle ear. The main function of the middle ear is an impedance matching of the air with the liquid in the inner ear. This is done with the lever of the ossicles and the areas' ratio between the tympanic membrane and the oval window. The oval window is the entrance membrane of the inner ear and is connected with the other end of the ossicles. Due to this impedance matching transmission at middle frequencies from 500 Hz to 4 kHz is most efficient. Moore [28, p. 22 f.] mentions further studies about possible additional functions of the middle ear. For instance, the ossicles may also reduce the influence of bone-conducted sounds from internal sources as chewing or flowing of blood transmitted to the inner ear and a reflex of muscles at the ossicles may protect the inner ear from damages.

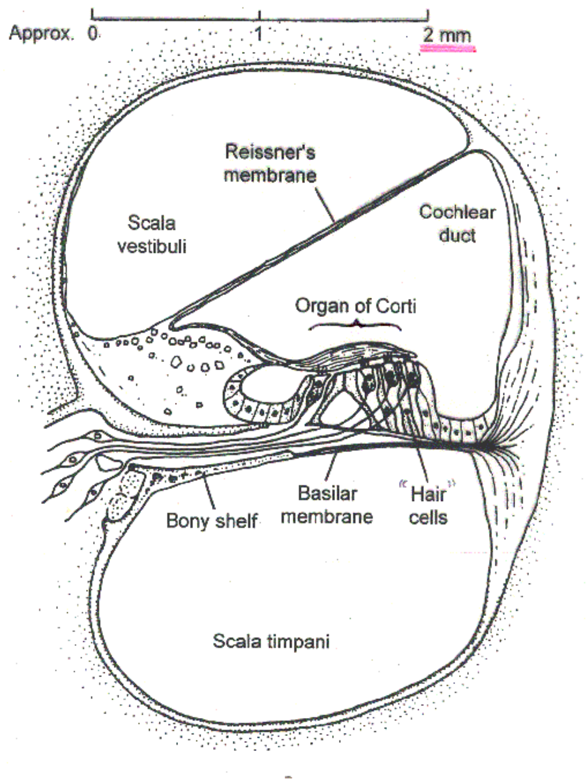

Figure 2.2: *Cross section of the cochlea. From [41].*

The inner ear consists of the cochlea, a snail-shaped tube with 2½ windings, where the sound pressure waves are transformed into nerve impulses. The cochlea is separated along its length by Reissner's membrane and the basilar membrane into several ducts (figure 2.2). When a sound wave hits the oval window, this wave travels through the scala vestibuli, reaches the hole in the basilar membrane at the end of the cochlea, the helicotrema, and travels back through the scala tympani. At the end of this duct lies the round window, where the wave leaves the cochlea.

While traveling through the cochlea, the sound wave displaces the basilar membrane. This displacements have their maximum at characteristic positions depending on the frequency of the wave. Higher frequencies displace the basilar membrane closer to the oval window, because the membrane is narrower there and more stiff than far away from the oval window. Lower frequencies have their resonance frequencies on the basilar membrane closer to the helicotrema. Therefore, the inner ear works mainly like a fourier analyzer which splits complex sounds into their spectral components. The frequency which displaces a certain point on the basilar membrane maximal is called the characteristic frequency of this point.

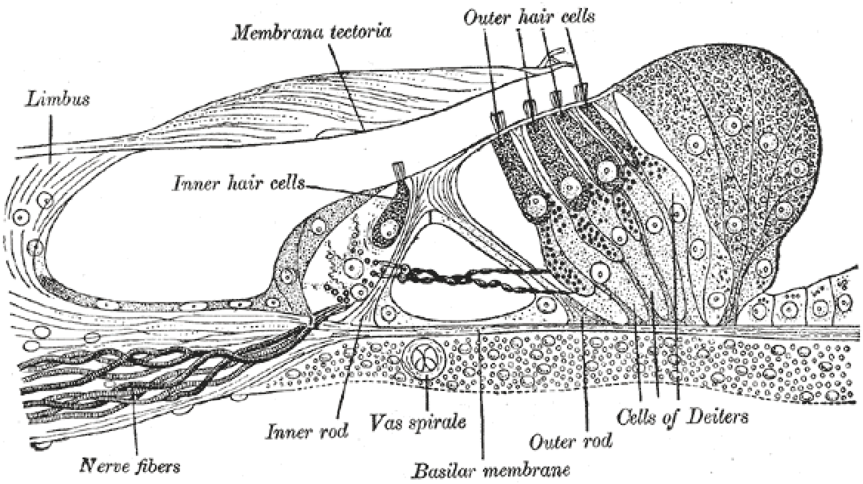

Figure 2.3: *Section through the spiral organ of Corti. Hairs are not pictured. From [44].*

The basilar membrane supports the organ of Corti, the sense organ of hearing, which has different hair cells connected with nerve fibers (figure 2.3). The hairs of the 3-5 rows of the outer hair cells are connected with the tectorial membrane and amplify the displacement of the basilar membrane. The hairs of the single row of inner hair cells, which are not connected with the tectorial membrane, transform the mechanical vibrations into electrical nerve impulses and send them to the brain.

Moore mentioned in [28, p. 37] that the nerve cells fire also spontaneous impulses without stimuli. Above a certain threshold of excitation the nerves fire more electrical impulses when the level of the stimuli increase. The nerve cells have also an upper limit in their firing rate, where the cells are saturated and do not increase their firing rate with increasing stimuli level. The range between these thresholds is called the dynamic range of a nerve cell. The dynamic ranges of three different types of nerve cells with different thresholds form the dynamic range of hearing.

Additionally, Moore describes in [28, p. 42 ff.] an effect called phase locking. For tones with frequencies below 4-5 kHz the nerve cells fire impulses according to the cycles of the stimuli. This means, the nerve's spikes occur roughly at the same phase of the stimuli's waveform each time. They occur not necessarily at each waveform cycle, but the time intervals between the spikes are approximately integer multiples of the period of the stimulating signal. This effect has not been observed for higher frequencies. It is still a controversy which kind and to which extend these effect is influencing the perception of sound.

The processing of complex auditory stimuli in the auditory cortex is still not very well known and subject of research. Moore mentions in [28, p. 48 ff.] studies which indicate, that some cortical neurons do not respond to single tone stimuli but more complex stimuli as clicks, noise bursts, repeating patterns or stimuli changing with time.

Apart from the phase locking effect, there is not certainty on the way the ear processes the phase information of the sound. Thus far, it is unknown how the phase distortion is detected but it is proven it is audible under certain conditions. Schroeder shows in [36] audibility of phase is not a simple analysis of the envelope of the signal's waveform. That leads to the assumption, that phase changes are recognized by means of higher processing in the brain using the timing-information of the nerve's phase locking in the organ of Corti among others.

2.2 Mathematical Definition of Phase Distortion

This section describes the mathematical background of distortions in a transmission system and defines the parameters delay, phase delay and group delay. Additionally, the terms differential time-delay distortion and differential phase-shift distortion are introduced.

Distortion is defined as the alteration of the original shape of a waveform. A causal, linear and time-invariant system which is not distorting signals, can be described in the time domain with its impulse response $h(t)$ as follows:

$$h(t) = K\delta(t-T), \tag{2.1}$$

where δ is the unit impulse and the constants $K > 0$ and $T \geq 0$. The Fourier transformation of $h(t)$ leads to the complex frequency response $H(\omega)$. The impulse response $h(t)$ can be calculated from the frequency response $H(\omega)$ with the inverse Fourier transformation respectively:

$$H(\omega) = \int_{-\infty}^{\infty} h(t)e^{-j\omega t}\mathrm{d}t \tag{2.2}$$

$$h(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} H(\omega)e^{j\omega t}\mathrm{d}\omega. \tag{2.3}$$

The complex frequency response of a system can generally be written in polar form as

$$H(\omega) = |H(\omega)|\, e^{j\phi(\omega)} \tag{2.4}$$

with the magnitude response $|H(\omega)|$ and the phase response $\phi(\omega)$, which characterize the system completely. Inserting the impulse response from equation 2.1 on the preceding page in equation 2.2 on the facing page gives

$$H(\omega) = Ke^{-j\omega T}. \tag{2.5}$$

Comparing equation 2.4 with equation 2.5 leads to the restrictions of an undistorted transmission system. Hence, the magnitude response of the system must be constant $|H(\omega)| = K$ and the phase response linear dependent from the frequency $\phi(\omega) = -\omega T$. If one of these restrictions are not fulfilled, the signals are distorted by the system. Phase distortion happens when the later restriction is not complied. Since this work is investigating the influences of phase distortions on audio quality the magnitude will be kept constant and only the phase will be manipulated.

The output signal is not phase distorted if only a pure time shift τ_d over all frequencies occurs, because the output signal is then just an exact time-shifted copy of the input signal. In this case the phase response of the system is a pure time delay:

$$\phi(\omega) = -\omega\tau_d \tag{2.6}$$

To get a better understanding on how much a signal is being distorted, some additional parameters are usually defined. The phase delay $\tau_p(\omega)$ is given as

$$\tau_p(\omega) = -\frac{\phi(\omega)}{\omega} \tag{2.7}$$

and indicates a time shift of single frequencies of the signal. Further, the group delay $\tau_g(\omega)$ is defined as

$$\tau_g(\omega) = -\frac{\partial\phi(\omega)}{\partial\omega} \tag{2.8}$$

and can be interpreted as a time shift of the envelope of the signal dependent on the frequency. With these definitions a system has no phase distortion if $\tau_d = \tau_p(\omega) = \tau_g(\omega)$. Figure 2.4 on the next page illustrates the phase delay and the group delay for an amplitude-modulated wave.

Additionally, Leach defines in [18] the differential time-delay distortion $\Delta\tau$ as the difference of phase delay and group delay,

$$\Delta\tau = \tau_p - \tau_g. \tag{2.9}$$

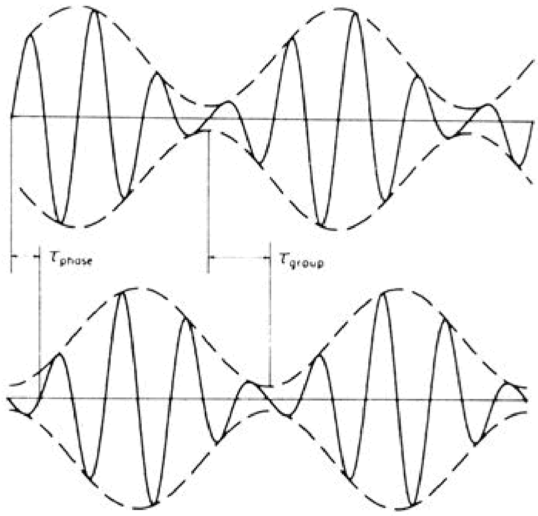

Figure 2.4: *Phase delay and group delay for an amplitude-modulated wave. From [34, Fig. 1].*

Consequently, $\Delta\tau$ measures the system's deviations from a system with a pure time shift which has a $\Delta\tau = 0$. The according phase angle in degrees

$$\Delta\phi = \frac{180}{\pi}\omega\Delta\tau \tag{2.10}$$

is called the differential phase-shift distortion and shows how much a system differs from the linear-phase system.

2.3 Phase Characteristics of Causal Systems

The following section will explain different types of filters and transmission systems. First, it is shown systems can be split up in a minimum-phase and an excess-phase part. Following, the terms linear-phase system and all-pass system are explained.

2.3.1 Minimum-Phase Systems

All causal systems have a minimum amount of phase shift $\phi_{min}(\omega)$ necessarily related to their magnitude response $|H(\omega)|$ [34]. This minimum phase shift and the magnitude response are related with the Hilbert-transform relation

$$\phi_{min}(\omega) = \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{\ln|H(\omega')|}{\omega' - \omega}\mathrm{d}\omega'. \tag{2.11}$$

Systems with only a phase shift $\phi(\omega) = \phi_{min}(\omega)$ are called minimum-phase systems.

Oppenheim et al. [32, p. 280] explain a minimum-phase system as a system which is causal and stable and has a stable and causal inverse system. Because the criterion for stability and causality are that the system's poles are inside the unit circle in the z-plane and the zeros are the poles of its inverse system, a minimum-phase system must have both poles and zeros inside the unit circle. They show that these systems have a minimum phase-lag, a minimum group delay and a minimum energy-delay [32, p. 287 ff.].

When reflecting one or more zeros of a minimum phase system on the unit circle to their conjugate reciprocals outside the unit circle, a nonminimum-phase system with the same magnitude response as the minimum-phase system is formed [32, p. 281].

2.3.2 Excess-Phase Systems

A causal, stable nonminimum-phase system can be split up in a minimum-phase component which contains the magnitude information, and an additional excess-phase or all-pass component which influences only the phase of the system:

$$H(s) = H_{min}(s) \cdot H_x(s). \tag{2.12}$$

Transforming this equation in the s-plane to the frequency domain with $s = \mathrm{j}\omega$ and only looking at the phase response of the system, it can be split up in a minimum phase part and an excess phase part:

$$\phi(\omega) = \phi_{min}(\omega) + \phi_x(\omega). \tag{2.13}$$

A practical definition (e.g. in [34], [21]) for the excess phase part is

$$\phi_x(\omega) = \theta_a(\omega) - \omega T + \theta_0 \tag{2.14}$$

where θ_0 is constant, for example caused by a polarity reversal, and $\theta_a(\omega)$ the frequency dependent phase shift of a nontrivial all-pass network with $\theta_a(0) = 0$. The pure time delay in equation 2.14 is represented with $-\omega T$ as in equation 2.6 on page 9.

Therefore, it can be easily seen that $\phi_{min}(\omega)$ from equation 2.13 and $\theta_a(\omega)$ and θ_0 from equation 2.14 contribute to the phase distortion.

2.3.3 Linear-Phase Systems

For an undistorted transmission through any system is a zero phase response desirable, but not attainable for causal systems. Therefore, it is a good compromise to allow a linear phase shift $\phi = -\omega T$, which introduces only a delay (as in equation 2.6 on page 9), but is not changing the envelope of the signal [32, p. 291 ff.]. Such systems with a linear phase response are called linear-phase systems and have a frequency response

$$H(\omega) = |H(\omega)|e^{-j\omega T} \tag{2.15}$$

with $T \geq 0$. From equation 2.8 on page 9 it follows that linear phase systems have a constant group delay.

2.3.4 All-Pass Systems

Systems with a constant magnitude response but a frequency-dependent phase-response are called all-pass systems. Therefore, they have a frequency response

$$H(\omega) = Ke^{-j\phi(\omega)} \tag{2.16}$$

where K is a constant. All-pass systems have all poles inside the unit circle in the z-plane and all zeros are conjugate reciprocals of the poles mirrored to the outside of the unit circle. In the complex s-plane all zeros are in the right half-plane and all poles are mirrored to the corresponding positions in the left half-plane .

First-order all-pass systems have in the complex s-plane the general transfer function

$$H(s)_{AP1} = \frac{s-\omega_0}{s+\omega_0}, \tag{2.17}$$

where ω_0 is the center frequency of the all-pass. The transfer function of second-order all-pass systems is defined as

$$H(s)_{AP2} = \frac{s^2 - \frac{\omega_0}{Q}s + \omega_0^2}{s^2 + \frac{\omega_0}{Q}s + \omega_0^2} \tag{2.18}$$

with ω_0 the center frequency and Q the quality factor. If a second-order all-pass has a $Q \leq 1/2$ it turns into two first-order sections [25]. Figure 2.5 on the facing page shows the phase and group delay of different second order all-pass filter centered at 1kHz, with Qs equal to $1/2$, $1/\sqrt{2}$, $\sqrt{2}$ and 4.

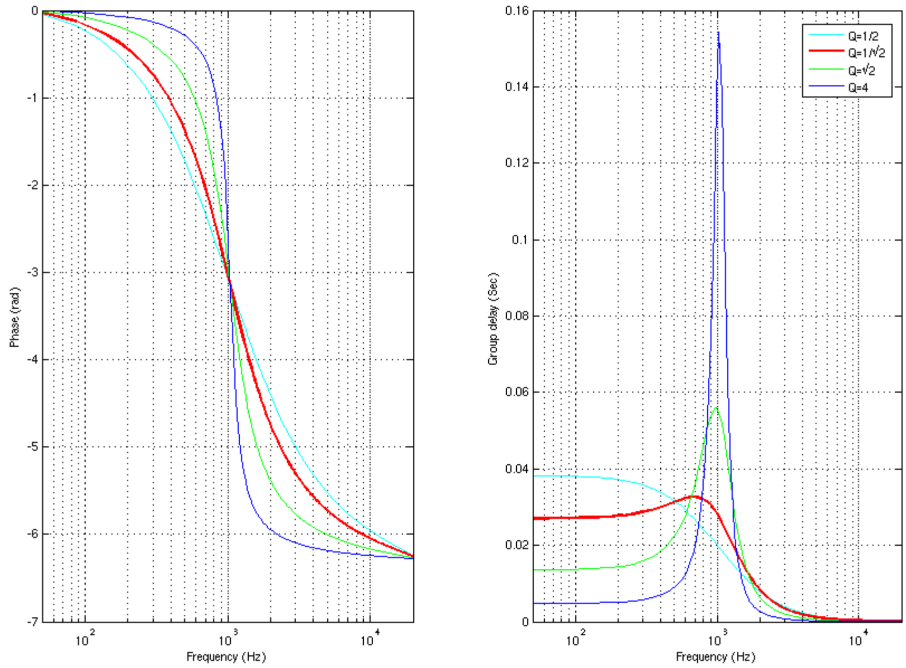

Figure 2.5: *Phase and group delay response of all-pass filters with* $f_o = 1Hz$.

It can be seen from the figure that the higher the Q the steeper is the change in phase around the center frequency and the higher the peak group delay.

2.4 Crossover Networks

In general loudspeakers are built with more than one driver. The reason for this is that the non-ideal physical characteristics of a single driver, such as resonances or bandwidth restrictions, make it incapable to handle frequencies within a wide range. In order to built a loudspeaker able to reproduce frequencies in the whole audio spectrum, 20 Hz -20kHz, it is necessary to use two or more drivers. Each driver reproduces only a part of the frequency range. To limit the reproduction of each driver to its optimal working range it is necessary to split the signal into different frequency bands. This set of electrical filters are called crossover networks [4, p. 195-217].

An ideal crossover network consists of a set of non-overlapping band limited filters. The total amplitude response of the filters summed up should be one and the phase zero. In real situations brick like filters are not realizable and there exists band interaction of the filters around the crossovers frequencies. Thus, care must be taken in the design of the crossover in order to obtain an overall response as close to the target function as possible. Figure 2.6 depicts a three way crossover network where the signals are split into low, middle and high frequencies.

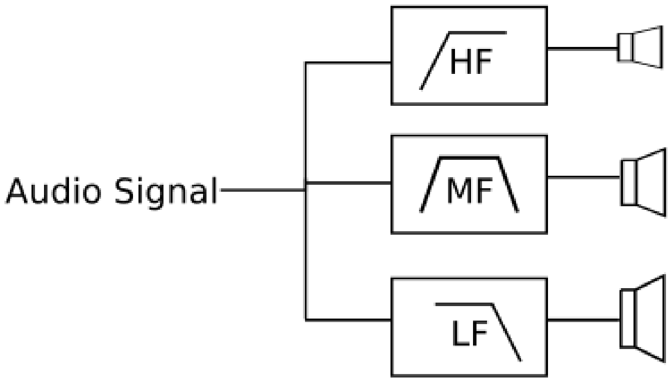

Figure 2.6: *Three ways crossover network.*

In general, the target function is defined such that the summation of HF, MF and LF is equal to one. To simplify the mathematical derivations a two way crossover, where the signal is just split into low and high frequencies, will be considered. In this case the target function can be defined by:

$$H_t(f) = HF(f) + LF(f) = 1 \tag{2.19}$$

There exist some design approaches to achieve this criteria. In the following a rough description of some of the most commonly mentioned designs is presented.

Butterworth or Maximally Flat

The simplest configuration to design a crossover network consists in two first order Butterworth filters defined as follow:

$$LF(s) = \frac{1}{1+s_n} \tag{2.20}$$

$$HF(s) = \frac{s_n}{1+s_n} \tag{2.21}$$

where

$$s_n = \frac{s}{2\pi f_o}$$

is the complex frequency, $s = \sigma + j\omega$, normalized to the nominal crossover frequency f_o. It can be shown that the summation of LF and HF leads to first order all-pass filter, where the magnitude is one and the phase is zero [23]. However, this is not consider to be a practical design due to the slow decay slope of the filters, which is -6dB per octave.

Using a third order Butterworth filters the roll-off is increased up to -18dB/octave. In this case, the equivalent filter for in phase connection is a second order all-pass section whereas for out of phase connection its equivalent is of first order. One of the characteristics of this design is that at the crossover frequency the output is at -3dB making the power output is constant [4]. However, the radiation pattern of the loudspeaker is not symmetrical with respect to the driver axis because the outputs of LF and HF are in phase quadrature [23].

Linkwitz-Riley

The Linkwitz-Riley (LR) approach is often used in crossover design. It consist in two Butterworth sections in cascade. It has the property that the outputs are in phase at all frequencies and the magnitude drops $-6dB$ at the crossover frequency, therefore there is a 3dB dip in the power response at this frequency [4]. The most commonly used configuration is the fourth-order LR described as follows (see [23] for more details):

$$LF(s) = \frac{1}{(1+\sqrt{2}s_n + s_n^2)^2} \tag{2.22}$$

$$HF(s) = \frac{s_n^4}{(1+\sqrt{2}s_n + s_n^2)^2} \tag{2.23}$$

The result of the summation of both filters is equivalent to a second order all-pass filter with $Q = 1/\sqrt{2}$:

$$H_{LR}(s) = \frac{1-\sqrt{2}s_n + s_n^2}{1+\sqrt{2}s_n + s_n^2} \tag{2.24}$$

In general, crossover networks resemble all-pass filters which were described in section 2.3.4 on page 12. The magnitude of this filters is unity but their phase might vary depending on their Q and their center frequency f_o.

2.5 Previous Research

This section describes the research on the audibility of phase changes and phase distortions from early to recent studies. It reviews chronological a selection of the most relevant studies and points out their main findings.

The first who did serious research on the audibility of phase was Ohm in 1843 and 1844 [29], [30]. He stated in his auditory law that the ear detects frequencies and amplitudes of all components of complex sound waves without regard to its phase. In 1877 Helmholtz experimented with phase shifts in music and discovered no audible differences between sound samples with phase differences. Therefore, he confirmed Ohm's auditory law but restricted it to continuous signals as music: *"Differences in musical quality of tone depend solely on the presence and strength of partial tones, and in no respect on the differences in phase under which these partial tones enter into composition* [15, p. 127]." These experiments were confirmed by a few authors but also questioned very soon.

Later experiments, for example from Mathes and Miller in 1947 [26], showed phase shifts are clearly audible under certain conditions, for example for amplitude- and frequency-modulated tones with enough phase difference. Thus, Ohm's and Helmholtz' results are explained with the simple mechanical test setup they used in these early days and with the only subtle difference between sounds with a small phase shift.

Since the quality of hi-fi audio equipment increased and flat magnitude responses were achieved, more detailed work about the influence of phase distortions was done in the 1970s and 1980s. In 1974 Hansen and Madsen showed that a difference between a signal and its time-inversion is audible [13]. Additionally they found in [14] frequency- and sound-pressure dependent thresholds for the audibility of phase differences of tones with three harmonics and showed that phase differences are more audible with headphones than with loudspeakers in semireverberant rooms. They explained this effect with the additional information from the reflections in the room reaching the ear.

Fleischer [12] investigated in 1976 the influence of the duration, the amplitude spectrum and the phase spectrum of test signals with three equidistant tones on the audibility of phase changes. He found that the threshold for audible phase changes is constant for test signal durations from 2 s to 0.25 s with a pause of 0.3 s between the test signals of a comparison pair. For test signals shorter than 0.25 s phase changes were less audible.

The lowest threshold for audible phase changes was observed with test signals where all three components have nearly the same amplitude, when the components had different amplitudes the thresholds increased. When the phase spectrum of the test signals was changed from $0°$ to $90°$ threshold differences were observed mainly in high frequencies, but the influence of the test signal's phase spectrum is much smaller than that of the amplitude spectrum.

Further, Fleischer compared phase change audibility thresholds using headphones and with loudspeakers in an anechoic, a reverberant and a semireverberant room. Phase changes were much more audible with headphones and with a loudspeaker in an anechoic room than in the semireverberant room and the reverberant room, where the thresholds were 2.3 times higher than in the anechoic room. This supports Hansen's and Madsen's theory that room reflections decrease phase sensitivity.

Because phase distortions in the audio chain are mainly caused by the all-pass characteristic of loudspeaker's and headphone's crossover network, Blauert and Laws studied 1978 the influence of all-pass filters on sound signals [3]. They compared with a cascade of four first order and 16 second order all-pass filters distorted signals (speech, music, noise, harmonic series, short impulses) with the undistorted signals. They found the difference between these test pairs most audible for impulsive signals in an anechoic room.

Additionally, they investigated the influence of training on the ability to recognize phase differences. The training period was 15 days long with a 30 min training session per day. The highest improvement was observed in the first sessions, the subject's thresholds for phase distortion detection were at the second day decreased more than one third compared to the first session. After 8 days the thresholds were stabilized. They concluded, phase distortions caused by all-pass filters are audible and need to be corrected under extreme conditions, e.g. trained listeners, critical test signals and headphone presentation.

Oppenheim et al. showed in 1979 phase informations in speech are important for the intelligibility [33]. Thus a phase-only speech sample with a magnitude set to unity was still good to understand, while the same sample with only magnitude information contained no speech intelligibility.

Suzuki et al. found no significant differences in phase shift audibility between anechoic and listening rooms, but again lower audibility thresholds when using headphones [37]. They used all-pass filters to distort their test signals like square, sinusoid and sawtooth waves. They found big differences in the test subject's individual detection rates. None of their subjects could detect phase distortions in music.

On their research Lipshitz et al. [24] focused on the audibility of phase distortion in the midrange frequencies. They used first and second order all-pass filters with different Qs, centered at frequencies within 300 Hz - 3 kHz. Low frequency square waves were used for testing as well as some common sounds such as handclaps and speech. They concluded that the phase distortion can be heard in most common acoustical signals, but they are easier to detect in artificial sound samples. This audibility was found to be more pronounced through headphones, but neither the number of subjects nor their experience was mentioned. In his conclusions is also stated that the phase distortion might have a subtle effect and that their results must not be refer as enough argument to demand linear phase transducers for high quality sound reproduction. More research was also suggested.

Preis and Bloom [35] investigated distortions of minimum-phase anti-alias filters using a $10\mu s$ - wide rectangular pulse and headphones and found the ear more sensitive to group-delay distortions in the middle of the audio band (4 kHz) than at the upper edge (15 kHz) for the same filters but with different cut-off frequencies.

Toole [38], [39] directed his study to sound quality of loudspeakers. He compared different measurement methods and proposed a parallel system of subjective and objective measurements. A set of loudspeakers was evaluated by 42 expert listeners. It was concluded that experienced listeners with normal hearing prefer wide bandwidth, flat and smooth amplitude response as well as uniform dispersion. It is suggested to recreate the high fidelity rating using normal listeners who will be a better representation of the average user. No depth analysis on phase distortion was done in this study, but it is shown that the best rated speakers had smooth phase responses, but without any particular overall shape.

Johansen and Rubak [17] analyzed the excess phase effects in loudspeakers/room deconvolution in 1996. They tested signals convoluted with whole impulse response as well as the excess phase part alone. This signals were assessed by 10 normal listeners with headphones. The results showed that the excess phase in a loudspeaker/room transfer function is audible but it is masked by the minimum phase part of the impulse response. They also concluded that when a room with a longer reverberation time is used for the room impulse response the phase effect is more audible. The same effect occurs for longer impulse responses. The most significant impact was found for male speech, followed by female speech, while guitar music showed less audible differences.

Koya on his study [21] in 2000 focused on the significance of the audibility and compared different synthetic signals (sawtooth shaped, impulsive) convolved with all-pass filters centered at each signals frequency. He also tested music and percussion with experienced subjects. His results were not as significant as Lipshitz [24] and he concluded that the audibility of phase distortion is a subtle effect and that it is highly dependent on inter-individual differences, listening environment and test signal. No analysis on phase characteristics of the transducers used was done.

Banno et al. [2] investigated in 2002 the influence of group delay peaks and group delay bandwidth on signals with a flat amplitude response. They showed that the auditory system is sensitive to the difference between zero phase and nonzero phase signals. Further, their tests indicate that the auditory system is less sensitive to group delays with a narrow bandwidth.

In 2002 and 2007 some studies carried by Møller, Minnaar et al. [27], [25] have been oriented to find the thresholds of audibility of phase distortion. Second order all-pass filters centered at different frequencies within 1kHz and 12kHz were tested with an impulse signal centered at the sampling frequency. The thresholds were measured on normal listeners. It is stated that the audibility of second order all-pass section depends not only on the Q but also on the center frequency, whether is presented in both or just one ear, causality and stimulus signal. Two kind of thresholds were found: ringing and lateralization. The latest occurs when the all-pass section is present in only one ear. Values of the thresholds were found and it is stated that in the case of ringing the threshold can be defined as a single value independent of frequency. The same characteristics were found in low frequency for the lateralization threshold. It was also found that the detection of the distortion is done as a monaural process.

Flanagan et al. [11] analyzed in 2005 thresholds and psychometric functions for the group delay audibility of clicklike sound signals with 5 test subjects. They achieved the group delay by filtering the test signals with second-order all-pass filters at center frequencies of 1, 2 and 4 kHz and a bandwidth of approximately 32% of the center frequency. The thresholds for listening with headphones and with loudspeakers in a low reverberant room were about the same values while they were for listening with loudspeakers in a reverberant room more than double than with headphones. They found no significant differences in the thresholds and psychometric functions of the different center frequencies when listening with heaphones and with loudspeakers in a low reverberant room, but in a reverberant room the thresholds of lower center frequencies were significantly higher than for high center frequencies.

For monaural and binaural listening the difference was only significant for the reverberant room, where group delay was easier to detect for binaural listening. The same was found for the difference between cone loudspeakers and distributed-mode loudspeakers (DML), where cone loudspeakers had a lower threshold in the reverberant room, but both loudspeaker types showed the same behaviour for the most other parameters. The only time the DML showed a significant higher threshold than the cone loudspeaker was for monaural listening in the reverberant room. They investigated additionally the influence of the sensation level on the audibility of group delay using headphones, where they found a minimum sensation level necessary to discriminate a group delay and increasing thresholds for decreasing center frequencies of the group delay.

In summary, the recent studies found all phase distortions or phase differences audible, but the differences between undistorted and phase distorted signals are subtle and only recognized under certain conditions. Some of the studies showed only the general audibility while other tried to find thresholds of audibility by changing the Q-factor of all-pass filters or other parameters. The mentioned authors used different processes and filters to distort the phase of test signals. Often artificial sound samples were used, where impulsive sounds showed to be more sensitive, while for natural signals as music or speech the thresholds were significantly higher. Furthermore, the listening environment seems to have a significant impact on the audibility, in listening test with headphones and with loudspeakers in anechoic rooms the differences were more audible than in normal reverberant listening rooms. Some studies showed additionally large individual differences in subject's perception of phase distortions and that trained listeners detect these distortions easier.

2.6 Problem Statement

Since most of previous studies used mainly artificial sound signals in an artificial environment to investigate phase distortions, this thesis is conducted under as realistic as possible conditions to achieve results which are applicable to real listening situations. Hence, both headphones and loudspeakers in a standard listening room shall be used. Further, phase sensitive but realistic sound samples out of a wide range of music, speech and other common sound samples will be select.

All previous investigations were looking only for thresholds for the discrimination between undistorted and distorted signals under different circumstances. Therefore, shall here be investigated how big the influence of phase distortions on audio quality is. To achieve this, a test will rank undistorted signals and different amounts of audible phase distortions on a quality ratio scale. If phase distortions decrease the audio quality, thresholds for the degeneration might be found.

It is expected from listeners well-known everyday experienced sounds, such as speech, will be ranked best undistorted, because the distorted signal might sound unnatural or unusual. For other signals, for example electronic music or not very common instruments, the distorted sample might be ranked as good as the undistorted or even better. This can be because a slight coloration will not be recognized as “wrong” if the listeners do not know the “correct sound”.

As phase distortions in the audio chain stem mainly from the all-pass characteristic of crossover networks, the introduced phase distortions will be achieved from all-pass filtering the test signals with varying Q parameter. The center frequencies will be defined according to the signals’ spectrum. Additionally, only second order all-pass filters will be used in order to make a fair comparison with the crossover networks described before.

CHAPTER 3

MEASUREMENTS

To ensure that the listening tests performed in this thesis are not influenced by the test setup, it has to be guaranteed all setup components have a satisfying flat magnitude response and linear phase response. Since the only parts of the audio reproduction chain which can not be assumed as ideal and could introduce a reasonable amount of magnitude or phase distortion are loudspeakers and headphones, impulse response measurements on these were performed. The impulse response measurements will show also if additional equalization is needed. Further, the directivity pattern of loudspeakers were measured to ensure the relation of direct sound to reverberant sound in the listening test is not influencing the test in a normal semi-reverberant room in an unusual way. From the results of these measurements the loudspeaker and the headphones with the best performance are chosen for the listening test.

Therefore, the following section 3.1 lines out the measurements of the impulse responses of a few different loudspeakers and discuss their results. Section 3.2 on page 24 is describing that for headphones respectively. Detailed measurement reports can be found in the appendix A.1 on page 67 and A.2 on page 75.

3.1 Loudspeaker Measurements

To find the loudspeaker which fits best the requirements of the listening test, the impulse responses of five high quality loudspeakers, two active and three passive, were measured. These loudspeakers are listed in table 3.1.

Manufacturer	Model	Type
Bang & Olufsen	Beolab 6000	active
B&W	DM60152	passive
DALI	2A	passive
Genelec	1031A	active
KEF	Reference Series 107	passive

Table 3.1: *List of measured loudspeakers.*

Figure 3.1 shows the on axis magnitude response and figure 3.2 on page 23 the phase response.

It can be noted from the graph that the Dali and the KEF have prominent dips above 4kHz. B&W, B&O and Genelec show fairly flat frequency responses. Nevertheless, the B&W shows a dip of about 5dB at 5.5kHz approximately. The B&O loudspeaker presents a roll-off of 10dB between 4kHz and 10kHz. In general, the Genelec's 1031A frequency response is the most flat.

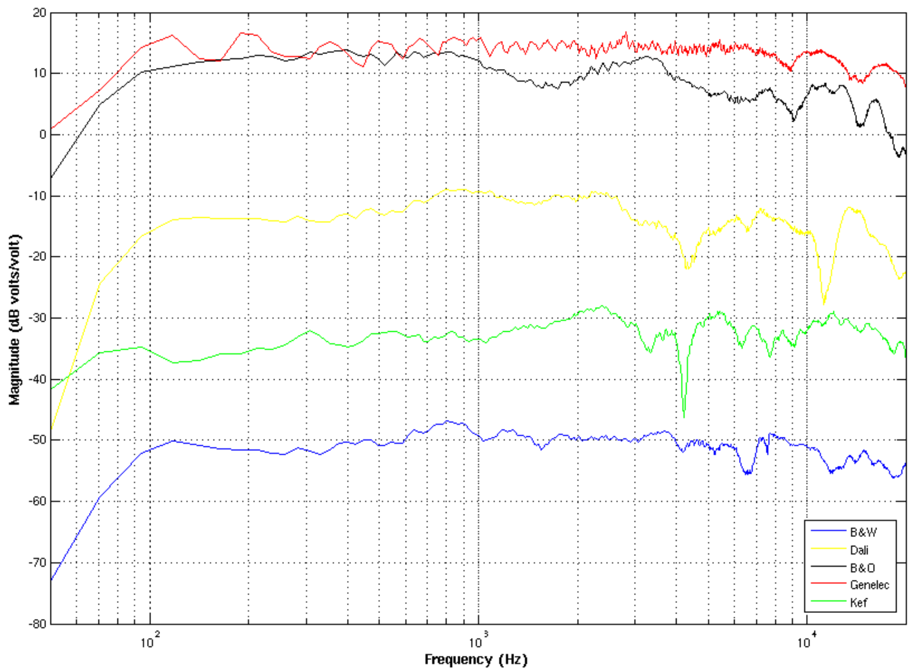

Figure 3.1: *Magnitude responses of measured loudspeakers.*

Figure 3.2 on the facing page depicts the phase response of the loudspeakers. From the graph it can be observed that all the measured loudspeakers have an approximately linear phase. Thus, from the magnitude responses it can be concluded that the Genelec is the best suitable loudspeaker for the listening test.

To ensure this speaker's characteristics in a not anechoic room is not influencing the test, its directivity pattern were measured and are plotted in figure 3.3 on the next page. It can be seen it is distributing the sound equal to the front and to the side, therefore it fits all requirements for the listening test.

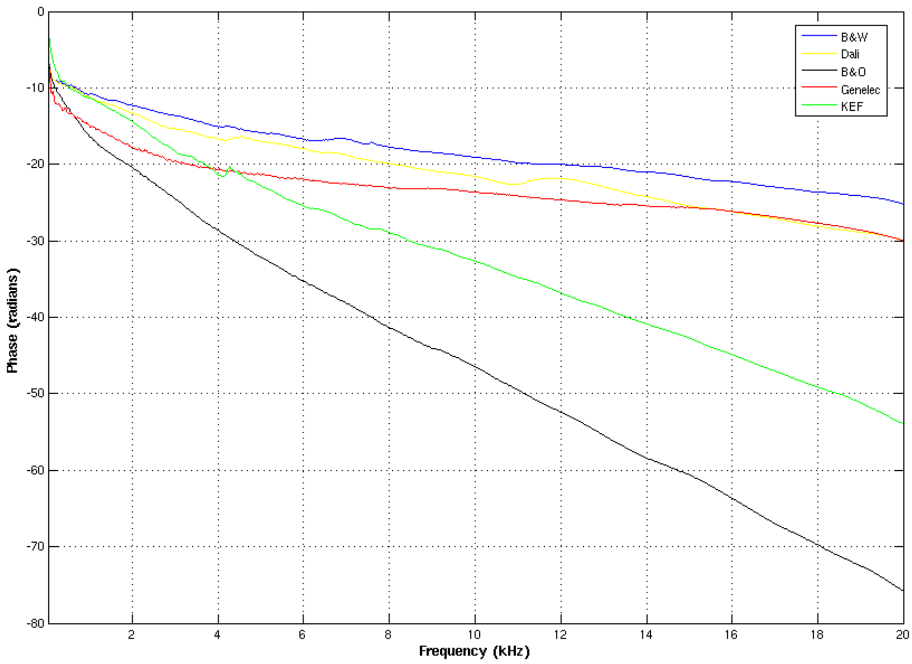

Figure 3.2: *Phase response of measured loudspeakers.*

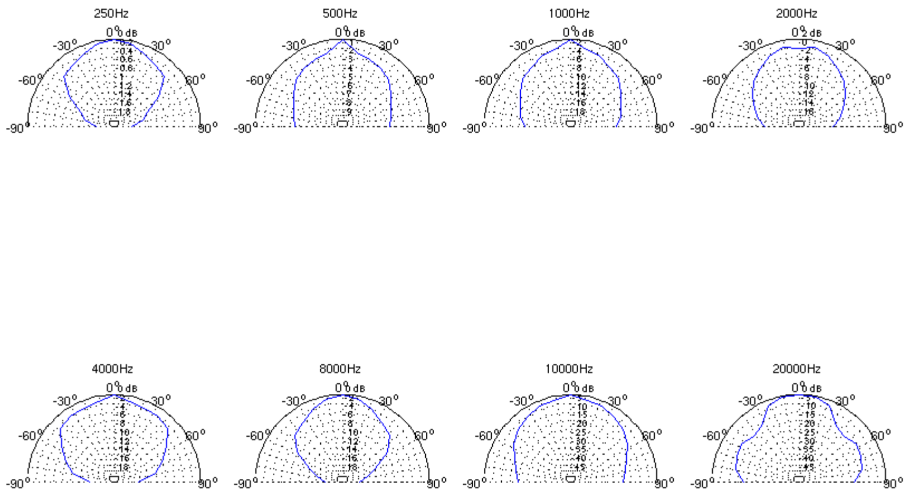

Figure 3.3: *Directivity pattern of the Genelec loudspeaker.*

3.2 Headphone Measurements

For the test parts using headphones three different headphones were measured, two electrostatic and one electrodynamic (see table 3.2). All of these headphones have only one driver in each ear which is common for nearly all headphones. Impulse response measurements were performed using the MLSSA measurement system and both an artificial ear and a VALDEMAR dummy-head.

Manufacturer	Model	Type
Beyerdynamic	DT 990 Pro	electrodynamic
Sennheiser	HE 60	electrostatic
Stax	SR Lambda Pro	electrostatic

Table 3.2: *List of measured headphones.*

Figure 3.4 shows the magnitude responses of the headphones and 3.5 on the facing page the phase delay, the group delay and the differential time delay distortion for all headphones, measured with VALDEMAR. From the figure it can be seeing that the Sennheiser has the steepest roll-off at low frequencies at around 300Hz. The Stax headphones have a dip around 3kHz. The Beyerdynamic DT 990 Professional headphones have the smoothest frequency response.

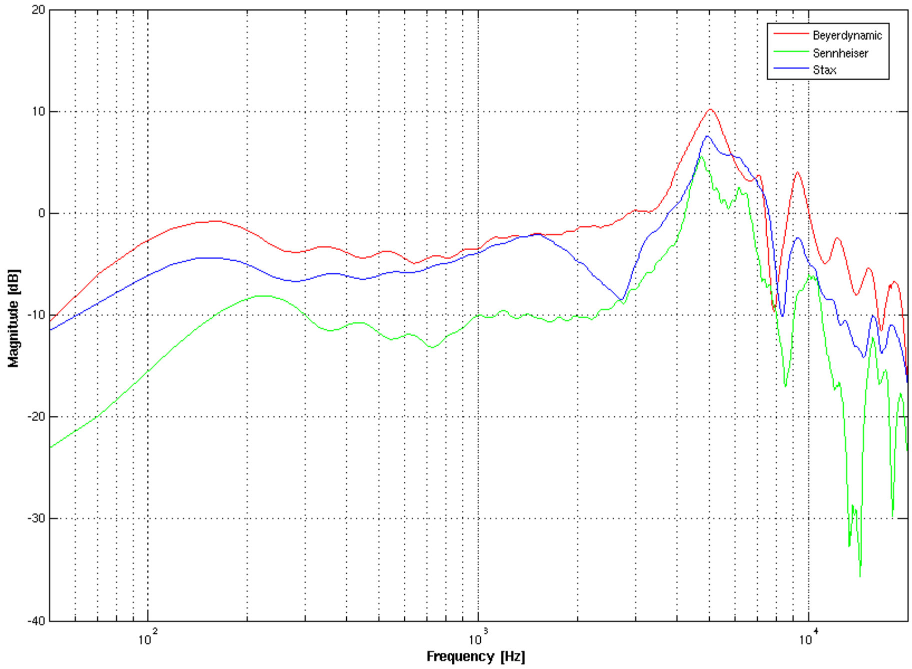

Figure 3.4: *Headphones magnitude response with VALDEMAR.*

From the phase results it can be observed that the Beyerdynamic has a zero phase delay whereas the Sennheiser and the Stax have zero phase delay only above 1kHz. This indicates that the Beyerdynamic headphones are linear phase. Due to this characteristic they were selected for the listening test. Additionally, its smooth magnitude response makes it less difficult to equalize.

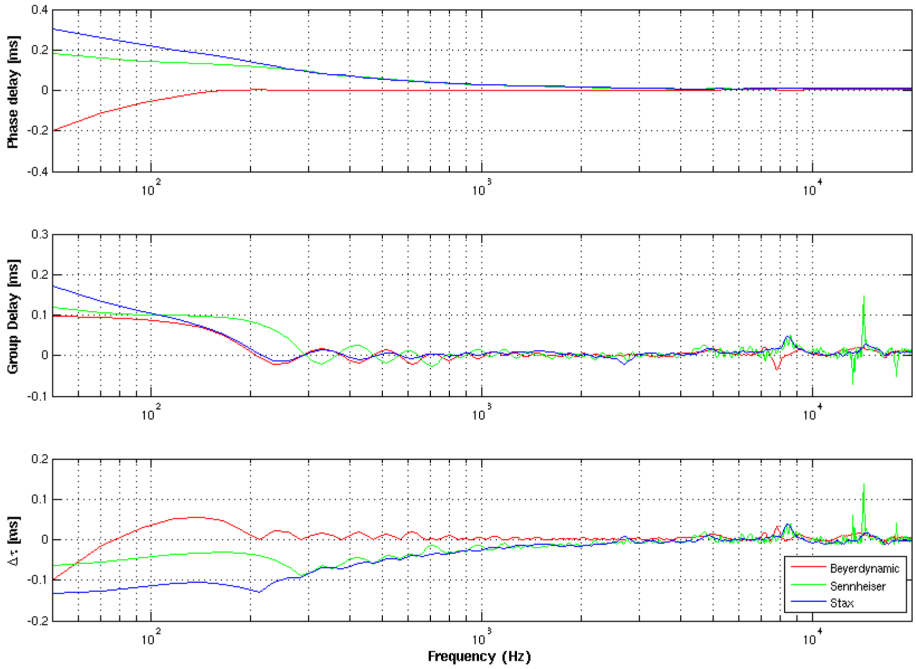

Figure 3.5: *Phase delay, group delay and differential-time delay distortion.*

Additionally, an equalization using the inverse filter described in section A.3 on page 78 was performed for the headphones. This was done to have a nearly ideal transmission system for the measurements of phase distortion audibility thresholds. The results of that equalization is shown in figure 3.6 on the following page. The picture shows that while the equalized magnitude response is approximately flat the phase delay is almost not affected by the equalization. This is, the headphones are still linear phase after equalization.

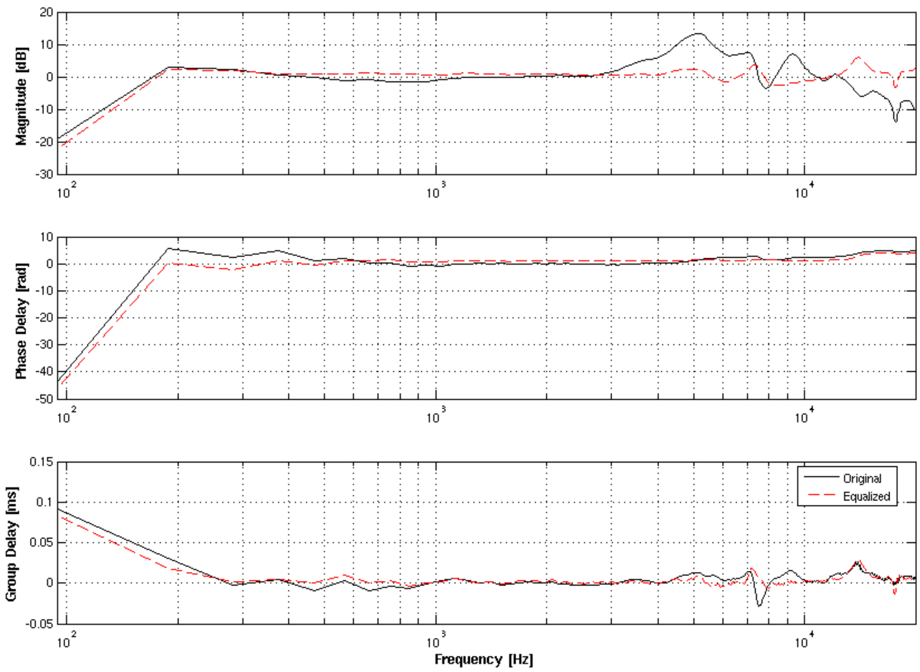

Figure 3.6: *Equalization results for the headphones Beyer Dynamic DT-990 PRO .*

CHAPTER 4

LISTENING TEST DESIGN

This chapter contains the steps carried out on the listening tests design. The first section describes the threshold listening test and the design criterion used. Some threshold measurement methods are mentioned and the final design is presented. Results from this test are shown and some conclusions are drawn. The second section describes the design and methods used for the main test. This test was based on the outcomes of the threshold test.

4.1 Thresholds Listening Test and Results

The purpose of this test was to select a set of signals in which phase distortion is more audible. This was done by measuring the thresholds of different sounds taken from the CDs *Sound Quality Assessment Material (SQAM)* from the EBU [40] and *Music for Archimedes* from B&O [31]. Some signals taken directly from commercial recordings were also used.

A threshold is defined in [16] as "*the limit below which a given stimulus or the difference between two stimuli cease to be perceptible*". In practice, these subjective thresholds are not absolute and subjects tend to show some variability. This is, there will be values where the differences are always or never perceptible. Whereas there are some values where the differences are sometimes detected. Due to this variability the thresholds are usually defined as a statistical measure [16]. One can define the threshold as a specific point in the psychometric function e.g. 50% or 75%. This function plots a percentage of correct answers against stimulus values. Figure 4.1 on the next page shows an hypothetical psychometric function and thresholds defined at 50% and 75%.

There exist a number of methods to measure and draw the psychometric function. These methods can be split into two groups: detection and discrimination methods. The former measures the minimal level required for a stimulus to be detected whereas the later measures the minimal difference required between two stimuli to be discriminated [16]. In the detection experiments the absolute threshold is measured and methods such as the constant stimuli and method of limits are used. In the discrimination experiments the just noticeable difference is measured instead and force choice methods are commonly used.

The thresholds are usually defined as the intensity where the stimulus is detected or the stimuli are discriminated 50% or 75% of the times. Given the scope of this thesis the discrimination

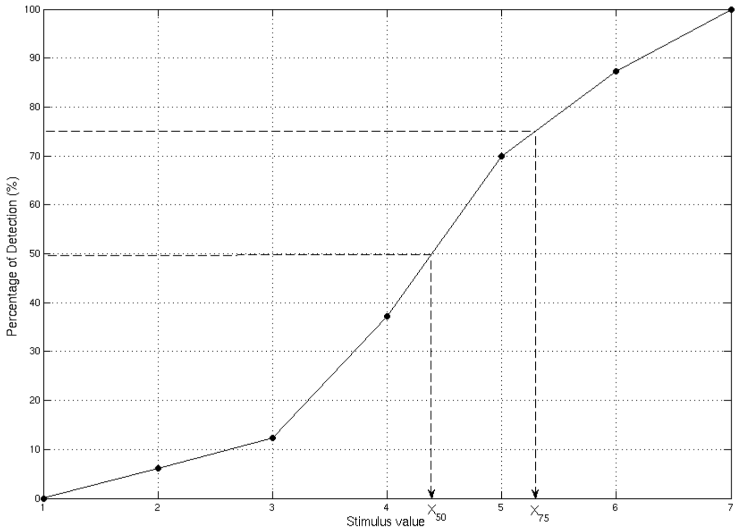

Figure 4.1: *Hypothetical psychometric function. X_{50} and X_{75} denotes the thresholds at the 50% and 75% respectively.*

experiments are consider to be more suitable. In the following, a rough description of this methods is presented and the design criteria used is described.

4.1.1 Choice Of Method

One of the methods commonly used for discrimination experiments is the two category forced choice (2AFC). In this method two stimuli are presented to the subject: a standard and a comparison. The task of the subject is to response whether the two stimuli are equal or not or which of the stimuli is larger, greater, smaller than, etc. In this method there is a 50% probability of guessing the right answer.

Another method is the three category forced choice (3AFC). This method consists in presenting to the subject a sequence of three stimuli where just one differs from the others. This is, the sequence can be two times the standard and one time the comparison and vice versa. The order of the stimuli is presented in a random fashion and the task of the subject is to select which of the stimuli is different. The probability of guessing the right answer with this method is $^1/_3$.

For the nAFC methods the threshold is often defined as the signal level at which the probability of correct responses is half way between perfect performance (100%) and probability of guessing [20]. In the case of the 2AFC this will be 75% whereas for the 3AFC it will be 66.6%.

Even though the 2AFC is commonly used for threshold tests the 3AFC was chosen instead due to two main drawbacks of the former. The foremost is the high probability of chance of the

2AFC which is 50%. The second reason is that if a 2AFC test with the question "Are the sounds different?" requires double amount of sequences (sequences with and without difference) which leads to an increase of time and reduction in the test efficiency. If a 2AFC test with the question "Which sound is more distorted?" or any other question of the kind was designed, the subject's attention would have been directed toward specific characteristics instead of overall differences. It was found in informal preliminary tests that the phase distortion can be perceived as different effects depending on the stimulus, the frequency and the subject. Therefore, in order to take into account any perceived difference, a 3AFC was designed and the simple question "Which one sounds different?" was asked.

4.1.2 Design Criteria

A simple adaptive test with the weighted up-down method was implemented. This was done in order to reduce the test time and increase efficiency. The adaptive procedure consist in a test where the stimulus level on one trial is determined by the preceding response [22]. In general this methods have a fixed step size and converges to a prescribed point in the psychometric function.

The simple up-down method applies the rule of one step down for every correct response and one step up for every incorrect response. This method has a constant step size and it converges to the 50% of the psychometric function. Variations on the level depending on the outcome of two or more preceding trials can be also applied [22]. These kind of methods are called the transformed up-down methods and converge to different fixed points in the psychometric function. Kaernbach [19] suggested a method which converges to any desired point in the psychometric function. This method consist on using the simple up-down rule but giving different step sizes to the up and down groups. The step sizes depend on the desired convergence point X_p and are defined as follows:

$$S_{down}p = S_{up}(1-p) \tag{4.1}$$

where S_{down} and S_{up} are the step size of the down and up groups respectively. As mentioned before for the 3AFC test the threshold can be defined at the 66.6% of the psychometric function. Then, applying the method suggested the adaptive procedure must comply with the rule $S_{down}/S_{up} = 1/2$. This means that the step size of the up group is twice the step size of the down group. In other words, the algorithm must go one step down for each correct response and two steps up for each incorrect response.

Figure 4.2 on the following page shows an example of how this method was implemented. A reversal is defined as the point where the test changes from down rule to the up rule. In this example a set of 12 hypothetic Qs are shown. At a Q of Q12 the sound is assumed to be highly distorted and thus clearly audible. At a Q of Q1 the distortion is assumed to be not perceivable. In the first run the step size is set to two in order to get a faster approach to region where the target value lies. After the first reversal the applied rule is one-down-two-up. A way of estimating the threshold is by calculating the median of the reversals. Despite of being a rather optimistic result, estimating the threshold in this way is more resilient to possible loose of attention from the subject. The median was chosen instead of the mean because of giving a measure that is more robust to lucky or unlucky guessing.

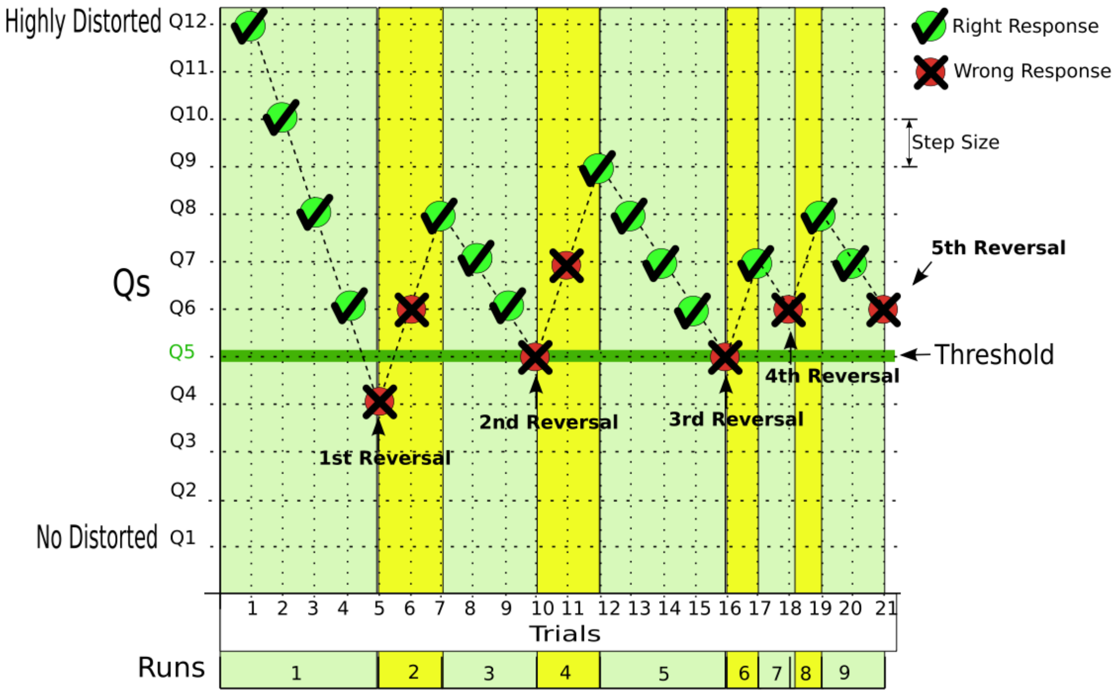

Figure 4.2: Data example of the simple up-down weighted method (rule: 1 down, 2 up).

For the simple up-down method a recommended procedure is to continue the test until at least six or eight reversals are obtained [22]. However, for the case of transformed up-down methods it is desirable to stop the test when clear convergence is reached. This might require a considerable amount of time and in most of the cases a compromise should be made.

4.1.3 Stimuli And Setup

Different music and speech signals were used as stimuli for this test. Each stimulus was convoluted with a second order all-pass filter centered at different frequencies. The raw signals were used as the standard and the signals convoluted with the all-pass filter as the comparison. The main criteria used to select the stimuli was sounds which are normally heard through loudspeakers. Previous studies had suggested that the phase distortion is more audible with impulsive sounds [24], [25], [35], [15], [3]. Therefore, some percussive sounds were included in the set of stimuli.

A preselection of signals was carried out in order to reduce the amount of stimuli presented. These signals were analyzed through short time Fourier analysis and were selected according to their spectral characteristics. In general, band limited and narrow band signals were selected. Additionally, an informal listening test was done in order to select those signals in which, at a considerable high Q, the distortion was clearly perceptible.

In broadband signals such as orchestral music not even high phase distortion was found to be perceptible, therefore they were discharged. However, in order to have a broader conclusion about the thresholds some sounds in which the distortion was found to be difficult to perceive were include. Table 4.1 shows the signals used on the test. The selection of the frequencies was done by means of the spectrogram. Only some frequencies where most of the energy was concentrated were tested.

Signal	Type & Duration	Track/Sentence	Filter Center Frequencies [Hz]
Capriccio	Anechoic guitar solo [4s]	–	261, 311, 440
Castanets	Anechoic castanets [3s]	–	880
Eddie Rabbit	Pop music [5s]	70 - SQAM	750, 1000
Electronic gong	Synthetic signal [3s]	06 - SQAM	500
Norah Jones	Vocal music [5s]	The nearness of you	200, 311, 580
Radiohead	Percussive electronic music [5s]	Idioteque	100, 500, 1000
Speech	Anechoic female speech [4s]	*"The numerals take different place values depending on position"*	150, 300, 500

Table 4.1: *List of signals and frequencies used on the test.*

Filter selection

The type of filter chosen is a second order all-pass filter (see equation 4.2). The criteria used to choose this filter is that in some cases resembles the total magnitude and phase response of a typical crossover network, e.g. with a $Q = 1/\sqrt{2}$ it is equivalent to the fourth-order Linkwitz-Riley filter (see section 2.4 on page 13).

$$H_{AP2}(s) = \frac{s^2 - (\omega_o/Q)s + \omega_o{}^2}{s^2 + (\omega_o/Q)s + \omega_o{}^2} \tag{4.2}$$

where $s = \sigma + j\omega$ and $\omega_o = 2\pi f_o$. The filters were centered at each of the frequencies described in table 4.1. The filters were obtained throughout an analog-to-digital conversion of equation 4.2 using the bilinear transformation method in Matlab.

The Q was varied from $1/2$ to 64. As explained in section 2.3.4 on page 12 for $Q \leq 1/2$ the filter turns into a first order all-pass section. The upper limit was set to Q=64 for being consider high enough to make the phase distortion clearly audible in most of the signals. Since the aim of this test is to get a rough idea of the phase distortion thresholds for the stimuli mentioned before and not to obtain absolute values, thresholds lying above Q=64 are consider to be out of the scope of this project.

In previous investigations regarding phase distortion audibility, all-pass filters were cascaded in order to increase the peak group delay. Some had argued that the audibility is determined mainly by the peak group delay [34], [11], whereas some suggested that the group-delay bandwidth might have some significant influence in the results as well [25],[2]. Therefore, in order to have complete control over the variables and avoid ambiguities in the results, no cascaded filters were used.

Since most of the practical applications which resembles all-pass sections have Qs multiples of $\sqrt{2}$, it was considered logical to set the Q scale as an exponential scale multiple of $\sqrt{2}$. Then, the Q values used were:

$$
\begin{aligned}
Q &= (\sqrt{2})^n \\
n &= -2, -1, \ldots, 11, 12
\end{aligned}
$$

which gives a total of 15 different Qs.

Setup

The phase distortion thresholds were measured for each signal and each frequency showed in table 4.1 on the previous page. This leads to a total of 16 signals. In order to reduce the time of the test, it was set to stop after 8 reversals or in the case the subject reached the ceiling of the staircase 8 consecutive times. Using this procedure the test takes around 10 minutes per sound which gives a total of 160 minutes for the whole test, without including breaks. Then, it was decided to split the signals among subjects, in such a way that one subject evaluated just 4 different signals.

The initial value of the adaptive test was set to Q=64. This was done in order to give the subjects cues of what to listen to. The step size of the first run was set to 2 in order to reduce the convergence time and after the first reversal the rule one down two up was applied (see figure 4.2 on page 30). The threshold was calculated as the median value of the reversals.

The test was ran through headphones. The chosen headphones were the Beyerdynamic DT-990 PRO due to their approximately constant phase and smooth magnitude response which make it easier to equalize. The minimum phase response of the headphones was equalized by means of the Yule-Walker approximation built in Matlab (see Appendix A.3 on page 78 for more details).

It was verified that there was no significant variation of the phase of the headphones after equalization.

A total of ten listeners -three female and seven male, including the authors- with normal hearing participated in the experiment. The test was carried out in an acoustic isolated cabin. Each subject was given an instruction sheet describing the procedure (see Appendix B.1 on page 81). A short training session of three trials was presented before the test for the subjects to get familiarized with the software, but no training regarding the signals characteristics was performed. The test was divided in four sessions and each session corresponded to a different signal. This is, no interleaving was applied. On each session the subject listened to a single signal and only the Q was varied. Table 4.2 shows a summary of the setup of the thresholds listening test.

Method	Adaptive weighted up-down with 3AFC
Filter	All-pass described in equation 4.2 on page 31
Center frequencies	Frequencies described in table 4.1 on page 31
Qs	$(\sqrt{2})^n$, n=-2,-1, ..., 11, 12
Stop criteria	After 8 reversals or reaching 8 times the upper limit
Threshold estimation	Median of reversals

Table 4.2: *Thresholds listening test general setup.*

The adaptive algorithm was designed in Matlab. All the filtered signals were generated offline and saved in a .mat file. The headphones' equalization was also done offline in order to save processing time. The software selected the signals randomly for each subject, thus there were subjects who listened to the same kind of stimuli distorted at different frequencies in some of the sessions. Picture 4.3 shows the graphical user interface designed.

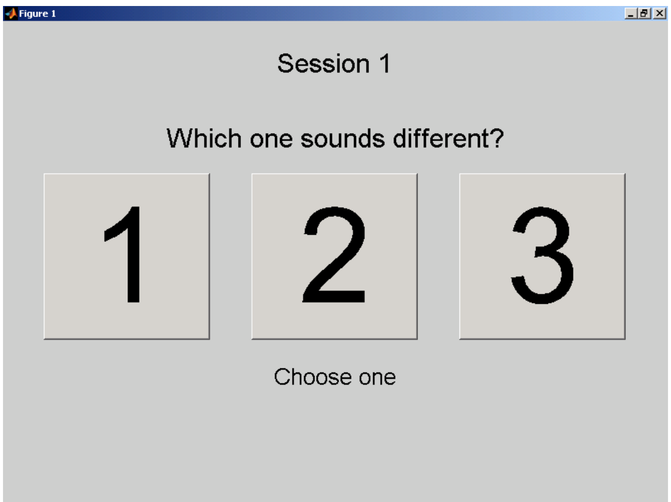

Figure 4.3: *Threshold listening test graphical user interface.*

4.1.4 Results

Each subject evaluated only four of the stimuli listed in table 4.1 on page 31. The selection of the stimuli that each subject assessed was done randomly. Due to the uneven amount of subjects some stimuli were evaluated three times whereas others were evaluated only two times. Thus, the stimuli were not necessarily balanced. This might lead to better mean values when the stimuli were evaluated three times than when they were evaluated twice. However, since the aim of this test was to get a rough idea of the distortion thresholds for certain stimuli, the results presented here are considered to be meaningful enough.

Thresholds were found for every signal and each subject was analyzed individually in order to make fair comparisons. The thresholds were calculated as the mean or average of the individual thresholds. Figure 4.4 shows the average thresholds found for each signal.

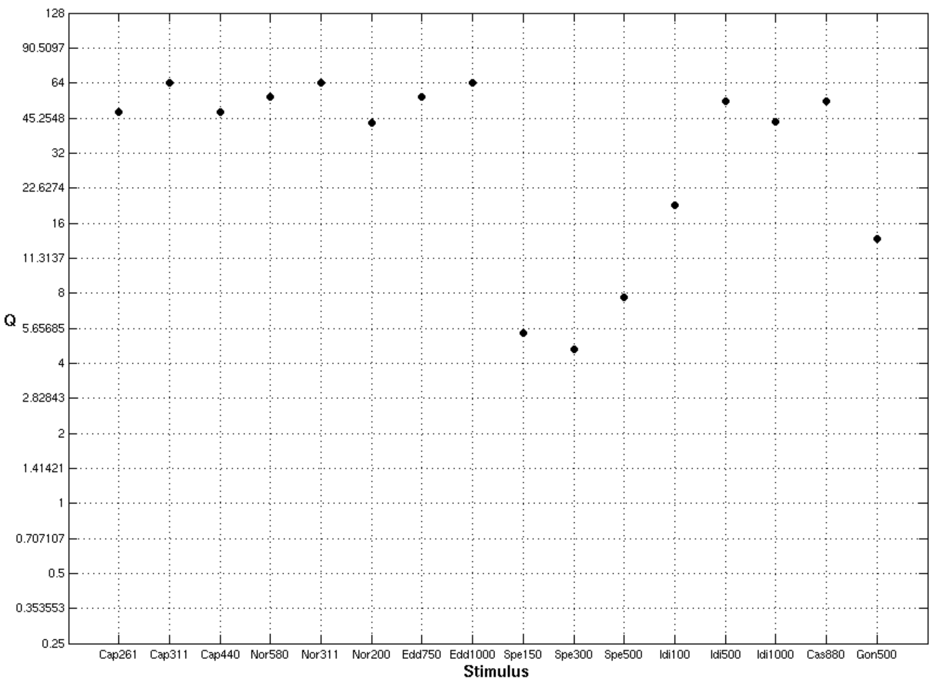

Figure 4.4: *Threshold results.*

The x-axis corresponds to the stimuli on table 4.1 on page 31. For practical reasons the names were abbreviated as the first three letters of the stimulus name and the frequency, e.g. Spe150 = Speech at 150Hz.

It can be concluded from the graph that phase distortion is most audible for speech. This was expected since it has been shown that speech carries most of its important features in the phase

[33]. Signals as the artificial gong and the Radiohead's sample show lower thresholds in comparison with the rest of the signals as well. Thresholds of 64 and 45.2 suggests that the distortion is not audible at all for the signals with these values. This is, such an outcome it is possible in the case where the subject can not hear difference and remains on the test ceiling of the staircase most of the time. Figure 4.5 shows the individual performance of the guitar solo at 311Hz which is an example of this case.

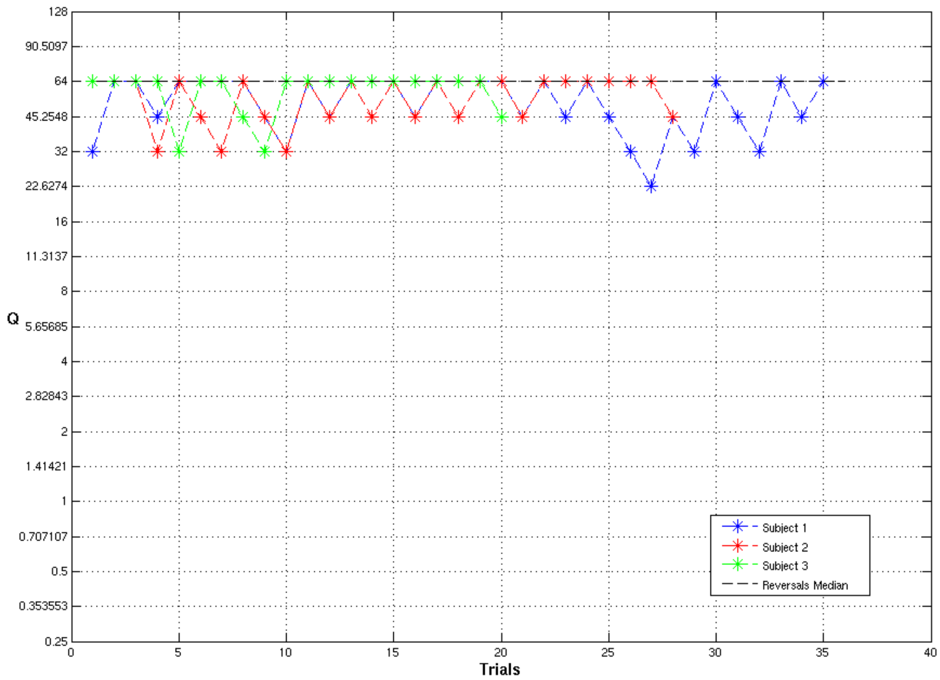

Figure 4.5: *Individual performance with the guitar solo at 311Hz.*

Note that in most of the trials the outcome is 64. In this case, the estimation of the threshold did not only take into account the reversals but also the amount of consecutive trials where the outcome was 64. From this it can be concluded that phase distortion is not audible for Qs lower than 64 for most of the signals evaluated.

Figure 4.6 on the next page depicts the individual performance with the speech signal at 300Hz. The graph shows a significant variability among subjects.

Figure 4.7 on page 37 shows the individual performances for one of the signals which overall threshold suggests that the phase distortion is not audible. It is evident from the graph that whereas for two of the subjects the distortion was not audible, one of the subjects performed with a comparable lower threshold. The main difference between the third subject (one of the authors) and the rest is training. Thus, it can be concluded that training might be a key factor to

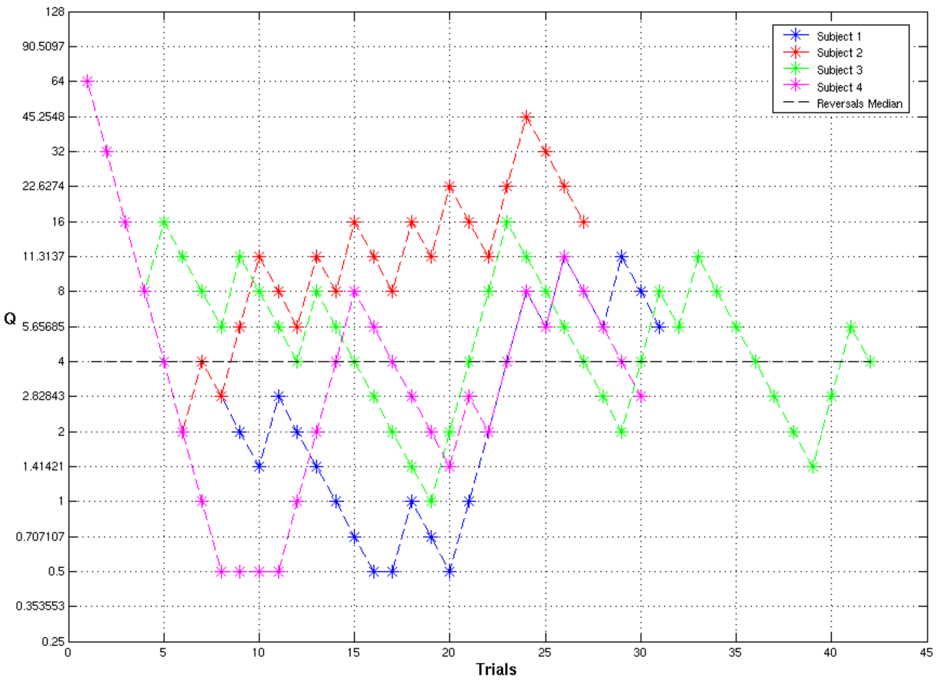

Figure 4.6: *Individual performance with the speech signal at 300Hz.*

measure phase distortion thresholds.

As mentioned before, most of studies suggests that phase distortion is more audible for impulsive sounds. It was expected that in signals such as the castanets the distortion would be clearly audible and its threshold would be one of the lowest of the set of stimuli. Surprisingly, it was found that for Qs lower than 64 the distortion was not audible.

One possible hypothesis to these results is that sounds such as the castanets presents a very narrow bandwidth and only certain frequencies are excited. In all-pass filter with high Q, the group delay bandwidth becomes narrower and more selective. Thus, if the filter is not centered at the exact frequencies presented in the sound they might not be affected by the distortion and therefore make it inaudible [2]. In informal test different frequencies were assessed without a perceivable effect. However, in order to support this statement a more rigorous analysis should be carry out.

How Subjects Perceived the Distortion

After the test each subject filled out an informal questionnaire in which they were asked to describe what did they hear when they heard a difference. It was found to be rather difficult to

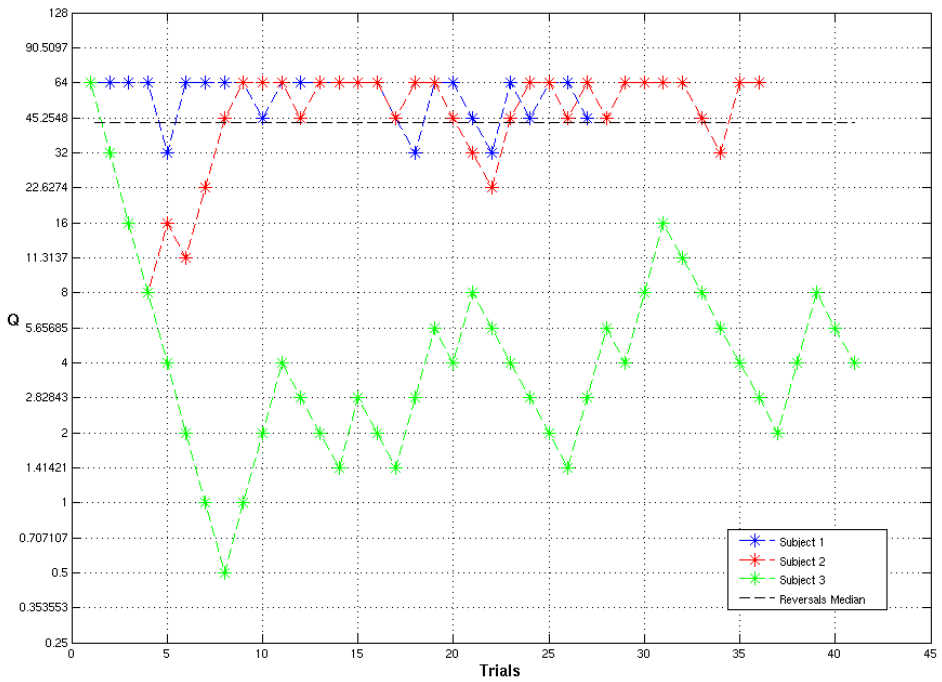

Figure 4.7: *Individual performance with the Norah Jones sample at 200Hz.*

describe the differences and most of the subjects answers varies between coloration, a change of pitch, ringing, loudness variation, bass enhancing and reverberation. Even though most of the subjects were acoustic students and were familiar with the acoustics terminology, no uniform description of the artifacts was found for most of the signal. Only the speech signals and the electronic gong showed agreement between subjects. The distorted speech signals were generally described as a “synthetic” sound and the differences in the electronic gong signal were described as a pitch change.

Summary of Results

In general, broadband sounds such as the Eddie Rabbit, the guitar and Norah Jones samples were found to be resilient to phase distortion. As expected, uncorrelated signals such as the speech were found to be rather sensitive to the distortion.

The main purpose of the test was to select those signals in which the distortion is clearly audible. It was not the intention to carry an exhaustive study about the absolute threshold for each signal. In order to improve the significance of the results some enhancements to the test should be done. For instance, a larger number of subjects and a balanced set of stimuli can be used. Other suggestion is to carry a more rigorous spectral analysis of the signals in order to find the exact frequency at which the distortion is more audible, e.g. varying the frequency and keeping the Q constant.

Even though some drawbacks of the method were found and possible improvements are suggested, the obtained results are sufficient to select the stimuli for the main test. The selected stimuli are thus the speech at 300Hz, the artificial gong and the Radiohead's sample at 100Hz.

4.2 Main Test Design

In the project goal it was queried whether the phase distortion degenerates the reproduction quality or not. The main purpose of this test is to assess the acceptability of different amounts of distortion in different acoustical signals.

In the previous section it was shown that for signals of certain characteristics the phase distortion is audible. It was found that specially for speech and synthetic signals the distortion was more perceptible. It was also found that subjects required a considerable amount of training in order to discriminate the subtle differences introduced by the distortion. The following describes the design of the main test on the basis of the results of the threshold test.

4.2.1 Design Criteria

The most straightforward method to rate sound quality with naive subjects is the pair comparison procedure or 2AFC. This kind of method makes the task easier for the subject and still makes it possible to obtain enough data for a complete statistical analysis. In this case the subject is presented with two sounds and the subject's task is to choose which one of the stimulus is preferred over the other. This kind of method is commonly used in preference studies with untrained subjects and where the rating sensory perception is rather difficult [9].

During the threshold test all the subjects were informally asked what did they heard when they perceived a difference. The answers varied among the subjects and most of the words used were very particular of each subject. This suggest that for non-expert subjects is rather difficult to rate how much of certain characteristic is present in each sound. Therefore, a simple pair comparison test where the subject's task is choosing preference was designed.

Other of the outcomes of the threshold test was the significant effect that training has on the subject's performance. Thus, it was decided that a short training session should be carried out before the preference test. This training ensures that the subject is aware of differences between distorted and undistorted signals. Additionally, this was used as a selection criteria for the subjects and for the signals that the subject would assess.

Since this project is focus specifically on the influence of phase distortion in normal listening situations, it was decided to run the pair comparison test through a loudspeaker and through headphones. The assessment with headphones serves as a reference measure, where no interaction of the room is affecting the perception. Furthermore, the transducer interaction with the stimuli was also diminished by means of equalization. Thus, the test was divided into three parts: training session, pair comparisons with headphones and pair comparisons with loudspeakers.

Training Session

For the training session a modified version of the threshold listening experiment was implemented. The simple adaptive test with weighted up-down method with the rule one-down-two-up was used. The purpose was not to converge to a point in the psychometric function but rather to teach the subject where the difference lies. Therefore a fixed amount of trials was set and the same step size was maintained in every run. This is, the test was setup to stop after ten trials disregarding the amounts of reversals and the rule one step down for correct responses was applied from the first run. The subject evaluated one signal at a time, which means that no interleaving was implemented.

The step size was doubled, decreasing in that way the amount of Qs. Additionally, just Qs from 2 to 64 were evaluated leading to just six possible values:

$$\begin{aligned} Q &= 2^n \\ n &= 1,2,\ldots,6 \end{aligned}$$

In order increase the efficiency of the training, subjects received instant feedback from the test. After selecting which of the signals sounds different, the subjects where able to know whether their choice was correct or not and which one was the correct one.

The outcome of the test was also used to select the signals and the Qs each subject assessed. If the subject was able to perceive the differences with the speech signal but not with the electronic gong or Radiohead, then he or she will only assess the speech. Moreover, if the subject's answers lie above a certain Q e.g. 8, then that subject will just assess the speech signal for Qs larger than 8. The idea behind this selection is to reduce the inter-individual inconsistencies. Given the small differences produced by the phase distortion, the subjects are very likely to chose a preference randomly when no difference is perceived. Therefore, by training the subjects and constraining the stimuli to be compared we decrease the probability that the subject's choice is a random answer.

Paired Comparison

For this part of the test the minimum possible Q was set to 4. This criteria was used due to the results obtained from the threshold test and in that manner to reduce the amount of stimuli to be assessed. The simplest approach to a pair comparison experiment is to present the subject all possible pairs [6]. Thus, if there are n stimuli to be compared the total amount of pairs presented to each subject is $\frac{n}{2}(n-1)$. Taking the set of stimuli the maximum number of pairs per signal a subject assessed was 15.

One important factor to take into account when designing a pair comparison test is the within order effect. When a sequence of sounds is presented to a subject the choice might be influenced by the order of the stimuli [42]. Thence, in order to account for this effect, each pair was presented twice. Once in a AB configuration and second in a BA configuration. This gives a total of $n(n-1)$ comparisons. All the pairs, including reversed repetitions, were presented in a random order to avoid carry over effects. Additionally, the subjects were able to play the sequences again but were instructed to do it just in case of not being 100% assure of the answer and to trust their first impression. This was done in order to give the subject opportunity to repeat the sequence in the case of a lost of attention or when the differences were very subtle.

4.2.2 Test Setup

The main test was performed in the standard listening room of Aalborg University, which conforms to the IEC 268-13 standard for an average listening room. It was verified that the room complied with the dimensions and reverberation time suggested in the ITU recommendation BS.1116. Details of the measurements carried out and the obtained results can be found in appendix C on page 85. This room has a very low background noise level. In figure 4.8 on the facing page the listening room and the test setup are diagramed. Both the loudspeaker and the listening position were on the symmetry axis of the room and the loudspeaker was placed in a distance of 2 m from the listening position.

The hight of the loudspeaker was adjusted in such a way that the tweeter was 1.3 m above the ground. At the listening position the subjects were seated on a chair with a headrest they were instructed to use. The hight of the chair was adjusted for every subject such that the subjects ears were at the same hight as the tweeter of the loudspeaker. That was done to ensure all subjects will receive the same direct sound and the sound which is reflected from the walls.

In front of the seat a touchscreen connected to a computer outside the room was placed to perform the test using an in Matlab programmed graphic user interface (GUI). The Genelec 1031A active loudspeaker was connected to an Pioneer A-616 stereo amplifier, while the Beyerdynamic DT 990 Professional headphone was connected to an Fostex PH-5 headphone amplifier. Both amplifiers were placed outside the room and also connected to the computer.

The playback levels were adjusted with the amplifiers such a way, that the level was comfortable and the perceived level was the same for listening with loudspeaker and with headphones.

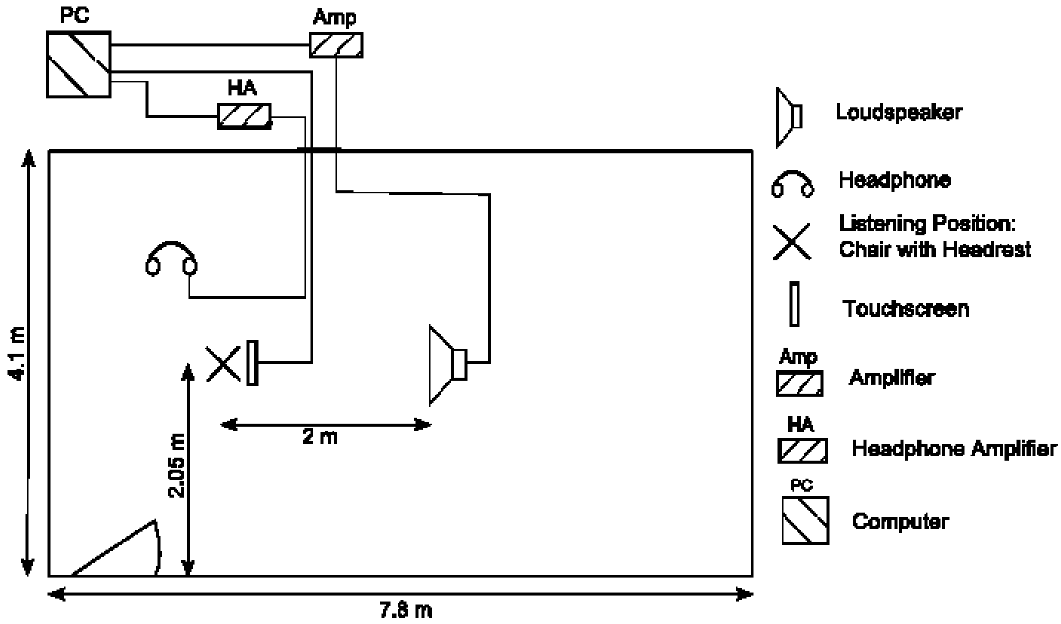

Figure 4.8: *Listening room and test setup.*

4.2.3 Test Subjects

The main test was performed by nineteen subjects, eight female and eleven male. Eight of these were students or staff of the acoustics department of Aalborg University and can be considered as more experienced in formal listening tests. The subjects were between 21 and 38 years old.

Before the test was started an audiometry was done with each subject to ensure no subject is influenced by a hearing damage. The results can be found in the appendix D on page 89. An hearing loss of 20 dB was chosen as the maximal acceptable hearing loss. Therefore subject YMI (see audiometry in figure D.19 on page 98) had to be excluded from the analysis due to her hearing loss up to 40 dB in the right ear.

Additionally all subjects were asked to fill out a short questionnaire (see E on page 99) to find out if they have any known hearing disorders which can not be detected with an audiometry. Thereby was found out three subjects suffering from a mild form of tinnitus, which was assumed not to influence the test.

4.2.4 Test Procedure

Upon arriving to the test a brief instruction sheet with instructions for the first test part was handed out to the subject (see B.2 on page 82). After the subject finished reading, the questionnaire was filled out, the audiometry was performed and the first of three test parts was started.

Between the three test parts the subjects had the chance to rest approx. 5 min to ensure they are concentrated during the whole test. After finishing the first part another instruction sheet with instructions for the second and third test part explained the subject the following tests.

Each test part started with three training comparisons which were ignored in the analysis to introduce the test GUI to the subject. The following describes these three test parts.

Part 1: Training Session

As explained before, this session provided feedback to the subject. A GUI where the subject had to push one of three buttons to indicate which of the three presented sounds was different. The feedback consisted of a smiley depending on the correct or wrong answer and the correct answer was highlighted immediately after the subject's choice. A screenshot of this GUI is shown in figure 4.9. The whole threshold test was done using headphones. In total 30 of such comparisons were performed and this test part took approximately 8 min.

Figure 4.9: *Training session GUI.*

While the subject had a pause between the first and the second test part, the subject's answers from the threshold test were analyzed manually. The thresholds for every sound sample were used to set the lower limits of Q for the preference comparison in the following parts. Therefore, only samples which could be discriminated from the undistorted signal by the subject were used.

Part 2 & 3: Paired Comparison

The second and the third sessions consisted of the paired comparison test as described before. The only difference between test part two and test part three is that former used again the headphones and later the loudspeaker. The subject was asked to chose the sample he preferred out of the test pair by clicking on the according button. Additionally a third button to repeat the whole comparison pair was enabled. Figure 4.10 shows the GUI for this tests.

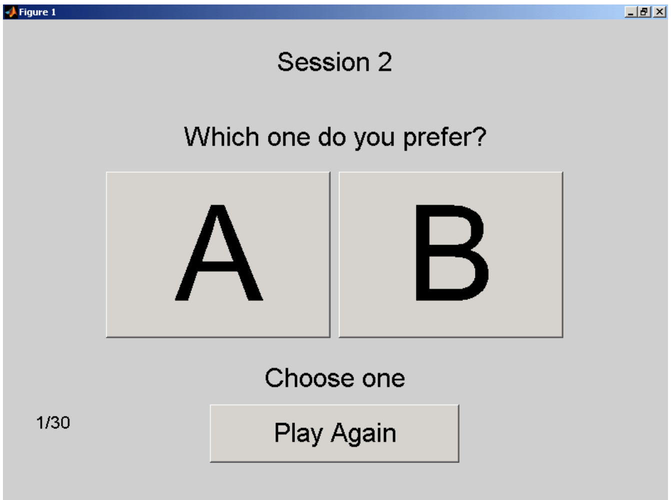

Figure 4.10: *Preference listening test GUI.*

The duration of the test was about 15 minutes per session. This gives a total of 50 minutes for the whole test including breaks.

CHAPTER 5

RESULTS AND ANALYSIS

This chapter contains the results of the main listening test. The chosen probabilistic choice model is described and the tools used to analyze the data are presented. Results are shown for each stimuli. The complete tables and figures of the results can be found in appendix F on page 101.

5.1 Probabilistic Choice Model

There exists a number of models to analyze the data from a paired comparison test. These methods aim to obtain statistical measurements significant enough to assess the differences between the compared elements in the test. One of the most commonly used is the Bradley-Terry-Luce (BTL) model. This model, which is based on some structural assumptions, makes it possible to extract a ratio scale from the data. If there are n compared stimuli, BTL states that for each one there is a true rating $\pi_1, \ldots, \pi_n$. The model assumes that when the stimulus i appears with the stimulus j in a block, the probability π_{ij} that stimulus i is preferred over stimulus j is defined as follows [5]:

$$\pi_{ij} = \frac{\pi_i}{\pi_i + \pi_j} \tag{5.1}$$

with the requirements that $\pi_i \geq 0$ and $\sum_{i=1}^{n} \pi_i = 1$.

With the pair comparison data one can define a frequency matrix of size $n \times n$, which elements describes how many times stimulus i (rows) was preferred over stimulus j (columns). Using this matrix and assuming independence between pairs, the likelihood function of the model can be defined as follows:

$$L = \prod_{i<j} \pi_{ij}^{N_{ij}} \left(1 - \pi_{ij}\right)^{N_{ji}} \tag{5.2}$$

where i and j are row and column indices of the frequency matrix, respectively, and N_{ij} is the ijth element. Consequently, the maximum likelihood estimates (MLEs) can be calculated as the values that maximize equation 5.2. In general, this maximization is not straightforward and the MLEs are found by numerical optimization using iterative methods [43].

Additionally, the model imposes strong restrictions with respect to transitivity violations. Transitivity occurs when, for example given a set {A,B,C}, A is chosen over B and B is chosen over C, then A is chosen over C. This permits to sort the items on a preference scale. Accounting for transitivity violations is a way to evaluate how well the BTL model fits with the data and how meaningful is the calculated rating. There exists three main restrictions regarding transitivity: the weak (WST), moderate (MST) and strong (SST) stochastic transitivity which are defined as follows [8]:

If $P_{xy} \geq 0.5$ and $P_{yz} \geq 0.5$ then

$$P_{xy} \geq \begin{cases} 0.5 & (WST) \\ min\{P_{xy}, P_{yz}\} & (MST) \\ max\{P_{xy}, P_{yx}\} & (SST) \end{cases} \tag{5.3}$$

If the paired comparison data complies with this requirements, then it is possible to extract a preference ratio from it.

A statistical way to check how well the model accounts for the data is by comparing the likelihood of the model with the saturated model that fits the data perfectly [43]. In the saturated model the probabilities are calculated as:

$$\hat{\pi}_{ij} = \frac{N_{ij}}{N_{ij} + N_{ji}} \tag{5.4}$$

A chi-square statistic can be used as a decision tool on whether or not the BTL model accounts for the data:

$$\chi^2 = -2\log\left[\frac{L_{BTL}}{L_{SAT}}\right] \tag{5.5}$$

where L_{BTL} and L_{SAT} are the likelihood functions of the BTL and saturated model, respectively. Thus, the model can be rejected if the p value is less than 10%.

5.2 Results

A total of 19 subjects participated on the experiment (8 female and 11 male) and one subject was excluded, as explained in section 4.2.3 on page 41. As mentioned there the test consisted in three sessions: training, paired comparison with headphones and paired comparison with loudspeaker. After the training session each subject's responses were analyzed in order to select the signals and the minimum Q that were going to be assessed in the paired comparison sessions. Thus, not all subjects compared every signal and every Q.

Since the BTL model requires that all items had been compared the same amount of times, it was necessary to divide the data into groups. This is, for every signal there were up to three groups, e.g. subjects who assessed Qs from 4 to 64, subjects who assessed Qs from 8 to 64 and so forth. In the following, the results are presented for each signal. A BTL analysis was carried out for each group and the most relevant results are presented. For a more detailed description of the obtained data see appendix F on page 101.

The Matlab function *OptiPt.m* developed by Wickelmaier and Schmid in 2004 was used [43]. This function allows to estimate the BTL model parameters and measures the goodness of fit. The input is the frequency matrix and it returns the scale values, the chi-square statistics, the likelihoods of the model and the saturated and a matrix with predicted values. The later can be used to calculate an alternative chi-square test for goodness of fit by applying:

$$\tilde{\chi}^2 = \sum_{i,j} \frac{(N_{ij} - \hat{\pi}_{ij}N)^2}{\hat{\pi}_{ij}N} \tag{5.6}$$

where $N = N_{ij} + N_{ji}$. This accounts for the lack of fit between observed and predicted values [43].

The validity of the results was additionally ensured by checking the consistency of the responses of each subject. This was calculated as a percentage of the maximum possible inconsistent answers. An inconsistency happens, for example, when the pair {AB} is presented and B was preferred, but when the pair {BA} of the same stimulus is presented A was preferred. Thus, a single subject's response matrix will be consider to be consistent if only twos and zeros are found in it. Thus, the percentage of inconsistencies was calculated as follows:

$$\text{Percentage of Inconsistent Answers.} = \frac{n_{ones}}{2} * \frac{100}{P_{max}} \tag{5.7}$$

where n_{ones} is the number of ones found in the individual response matrices and P_{max} is the maximum number of pairs evaluated by each subject.

5.2.1 Speech Signal (300Hz)

The results for the speech signal were divided into two groups. The first group of 14 subjects evaluated the signal convolved with all-pass filters with Qs from 4 to 64. The second group of 18 subjects evaluated the signal down to a Q of 8. Table 5.1 on the next page describes the frequency matrix for the first group with headphones.

The predicted values obtained from the function *OptiPt.m* are shown in parenthesis as a measure of how well the model accounts for the data. Table 5.2 on the following page shows the number of transitivity violations found in the data. The goodness of fit of the model, $[\chi^2(10) = 11.7105, p = 0.3049]$, indicates that the BTL model accounts quite well for the data

	Original	64	32	16	8	4
Original	-	28 (27.9)	28 (27.9)	28 (27.6)	21 (21.7)	17 (13.5)
64	0 (0.08)	-	14 (12.8)	4 (4.70)	0 (0.27)	0 (0.07)
32	0 (0.09)	14 (15.1)	-	5 (5.37)	2 (0.32)	0 (0.08)
16	0 (0.39)	24 (23.2)	23 (22.6)	-	0 (1.32)	1 (0.36)
8	7 (6.23)	28 (27.7)	26 (27.6)	28 (26.6)	-	6 (5.91)
4	11 (14.4)	28 (27.9)	28 (27.9)	27 (27.6)	22 (22.0)	-

Table 5.1: *Frequency table for the speech signal (300Hz) with headphones and predicted values in parentesis (Rows are preferred over columns).*

and it must not be rejected. Using equation 5.6 on the previous page the alternative goodness of fit can be calculated: $\tilde{\chi}^2(10) = 14.4135$. This confirms the good fitting between the observed and predicted values and therefore the found SST violation can be assumed to not be severed.

WST	MST	SST
0	0	5

Table 5.2: *Transitivity violations for the speech signal with headphones*

Figure 5.1 on the facing page shows the BTL scale values obtained from the data and their respective 95% confidence intervals. From the figure two groups of stimuli can be derived. The first, composed of the original signal and the signal distorted with Q equal 4, are the highest rated. Even though from the scale it can be deduced that speech signal distorted with Q equal 4 is preferred over the rest, the difference between the rating of it and the rating of the original is rather small. Looking at the 95% confidence interval, one can conclude that there is not significant difference between both stimuli and that a distortion up to Q equal 4 has the same acceptance as the original signal, approximately. The second group consists in the signal distorted with Qs equal 8, 16, 32 and 64. It is clear from the figure that no distortion above Q equal 8 is accepted.

The inter-individual consistency was also analyzed. For this case most of the subjects answers showed to be quite consistent. Figure 5.2 on the next page shows the inter-individual inconsistencies for subjects of the first group. It can be noted that all the percentage of inconsistencies lie below 32%.

The subjects who evaluated the former stimuli through headphones, evaluated the same stimuli through loudspeaker as well. Therefore it is possible to make a parallel between both results. Table 5.3 on page 50 shows the transitivity violations for the speech signal with the loudspeaker configuration. In contrast with the violations found with the headphones, the ones found with the loudspeaker are considerable large. However, a better goodness of fit between the model and the data was found: $[\chi^2(10) = 10.3350, p = 0.4116]$ with an alternative goodness of fit from equation 5.6 on the previous page equal to $\tilde{\chi}^2(10) = 10.2216$. This results suggests that

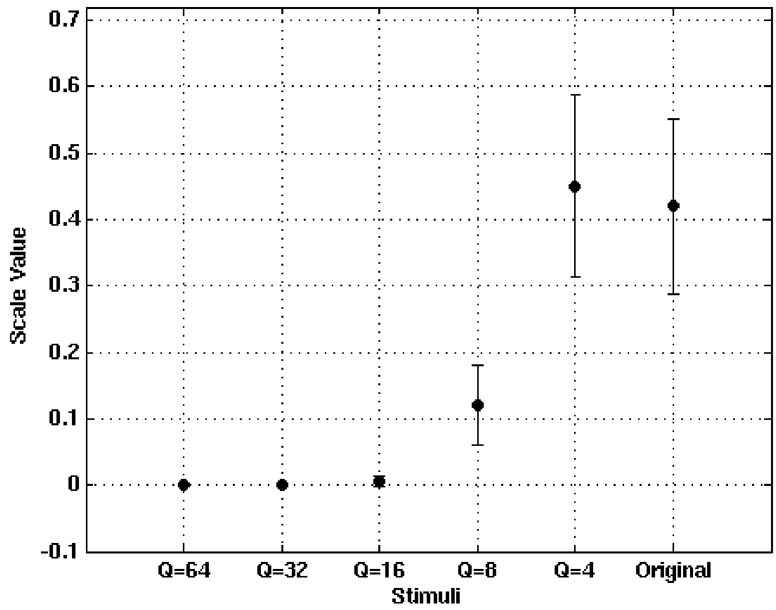

Figure 5.1: *Bradley-Terry-Luce scale values for the speech signal with headphones and the 95% confidence intervals.*

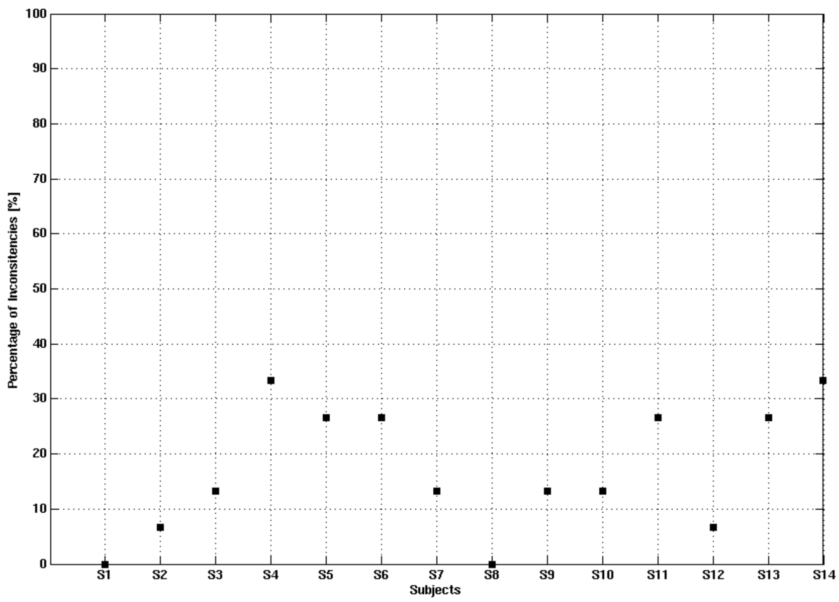

Figure 5.2: *Percentage of inter-individual inconsistencies for the speech signal with headphones.*

the BTL model accounts well from the data and can not be rejected.

WST	MST	SST
13	4	9

Table 5.3: *Transitivity violations for the speech signal with the loudspeaker.*

One possible explanation to the high amount of violations is the little differences perceived through the loudspeaker. This can be clearly seen in figure 5.3. All the stimuli, including the highly rated through headphones, were rated approximately the same which means that no significant difference were perceived through the loudspeaker.

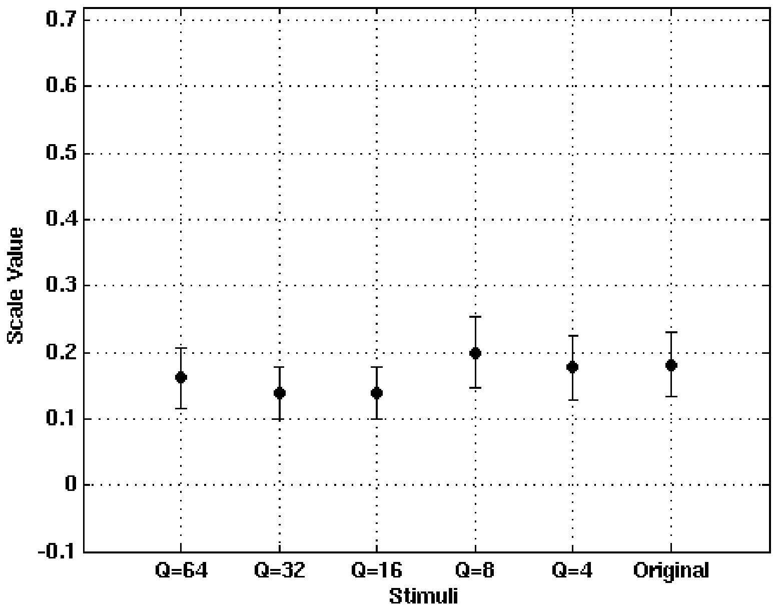

Figure 5.3: *Bradley-Terry-Luce scale values for the speech signal with the loudspeaker and the 95% confidence intervals.*

The hypothesis of non-noticeable differences between stimuli through loudspeakers can be verified by looking at the inter-individual consistency. Figure 5.4 on the next page depicts the inter-individual inconsistencies for the speech signal with the loudspeaker. It is observed that more than half of the subjects answers are highly inconsistent. Furthermore, most of the subjects commented after the test that while through headphones a difference was clearly perceivable through the loudspeaker they found it hard to discriminate the stimuli.

The results obtained from the other group of subjects are consequent to the first one. This implies that the results showed above can be consider to be representative for this signal. Details of the results obtained with the other groups of subjects can be found in appendix F.1 on page 101.

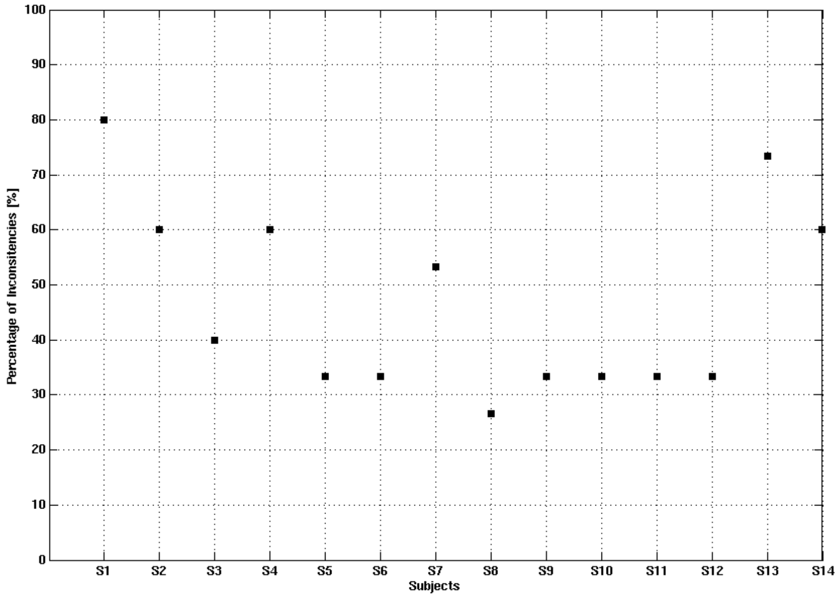

Figure 5.4: *Percentage of inter-individual inconsistencies for the speech signal with the loudspeaker.*

5.2.2 Electronic Gong (500Hz)

Similar to the speech results, the electronic gong results were divided into three groups: the first group of 5 subjects evaluated the original signal and the distorted with Qs from 8 to 64. The second group of 14 subjects evaluated the original signal and the distorted ones with Qs from 16 to 64. The third group of 18 subjects evaluated the original signal and the distorted with Qs equal to 32 and 64.

Table 5.4 shows the frequency matrix obtained from the second group with headphones. The calculated goodness of fit of the model, $[\chi^2(3) = 3.6458, p = 0.3023]$, and the alternative goodness of fit, $\tilde{\chi}^2(3) = 3.6706$, show that the model accounts well for the data. The absence of transitivity violations is a strong evidence that the BTL model can not be rejected.

	Original	64	32	16
Original	-	22 (24.0)	21 (21.0)	19 (16.9)
64	6 (3.98)	-	7 (9.35)	6 (5.66)
32	7 (6.95)	21 (18.6)	-	7 (9.40)
16	9 (11.0)	22 (22.3)	21 (18.5)	-

Table 5.4: *Frequency table for the electronic gong signal (500Hz) with headphones and predicted values in parenthesis (Rows are preferred over columns).*

Figure 5.5 on the following page plots the BTL scale values calculated throughout the function

OptiPt.m. In contrast with the speech signal, a monotonically decreasing with Q scale was obtained where the original signal has the highest rating and the most distorted signal (Q=64) has the smallest rating.

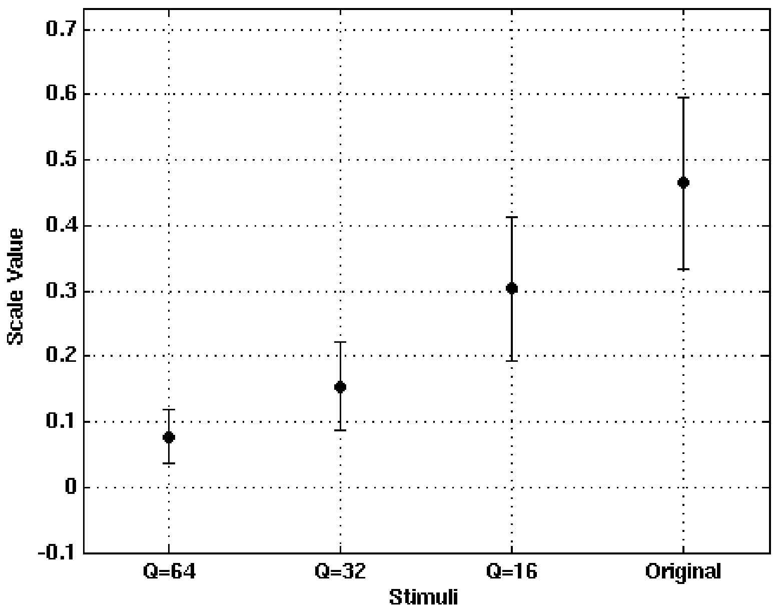

Figure 5.5: *Bradley-Terry-Luce scale values for the gong signal with the headphones and the 95% confidence intervals.*

Some subjects answers showed a high inconsistency percentage. Figure 5.6 on the next page shows the percentage of inter-individual inconsistencies. However, given the good fitting of the model with the data, those inconsistencies can be consider not to have a significant influence on the results.

In the case of the electronic gong played back with the loudspeaker the chi-square statistics suggest an approximately perfect fitting of the model with the data: $[\chi^2(3) = 0.4304, p = 0.9339]$ with an alternative goodness of fit $\tilde{\chi}^2(3) = 0.43004$. Therefore the transitivity violations shown in table 5.5 can be consider to be not severe and the model can not be rejected.

WST	MST	SST
4	1	5

Table 5.5: *Transitivity violations for the electronic gong with the loudspeaker.*

Figure 5.7 on page 54 shows the calculated BTL scale values for each assessed stimuli. In this case no clear rating is observed. Similar to the speech signal with the loudspeaker, there is not significant difference between the stimuli. This suggests that while through headphones there are noticeable differences, through the loudspeaker those differences become subtle.

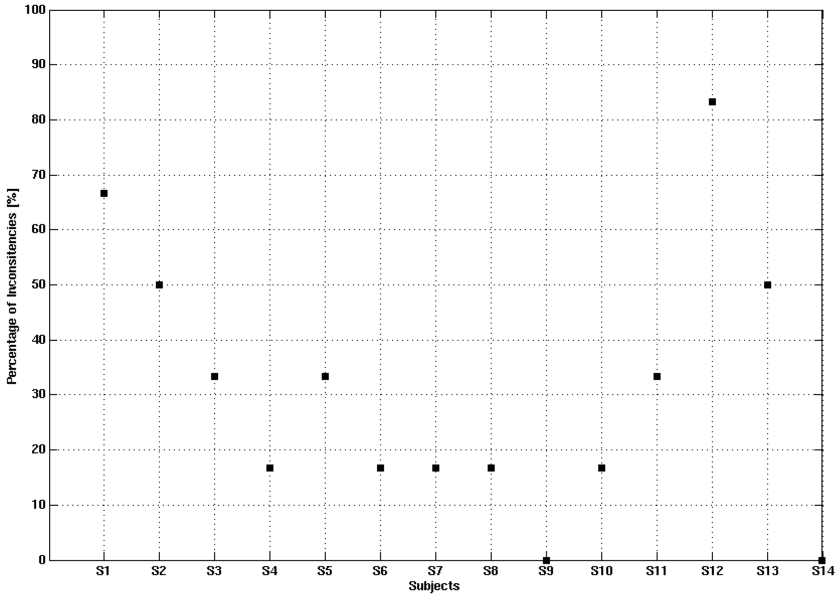

Figure 5.6: *Percentage of inter-individual inconsistencies for the electronic gong signal with headphones.*

The former statement is strongly supported by Figure 5.8 on the following page. It can be observed that just two subjects out of 14 had an inconsistency percentage below 50%, whereas for the answers of three subjects were completely inconsistent.

In the first group only five subjects evaluated the signal distorted down to a Q equal to 8. The obtained results are consequent with the ones obtained from the second group. This is, for the headphones a monotonically decreasing scale was obtained and for the loudspeaker no significant difference was observed. Their low percentage of inconsistencies indicates that the differences were audible most of the time and that the preference is relatively stable. Detailed results of this group and the third group can be found in appendix F.2 on page 105.

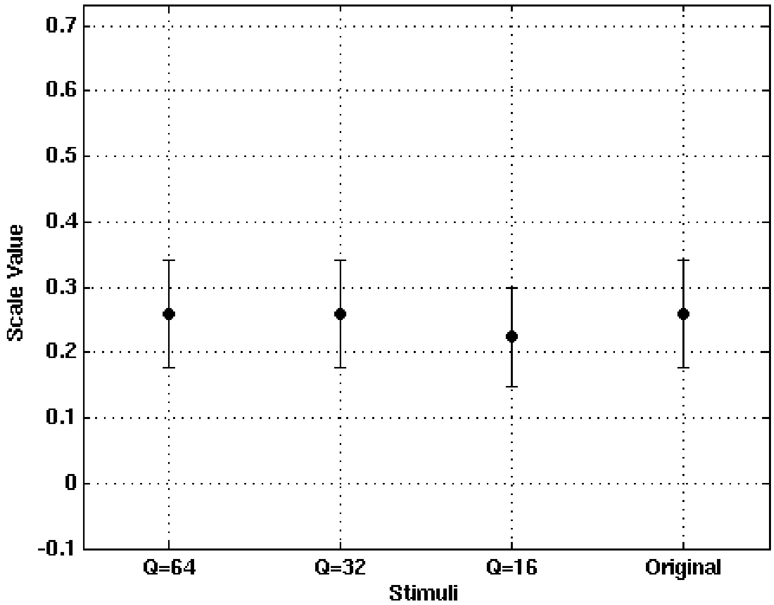

Figure 5.7: *Bradley-Terry-Luce scale values for the gong signal with the loudspeaker and the 95% confidence intervals.*

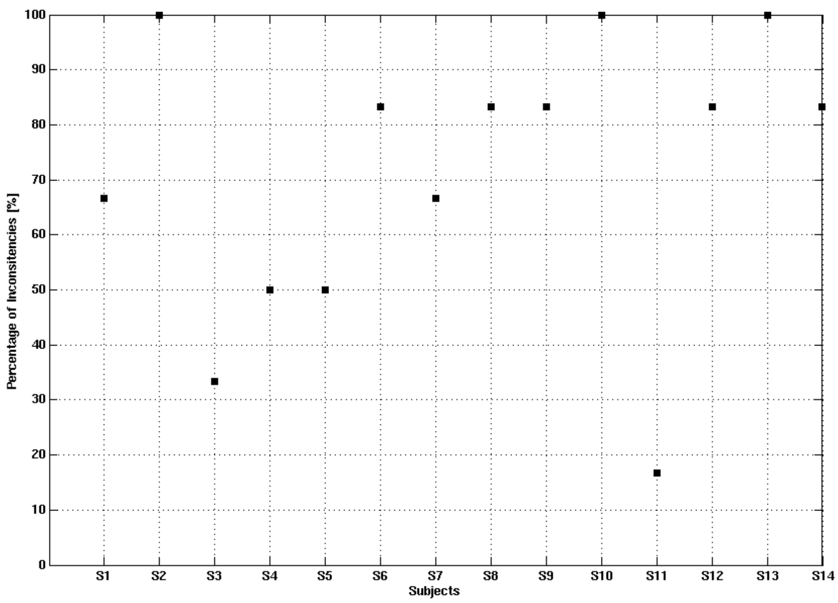

Figure 5.8: *Percentage of inter-individual inconsistencies for the electronic gong with the loudspeaker.*

5.2.3 Radiohead's Sample (100Hz)

Only ten subjects evaluated this signal. Four out of the ten assessed the signal distorted down to a Q equal to 16 and the rest only evaluated the signal distorted with Qs equal to 32 and 64. Table 5.6 shows the frequency matrix obtained with the second group.

	Original	64	32
Original	-	8 (9.31)	14 (12.6)
64	12 (10.6)	-	12 (13.3)
32	6 (7.31)	8 (6.68)	-

Table 5.6: *Frequency table for the Radiohead's sample (100Hz) with headphones and predicted values in parentesis (Rows are preferred over columns).*

The model accounts well for the data according to the chi-square statistics, [$\chi^2(1) = 1.1086$, $p = 0.2924$]. The alternative goodness of fit backs up this argument, [$\tilde{\chi}^2(1) = 1.1053$]. Additionally, only one transitivity violation was found, which means that the BTL model can not be rejected. Table 5.7 shows the transitivity violations for the signal with headphones.

WST	MST	SST
0	0	1

Table 5.7: *Transitivity violations for the Radiohead's sample with headphones.*

Figure 5.9 on the following page depicts the BTL scale values calculated for this group. It can be observed that the stimulus that got the highest rating was the signal distorted with Q equal 64. Nevertheless, the confidence interval suggests that differences in preference between the signal distorted with Q = 64 and the original signal are small.

The percentage of inter-individual inconsistencies, plotted in figure 5.10 on the next page, was found to be below 35% for all subjects. This indicates that the BTL scale found can not be disregarded. This indicates that the similarities found between the signal distorted with Q equal 64 and the original are not due to unnoticeable differences between the stimuli. Thus, it can be assume that the audible distortion is still subtle enough to affect the quality of reproduction for this kind of signals. Looking at the data obtained from the first group, who assessed the signal down to Q =16, one can sustain this argument (see appendix F.3 on page 112, table F.33 on page 113). Nonetheless, given the small number of subjects the data obtained from the first group can not be consider to be representative enough.

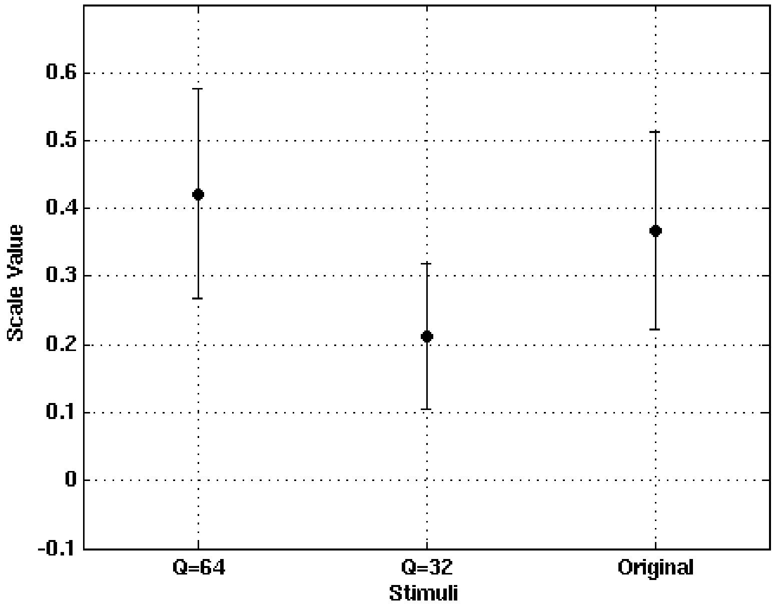

Figure 5.9: *Bradley-Terry-Luce scale values for the Radiohead's sample with the headphones and the 95% confidence intervals.*

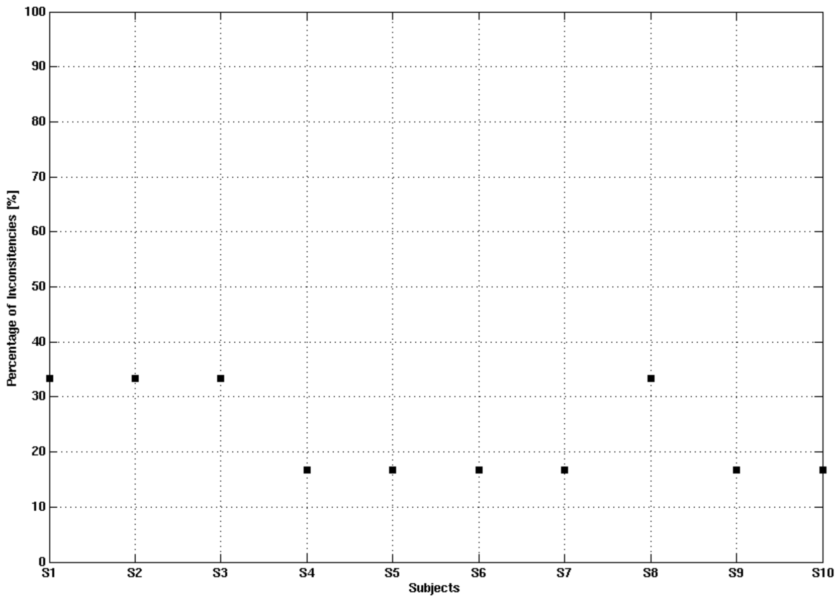

Figure 5.10: *Percentage of inter-individual inconsistencies for the Radiohead's sample with headphones.*

In the case of the loudspeaker the model was found to account for the data better than with the headphones: [$\chi^2(1) = 0.0665, p = 0.7965$], with an alternative goodness of fit $\tilde{\chi}^2(1) = 0.06646$. Additionally, no transitivity violations were found. Figure 5.11 shows the BTL scale values calculated for the second group with the loudspeaker. The scale shows no significant difference between stimuli. In a similar manner as with the speech and the electronic gong, the values lie close to mean and the confidence intervals suggest that there is not clear preference among the set.

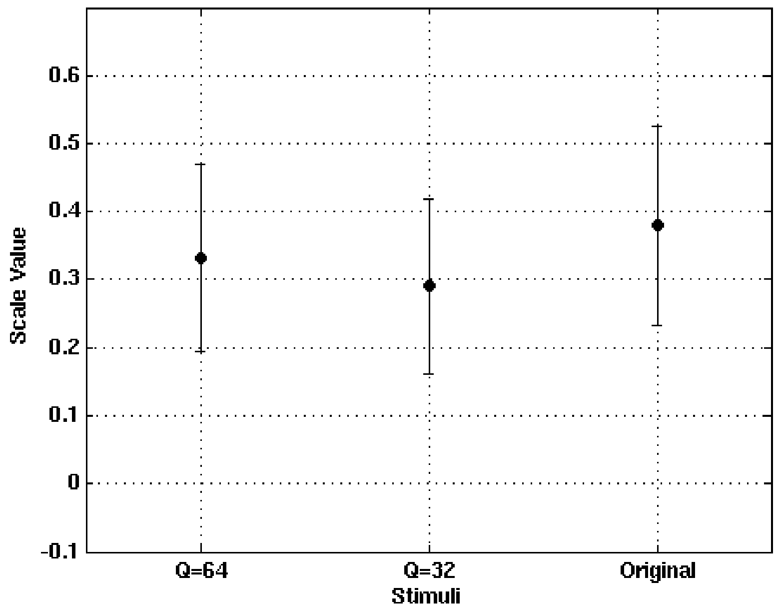

Figure 5.11: *Bradley-Terry-Luce scale values for the Radiohead's sample with the loudspeaker and the 95% confidence intervals.*

Despite the low percentage of inconsistencies found in figure 5.12 on the next page, results suggest that phase distortion is hardly audible for this signal through loudspeakers. Moreover, most of the subjects commented that specially for this signal it was impossible to hear a difference through the loudspeaker.

5.3 Summary of the Results

Three signals were assessed by means of a paired comparison experiment. The experiment was carried out through headphones and a loudspeaker and the data of the most representative results were presented. For the speech signal with headphones no scale could be derived. Subjects showed a high sensibility to the distortion and only a little amount of it was assessed as acceptable. In the case of the electronic gong a monotonically decreasing with Q scale can be assumed when played through the headphones. Subjects preferred the most the undistorted signal over

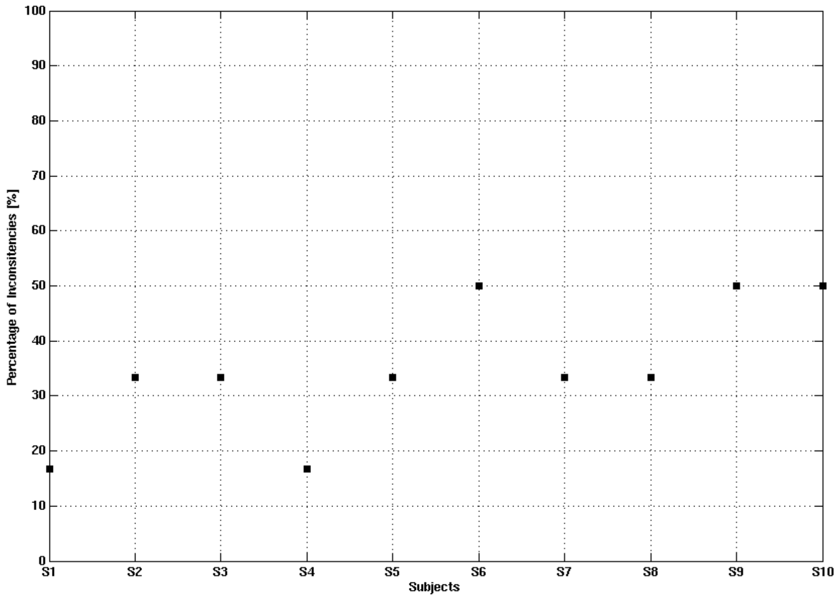

Figure 5.12: *Percentage of inter-individual inconsistencies for the Radiohead's sample with the loudspeaker.*

the others. For the Radiohead's sample no significant difference was found in the results, and subjects showed a low sensibility to the distortion. In general, no significant differences were found throughout the loudspeaker. Results showed that the distortion is hardly audible and subjects responses were not stable.

CHAPTER 6

DISCUSSION AND CONCLUSIONS

This chapter contains the discussion and conclusions of this thesis. A discussion about what has been researched in this work and which results have been found is described. Some parallels with previous studies are also presented. Afterwards, the general conclusions of this thesis are drawn and some suggestions for further studies are formulated.

Discussion

In chapter 1 on page 2 the following question was stated:

- Does phase distortion introduced by loudspeakers affects the perception of the reproduction quality?

In order to give an answer to this question, during the development of this thesis a number of issues regarding phase distortion were researched. A mathematical definition of phase distortion and related theory was given. Measurements of some transducers and an analysis of their characteristics were done. Furthermore, a threshold listening test was designed in order to select the stimuli which are more sensitive to phase distortion. Finally, a main listening test where preference among distorted and undistorted stimuli were measured was developed.

A considerable amount of previous studies regarding audibility of phase distortion were investigated. In general, these studies focused on the thresholds of the audibility of the distortion and made use of electronically generated signals to measure them. Some researchers studied the perception of the distortion in terms of coloration or artifacts, but none of the investigated works regarded the sound quality perception in normal listening environments.

In this project some measurements were carried out. High fidelity loudspeakers were analyzed with respect to their magnitude and phase characteristics. Three different kind of headphones were measured and analyzed as well. All the measured transducers showed approximately linear phase characteristics, and most of the loudspeakers presented fairly constant magnitude responses. Among the measured loudspeakers, the Genelec 1031A showed the flattest magnitude response and it was chosen for the main listening test. The Beyerdynamic DT-990 PRO headphones were selected for both listening tests due to their smooth magnitude response and linear phase characteristics. These headphones were additionally equalized in order to ensure control over the stimuli the subjects were about to assessed.

The first listening test consisted of a phase distortion threshold measurement of different signals. This test was carried out only with headphones. Seven different sounds were selected for the test among which were speech, pop music, electronic music and instrumental music. These signals were distorted with second order all-pass filters centered at different frequencies, giving a total of sixteen stimuli. An adaptive weighted up-down threshold algorithm where the Q was varied was implemented.

Results showed that for most of the signals the phase distortion was not audible for the selected frequencies. Signals such as speech and the electronic gong showed clearly audible differences, whereas for signals such as an anechoic guitar and pop-music the phase distortion was not audible. Previous studies concluded that phase distortion is more audible for impulsive sound [3]. However, this was not observed here for impulsive signals such as the castanets. According to Oppenheim [33] speech carries most of its information on the phase. The results found here are consistent with this and other studies where the phase distortion was found to be clearly audible in speech [17]. In most of the previous researches, electronically generated signals such as the electronic gong were also found to be sensitive to phase distortion.

Furthermore, it was found that training is a key factor in the ability to hear phase distortion. In addition, when the distortion was clearly discriminated some subjects described the differences as echos or reverberation while others described them as a synthetic effect or coloration. The signals which presented the lowest thresholds were the speech signal, the electronic gong and the Radiohead's sample at 100Hz. These signals were selected for the main listening test.

The purpose of main listening test was to evaluate the perceived quality of distorted signals. A pair comparison experiment was designed and it was ran through headphones and the chosen loudspeaker. Before the preference tests a training session was carried out using a modified version of the threshold listening test. The modification consisted of a smaller amount of trials and feedback.

For the paired comparison experiment a BTL analysis was done for each signal with each transducer. It was found that for the speech signal through headphones the distortion was not acceptable at all for Qs greater than four and no scale was observed. In contrast for the electronic gong a monotonically decreasing scale was obtained. The non-distorted signals were preferred over the distorted ones. In the case of the Radiohead's sample no significant differences were found and all the stimuli were rated approximately equal.

The obtained results with the loudspeakers showed that in general the phase distortion is not audible in normal listening environment. No significant differences were found and most of the subjects showed a high amount of inconsistent responses. This matches quite well with many researchers who stated that distortion was less audible through loudspeakers [21], [13], [14].

Conclusions

Based on the results of the main listening test the question formulated before can be answered: phase distortion does not affect the perception of the reproduction quality when played through loudspeakers in a normal listening environment. On reason might be that the reflected sound from the walls might mask the distortion and therefore make it inaudible. Additionally, since the measured loudspeakers have linear phase characteristics, it can be concluded that these loudspeakers do not have an impact on the phase of the played sound significantly.

On the other hand, it is concluded that phase distortion is audible and it degenerates the perceived sound quality for commonly heard sounds when reproduced through headphones. Although headphones are built commonly with only one driver for each ear, a few high-end headphones provide multiple drivers per ear, for example the Shure E500, the Ultimate Ears UE-10 or the Westone ES3, all with three drivers per ear. Therefore, crossover networks are also needed for these headphones and it might affect the phase of the reproduced sound.

In general crossover networks resembles all-pass filters with Q around $1/\sqrt{2}$. Thresholds measured in this work were found to be larger than Q equal four. Therefore, it can be concluded that typical crossover networks do not modify the reproduced sound in an audible way.

Nevertheless, a more exhaustive study of phase distortion thresholds should be carried out. In this work just single second order all-pass filter were evaluated and at high Qs they become more frequency selective. Thus, it would be desirable to make a more rigorous frequency analysis of the stimuli in order to find the exact frequencies at which the sounds are more affected by the phase distortion.

BIBLIOGRAPHY

[1] ISO 3382:1997. *Acoustics - Measurement of the Reverberation Time of Rooms with Reference to Other Acoustical Parameters*. ISO, 2005.

[2] Hideki Banno, Kazuya Takeda, and Fumitada Itakura. The Effect of Group Delay Spectrum on Timbre. *J. Acoust. Sci. Tech.*, 23(1):1–9, 2002.

[3] J. Blauert and P. Laws. Group Delay Distortions in Electroacoustical Systems. *Audio Engineering Society*, 63(5):1478–1483, May 1978.

[4] John Borwick. *Loudspeaker and Headphone Handbook*. Butterworth & Co., 1st edition, 1988. ISBN 0408013877.

[5] Ralph Allan Bradley and Milton E. Terry. Rank Analysis of Incomplete Block Designs. I. The Method of Paired Comparisons. *Biometrika*, 39(3/4):324–345, Dec. 1952.

[6] Thomas C. Brown, George L. Peterson, and Glen E. Brink. An Enquiry into the Method of Pair Comparison: Reliability, Scaling, and Thurstone's Law of Comparative Judgment. *Rocky Mountain Research Station, U.S. Forest Service. Forth Collins - Colorado 80526*, 18 January 2007. Discussion Paper.

[7] ITU-R BS.1116-1. *Methods for the Subjective Assessment of Small Impairments in Audio Systems Including Multichannel Sound Systems*. ITU, 1997.

[8] Sylvain Choisel and Florian Wickelmaier. Ratio-Scaling of Listener Preference of Multichannel Reproduced Sound. In *Proceedings of 31. Deutsche Jahrestagung für Akustik, DAGA '05*, pages 793–794, March 14-17 2005.

[9] Philippe Courcoux, Laurent Chaunier, Guy Della Valle, Denis Lourdin, and Michel Séménou. Paired Comparisons for the Evaluation of Crispness of Cereal Flakes by Untrained Assessors: Correlation with Descriptive Analysis and Acoustic Measurements. *Journal of Chemometrics*, 19:129–137, 2005.

[10] J. L. Flanagan. The Auditory System, 21 of March 2007. URL `http://cobweb.ecn.purdue.edu/~ee649/notes/figures/ear.gif`. After J. L. Flanagan, Speech Analysis and Perception, Springer-Verlag, Berlin, 2nd edition, 1965.

[11] Sheila Flanagan, Brian C. J. Moore, and Michael A. Stone. Discrimination of Group Delay in Clicklike Signals Presented via Headphones and Loudspeakers. *Audio Engineering Society*, 53(7/8):593–611, July/August 2005.

[12] H. Fleischer. Hörbarkeit von Phasenunterschieden bei verschiedenen Arten der Schalldarbietung. *Acustica*, 36:90–99, 1976.

[13] Villy Hansen and Erik Rørbæk Madsen. On Aural Phase Detection. *Audio Engineering Society*, 22(1):10–14, January/February 1974.

[14] Villy Hansen and Erik Rørbæk Madsen. On Aural Phase Detection: Part II. *Audio Engineering Society*, 22(10):783–788, December 1974.

[15] Hermann Helmholtz. *On the Sensations of Tone*. Dover publications, Inc., New York, 2 edition, 1954. ISBN 0-486-60753-4. Translated by A. J. Ellis.

[16] Martin Haydon Joan Gay Snodgrass, Gail Levy-Berger. *Human Experimental Psychology*. Oxford University Press, 1st edition, 1985. ISBN 0195035747.

[17] Lars G. Johansen and Per Rubak. The Excess Phase in Loudspeaker/-room Transfer Functions: Can it be Ignored in Equalization Tasks? In *100th Convention, Copenhagen, Denmark*. AES, May 1996.

[18] W. M. Leach Jr. The Differential Time-Delay Distortion and Differential Phase-Shift Distortion as Measures of Phase Linearity. *Audio Engineering Society*, 37(9):709–715, September 1989.

[19] Christian Kaernbach. Simple Adaptive Testing With the Weighted Up-Down Method. *Perception & Psychophysics*, 49:227–229, 1991.

[20] Christian Kaernbach. Adaptive Threshold Estimation with Unforced-choice Tasks. *Perception & Psychophysics*, 63(8):1377–1388, 2001.

[21] D. Koya. Aural Phase Distortion Detection. Master's thesis, University of Miami, 2000.

[22] H. Levitt. Transformed Up-Down Methods in Psychoacoustics. *The Journal of the Acoustical Society of America*, 49(2 (Part 2)):467–477, 1971.

[23] Siegfried H. Linkwitz. Active Crossover Networks for Noncoincident Drivers. *Audio Engineering Society*, 24(1):2–8, January/February 1976.

[24] Stanley P. Lipshitz, Mark Pocock, and John Vanderkooy. On the Audibility of Midrange Phase Distortion in Audio Systems. *Audio Engineering Society.*, 30:580–595, 1982.

[25] Henrik Møller, Pauli Minnaar, Søren Krarup Olesen, Flemming Christensen, and Jan Plogsties. On the Audibility of All-Pass Phase in Electroacoustical Transfer Functions. *Audio Engineering Society*, 55(3):115–133, March 2007.

[26] R. C. Mathes and R. L. Miller. Phase Effects in Monaural Perception. *Journal of the Acoustical Society of America*, 19(5):780–797, September 1947.

[27] Pauli Minnaar, Jan Plogsties, Søren Krarup Olesen, Flemming Christensen, and Henrik Møller. The Perception of All-Pass Components in Transfer Functions. *Proceedings of Forum Acusticum*, 2002.

[28] Brian C.J. Moore. *An Introduction to the Psychology of Hearing*. Elsevier Academic Press, 5th edition, 2004. ISBN 0125056281.

[29] G. S. Ohm. Über die Definition des Tones, nebst daran geknüpfter Theorie der Sirene und ähnlicher tonbildender Vorrichtungen. *Ann. Phys. Chem.*, 59:513–565, 1843.

[30] G. S. Ohm. Noch ein Paar Worte über die Definition des Tones. *Ann. Phys. Chem.*, 62: 1–18, 1844.

[31] Bang & Olufsen. CD: Music for Archimedes, October 1992. Details presented at the 87th Convention of the Audio Engineering Society (AES) by Villy Hansen and Gert Munch.

[32] Alan V. Oppenheim and Ronald W. Schafer. *Discrete-Time Signal Processing*. Prentice-Hall, Inc., 1999. ISBN 0130834432.

[33] Alan V. Oppenheim, Jae S. Lim, Gary Kopec, and Stephen C. Pohlig. Phase in Speech and Pictures. In *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '79*, volume 4, pages 632–637. IEEE, April 1979.

[34] D. Preis. Phase Distortion and Phase Equalization in Audio Signal Processing - A Tutorial Review. *Audio Engineering Society*, 30(11):774–794, 1982.

[35] D. Preis and P. J. Bloom. Perception of Phase Distortion in Anti-Alias Filters. *Audio Engineering Society*, 32(11):842–847, November 1984.

[36] Manfred R. Schroeder. Models of Hearing. *Proceedings of the IEEE*, 63(9):1332–1350, September 1975.

[37] Hideo Suzuki, Shigeru Morita, and Takeo Shindo. On the Perception of Phase Distortion. *Audio Engineering Society*, 28(9):570–574, September 1980.

[38] Floyd E. Toole. Loudspeaker Measurements and Their Relationship to Listener Preferences: Part 1. *Audio Engineering Society*, 34(4):227–235, 1986.

[39] Floyd E. Toole. Loudspeaker Measurements and Their Relationship to Listener Preferences: Part 2. *Audio Engineering Society*, 34(5):323–348, 1986.

[40] Europeand Broadcasting Union. CD: Sound Quality Assessment Material (SQAM) [Recordings for subjective tests], 1988. http://www.ebu.ch/en/technical/publications/tech3000_series/tech3253/index.php.

[41] Indiana University. Cross Section of the Cochlea, 21 of March 2007. URL http://physics.indiana.edu/~p105_f02/cochlea.gif.

[42] Florian Wickelmaier and Sylvain Choisel. Modeling Within-Pair Order Effects in Paired-Comparison Judgments. In *Proceedings of Fechner Day 2006, 22nd Annual Meeting of the International Society for Psychophysics*, pages 89–94, July 25-26 2006.

[43] Florian Wickelmaier and Christian Schmid. A Matlab Function to Estimate Choice Model Parameters From Paired-Comparison Data. *Behavior Research Methods, Instruments, & Computers*, 36(1):29–40, 2004.

[44] Wikipedia. Section through the Spiral Organ of Corti, 21 of March 2007. URL http://en.wikipedia.org/wiki/Image:Gray931.png. Reproduction from Gray's Anatomy of the Human Body, 20th U.S. edition, 1918.

APPENDIX A

MEASUREMENT REPORTS

A.1 Loudspeaker Measurement

Purpose

In order to design a proper subjective evaluation of the phase distortion added to the acoustical signals by loudspeakers, it is necessary to analyze and understand their phase and magnitude characteristics. The purpose of these measurements is to obtain the main characteristics which describes the acoustical performance of certain loudspeakers. Measurements of frequency response and radiation pattern of five loudspeakers were done.

Equipment

The table A.1 lists the equipment used during the measurement.

Equipment	Type	Serial Number	Model
Microphone	B&K	207418	4133
Microphone pre-amplifier	B&K	971091	2619
MLS system	MLSSA	26827	Toshiba
Power amplifier	Pioneer	08340	A-616
Measuring amplifier	B&K	08022	2636
Calibrator	B&K	08155	4230
Anechoic Chamber	B4-111	—	—
Rotating base and Controller	B&K	—	—

Table A.1: *List of equipment used in the measurements.*

Loudspeakers

Type	Model	Serial Number
B&W	DM60152	2144-03
DALI	Dali 2A	00236 – 05
Genelec	1031A	33987 – 00
B&O	Beolab 6000	—
KEF	Reference series 107	002733R

Table A.2: *List of measured loudspeakers.*

Setup

The measurements were carried out in the anechoic chamber. All the loudspeakers were placed 1 m from the microphone and mounted on a rotating base. The impulse responses were acquired through an MLSSA and the base was rotated with a resolution of 10° steps from 0° to 100°. Those angles were chosen based on symmetry assumption. A single lobe is also assume therefore no angles on the back where measured. Figure A.1 on the facing page depicts the setup and table A.3 shows the setup of the equipment.

Equipment	Parameter	Value
Measuring amplifier	Input gain	40 dB
	Output gain	20 dB
MLSSA	Length	65536 $samples$
	Sample rate	48.125kHz
	Antialiasing filter	Chebyshev
	Preaverage	16
	Bandwidth	22.05 kHz
Power amplifier	Gain	0 dB
Microphone	Sensibility	13.71 mV/Pa
Rotating base	Resolution	10°

Table A.3: *Equipment setup.*

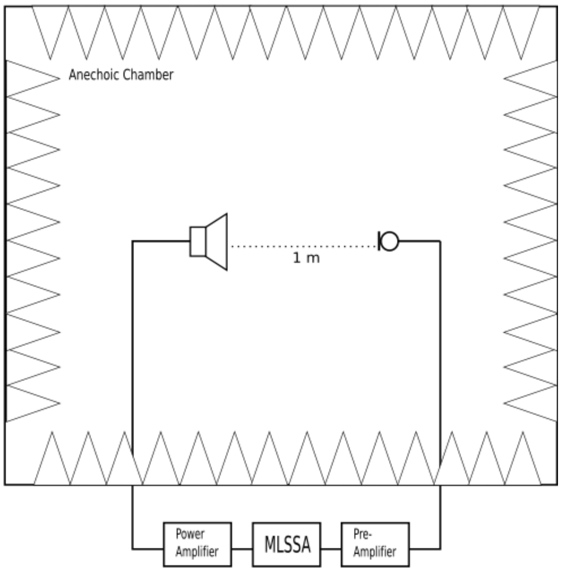

Figure A.1: *Measurement setup for measuring radiation characteristics of the loudspeakers.*

Results

Signal to noise ratio

In order to have control and ensure a good signal to noise ratio (SNR) in the measurements an acoustic analysis was performed through the MLSSA software for each loudspeaker. The software has an option which allows to do an IEC octal band analysis of the measured signals. Table A.4 shows the on axis results for each loudspeaker.

Loupspeaker	125 Hz	250Hz	500Hz	1kHz	2kHz	4kHz	8kHz
B&W	22,6	29,8	27,9	30,0	32,4	33,8	41,0
DALI	31,2	26,3	31,6	33,1	30,6	27,9	29,0
GENELEC	42,0	34,3	36,8	35,0	32,9	30,1	31,1
B&O	50	48,1	50,3	47,5	46,6	45,6	43,3
KEF	25,3	25,2	31,1	36,0	40,2	41,6	39,6

Table A.4: *IEC octal band signal to noise ratio [dB].*

Magnitude response

Figure A.2 on the next page shows the magnitude response of the five loudspeakers on the acoustic axis.

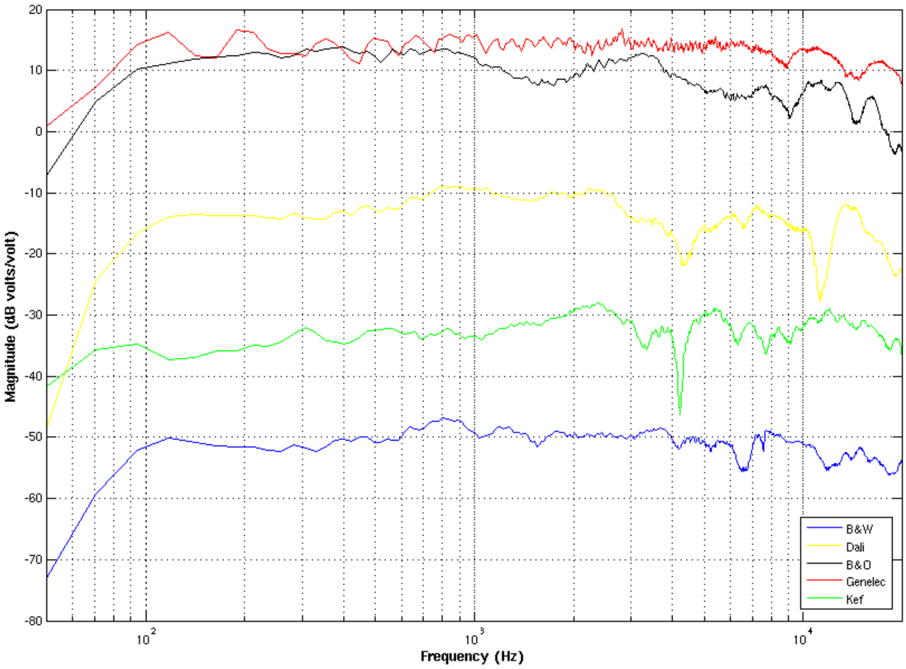

Figure A.2: *Magnitude response.*

Phase results

Figure A.3 on the facing page shows the phase response of the loudspeakers. It can be noted that all the measured loudspeakers present an approximately linear phase.

Figure A.4 on the next page shows the phase delay, the group delay and the differential time delay distortion ($\Delta\tau$ defined in equation 2.9 on page 9). The later is defined as the difference between the group delay and the phase delay, and it was suggested by Lipshitz et al. as a measure of the amount of distortion introduced. The smaller the $\Delta\tau$ is the smaller the first order distortion is [24].

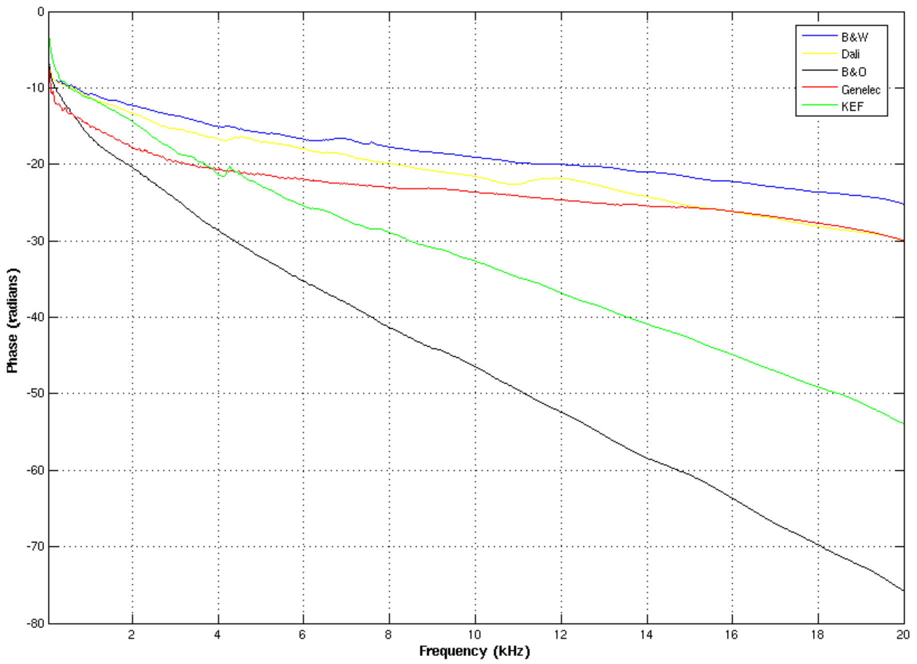

Figure A.3: *Phase response.*

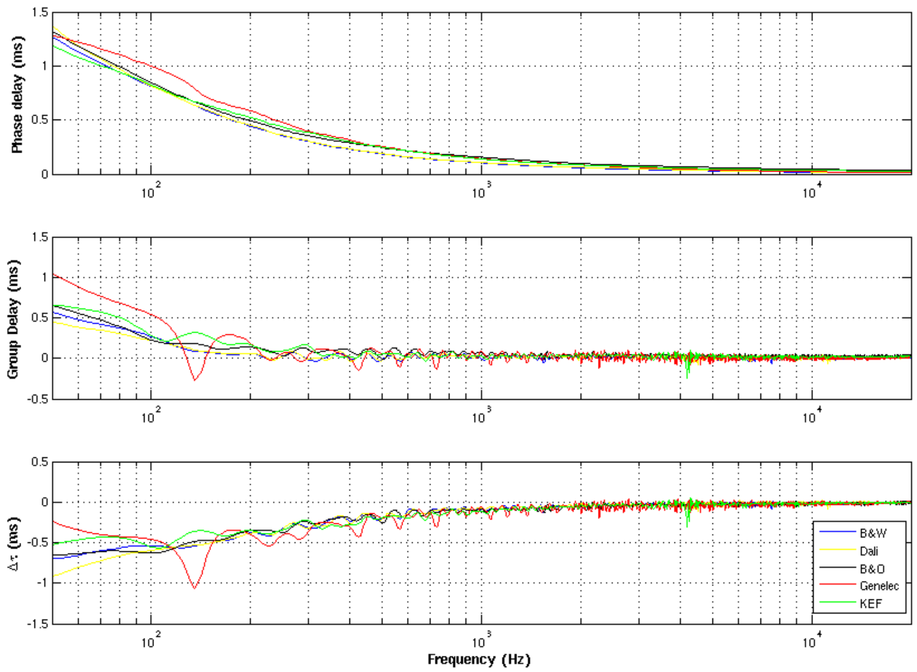

Figure A.4: *Phase delay, group delay and differential time delay distortion.*

Directivity pattern

The following graphs show the measured directivity pattern of the loudspeakers. Due to the size and weight of the KEF loudspeaker its directivity pattern was not measured for practical reasons.

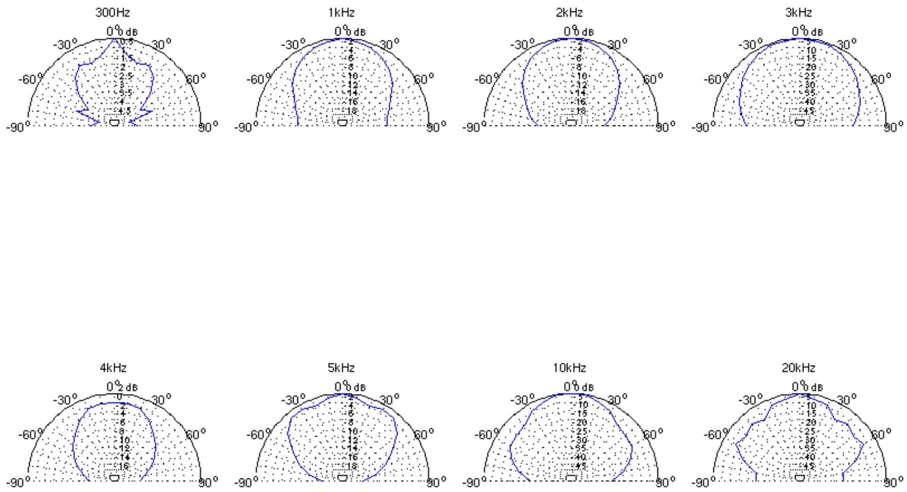

Figure A.5: *Directivity pattern of the B&W loudspeaker.*

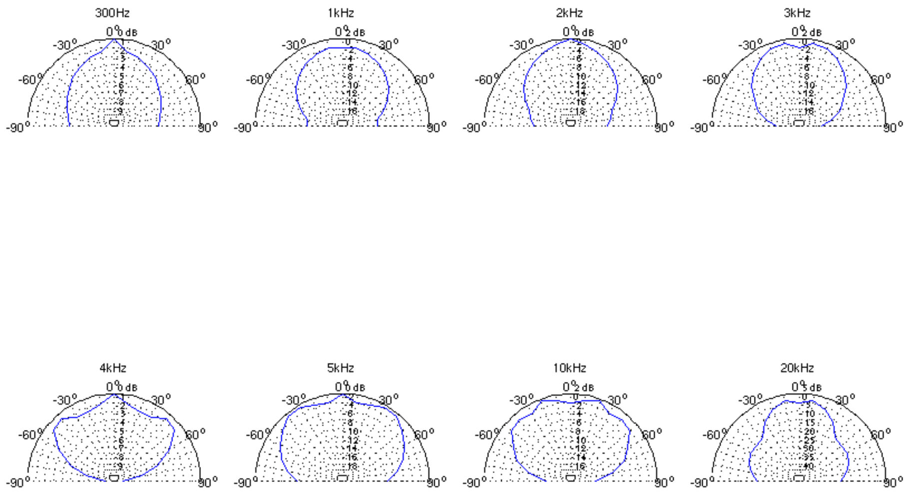

Figure A.6: *Directivity pattern of the DALI loudspeaker.*

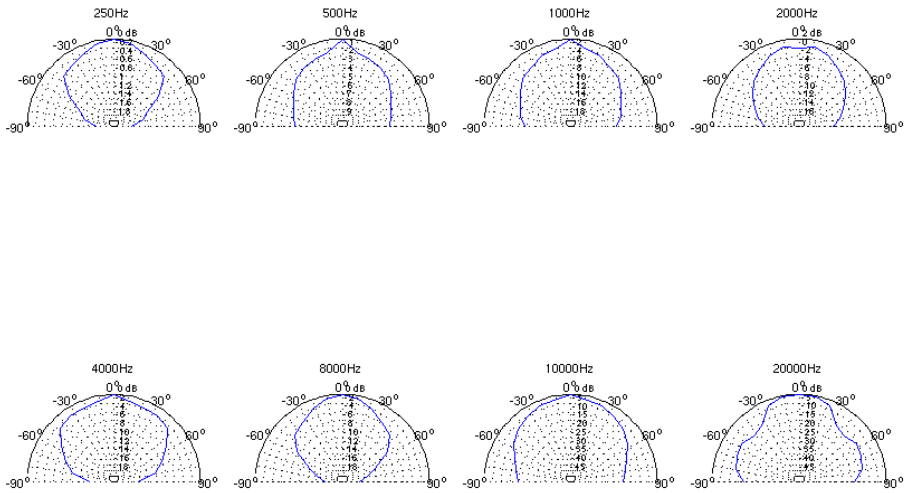

Figure A.7: *Directivity pattern of the Genelec loudspeaker.*

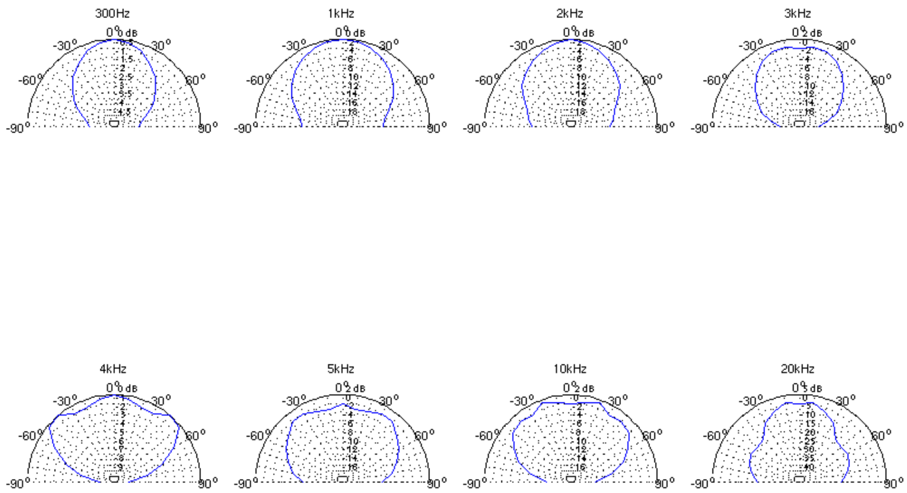

Figure A.8: *Directivity pattern of the B&O loudspeaker.*

A.2 Headphone Measurements

Purpose

The purpose of this measurements is to select the headphones that best fits the project requirements. This is, smooth frequency response and phase response close to zero. The smoother the frequency response the better the equalization that can be achieved.

Equipment

Table A.5 lists the equipment used during the measurements.

Equipment	Type	Serial Number
MLS system	Toshiba MLSSA	26827
Dummyhead	VALDEMAR	2150-01
Artificial Ear	B&K 4153	342467
Microphone	B&K 4134	1253168
Microphone Pre-Amp	B&K 2639	08639
Measurement Amp	B&K 2636	08022
Calibrator	B&K 4230	08155

Table A.5: *List of measurement equipment for headphone measurments.*

Headphones

Table A.6 lists the headphones selected for the measurement. Two electrostatic and one electrodynamic headphones were chosen for the measurement. All the headphones have only one driver, therefore there is no crossover network in any of them, and only the drivers physical characteristics will affect the phase.

Manufacturer	Model	Serial Number	Headphone Amp.	Serial Number
Beyerdynamic	DT 990 Pro	2036-21	Fostex PH-5	02092
Sennheiser	HE 60	33675-00	Sennheiser HEV 70	33676-00
Stax	SR Lambda Pro	2030-00	Stax SRM-1/MK2	33117

Table A.6: *List of measured headphones and used headphone amplifiers.*

A.2.1 Setup

Figure A.9 on the next page shows the setup of the headphones impulse response measurement.

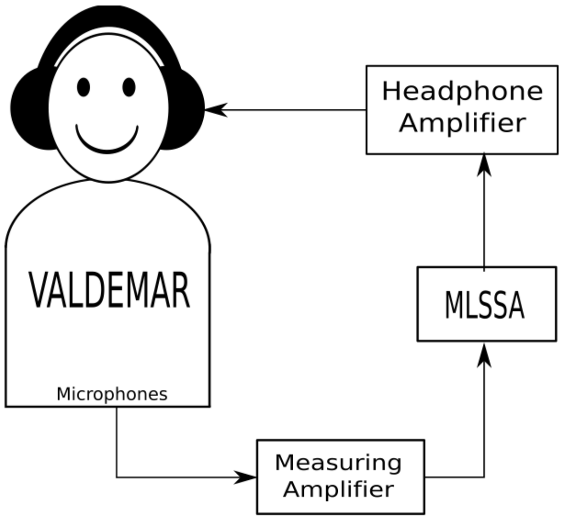

Figure A.9: *Headphones measurement setup.*

A.2.2 Results

The headphones were measured using an artificial ear and VALDEMAR. The artificial ear measurements were carried out in order to ensure that the results obtained with VALDEMAR were correct. VALDEMAR's measurements are considered more relevant given that resembles better a human head. Figure A.10 on the facing page depicts the headphones magnitude response obtained with VALDEMAR.

The figure shows noticeable differences between headphones. The Sennheiser shows a steep roll off at 300hz of about 8dB per decade, whereas the Beyerdynamic and the Stax shows a roll off of about 3dB around 150Hz. The Stax headphones presents a pronounced dip around 3kHz. The Beyerdynamics shows the smoothest magnitude response of the headphones selected. Figure A.11 on the next page shows the phase delay, the group delay and the differential time delay of the three headphones.

The Beyerdynamics shows zero phase delay above 300Hz. In contrast, the Stax and the Sennheiser phase delay begins to be zero above 1kHz. The group delay of the three headphones is nearly zero above around 300Hz, and below 0.2ms for lower frequencies. Only the Sennheiser presents some peaks and dips at certain frequencies. In general, the Beyerdynamics shows the smallest group delay and differential time delay among the selected headphones.

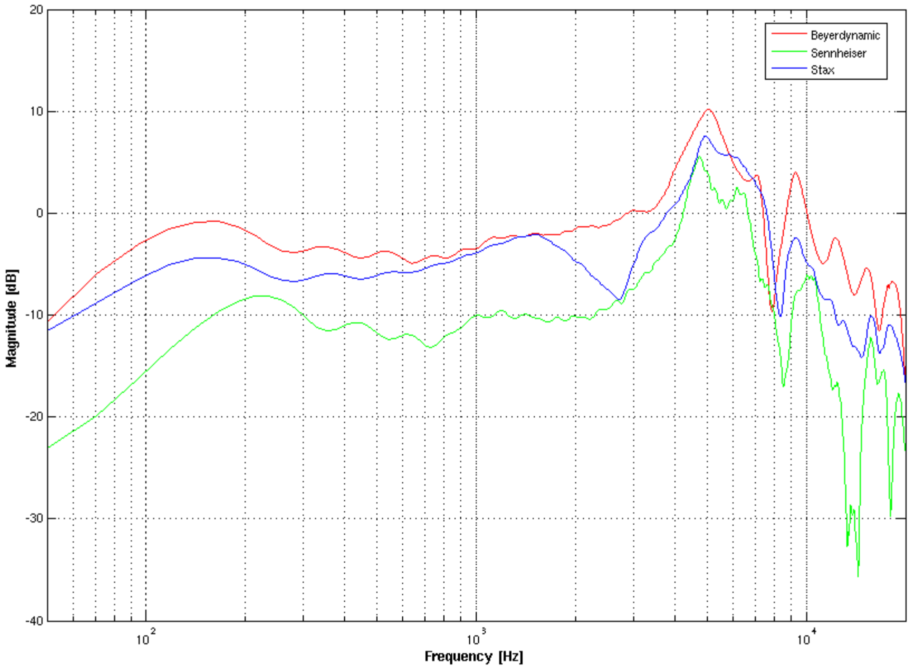

Figure A.10: *Headphones magnitude response with VALDEMAR.*

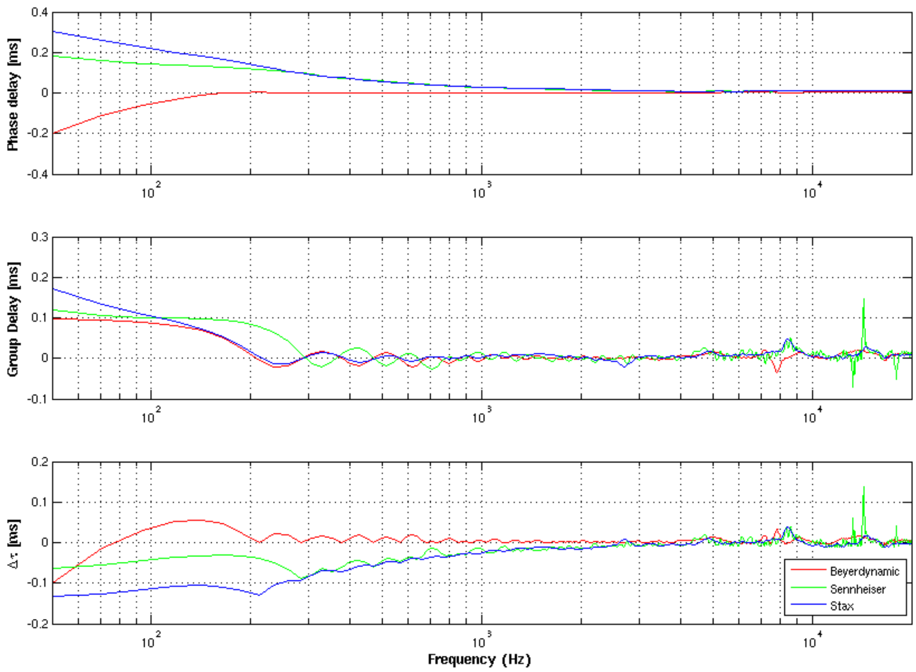

Figure A.11: *Phase delay, group delay and differential time delay distortion.*

A.3 Headphones Equalization

In order to ensure that the frequency responses of the stimuli are not affected by the headphones, they must be equalized. The Beyerdynamic DT-990 Professional headphones were chosen for the test given their smooth frequency response and their approximately linear phase characteristics (see figures A.10 on the preceding page and A.11 on the previous page). The equalization is achieved by including an inverse filter into the reproduction chain. The inverse filter has an impulse response such that:

$$\begin{aligned} H(z)H_i(z) &= 1, \quad \text{thus} \\ H_i(z) &= \tfrac{1}{H(z)} \end{aligned} \tag{A.1}$$

where $H(z)$ is the z-transform of the headphones impulse response and $H_i(z)$ is the inverse filter. In order to ensure an stable and causal inverse filter, $H(z)$ must be minimum phase [32]. In general, acoustic transducers such as headphones and loudspeakers have non-minimum phase characteristics. Therefore, before acquiring the inverse filter it is necessary to extract the minimum phase components of the impulse response. As explained in section 2.3 on page 10 any rational system can be decomposed into its minimum phase components and all-pass components. Thus, the equalization assumes that by equalizing the minimum phase components the phase of the system is not significantly affected.

The headphone calibration was accomplished by estimating the inverse filter of the headphones magnitude response throughout a Yule-Walker approximation. This method creates a minimum phase approximation and obtains an IIR digital filters which fits to a specified frequency response. The Matlab function *yulewalk* was used to derived a 35 order IIR digital filter. The criteria to choose the order was an order high enough to get a good magnitude approximation and low enough not to affect the phase.

A.3.1 Measurement and Results

To get a flat frequency response from the headphones when mounted on the head, one must carry out the measurements in such position. Thus, for practical reasons the Beyerdynamic DT-990 were measure using VALDEMAR. Even though VALDEMAR differs from human heads and the impulse response of the headphones are expected to vary among the subjects, this was considered a good approximation. In order to obtain a better approximation, the impulse was measured five times to account for differences due to fitting. Table A.7 on the next page shows the list of equipment used during the measurements.

Figure A.12 on the facing page shows the inverse filter obtained by Matlab's *yulewalk* function. The figure shows a good approximation to the desired magnitude response. Since the measurement equipment is band limited, the measurement data at 0Hz does not come from a real

Equipment	Type	AAU Number
MLS system	MLSSA	26827
Measuring Amp.	B&K	08022
VALDEMAR	AAU	2150-01
Microphone	GRAS 40AD - VALDEMAR right ear	−−
Microphone	GRAS 40AD - VALDERMA left ear	−−
Headphones	Beyerdynamic DT990	2036-21
Headphone Amp.	Fostex PH-5	02092-00
Calibrator	B&K 4230	08155

Table A.7: *Equipment used in the measurement of DT990 headphone impulse response.*

measurement but rather from noise introduced by the measurement equipment. Therefore, the 0Hz component of the filter was set to 0dB manually.

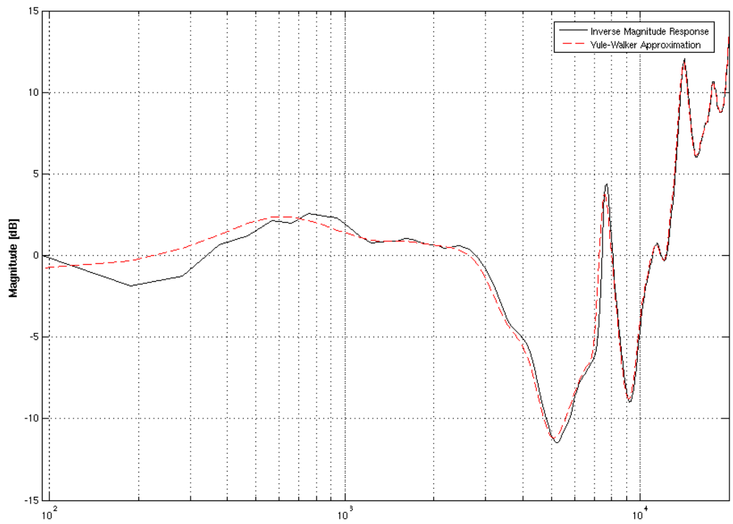

Figure A.12: *Yule-Walker approximation of the headphones inverse filter.*

One of the five impulses measured was used to test how well the equalization works. Figure A.13 on the next page shows the results. It can be noted that the Yule-Walker approximation, fits well the magnitude response and the equalized result is approximately flat. Additionally, the phase of the headphones is not significantly affected by equalization, which ensures that no artifacts or phase distortion will be introduced by the headphones during the play back.

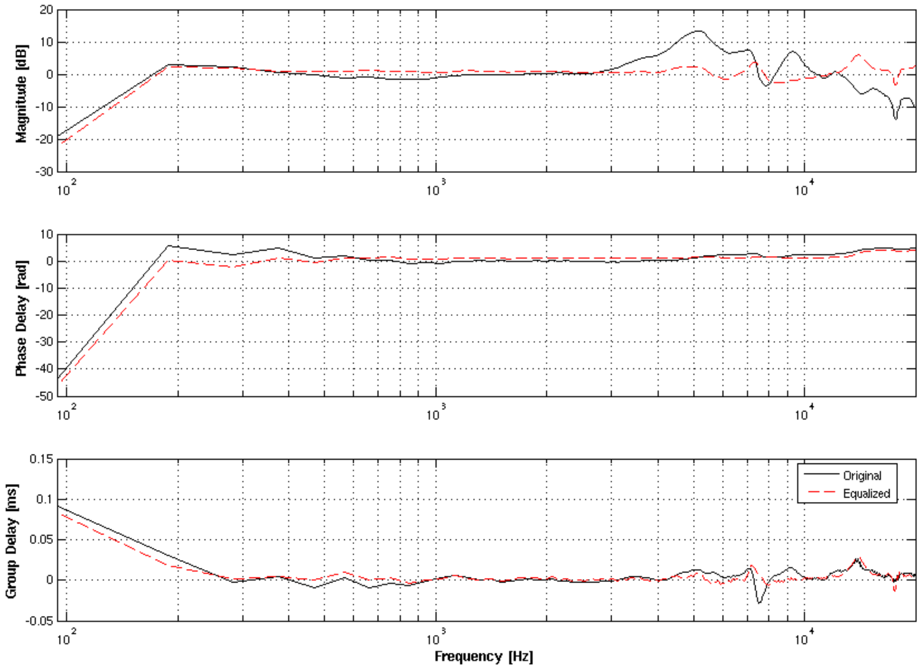

Figure A.13: *Equalization results for the headphones Beyer Dynamic DT-990 PRO.*

APPENDIX B

INSTRUCTION SHEETS

B.1 Threshold Listening Test Instruction Sheet

Listening Test Description

Welcome to our master's project listening test!

The experiment is divided in 2 main sessions of similar characteristics. Each session lasts approximately 20 minutes. There is a break between the two sessions where you can leave the room and eat some snacks. Each session is divided in two sub-sessions. Between the sub-session a message saying "it's time for a short brake" will appear. You are encourage to take off the headphones and relax for some minutes inside de room before continuing.

Each session consists on the presentation of a set of 3 consecutive sounds through headphones. As the sounds are played back the labels 1, 2 and 3 will appear on the screen corresponding to each signal. Two of the sounds you will hear are equal and one is of similar characteristics but different. This will be presented in a random way and your tasks consists in identifying which of the three sounds *you think* is different. There are no right or wrong answers. The difference may be subtle in some cases and we encourage you to try to detect the one which *sounds different to you* or do your best guess.

The test will start with a short training session for you to get familiarized with the task. If you have any questions regarding the procedure don't hesitate to ask.

Thank you in advance for your cooperation.

Group 1064

B.2 Main Test Instruction Sheet

Welcome to our master's project listening test!

Listening Test Description

You are about to participate in the main test of our master's project. The aim of the project is to study the effects of phase distortion in sound quality perception. Your participation is really important for the development of it. All the sounds you will hear will be played at comfortable levels where your hearing won't be compromised. You are allowed to quit the experiment at any moment without need of justification.

The experiment is divided into 3 sessions. The duration of the first session is around 8 minutes. The duration of the second and third sessions is approx. 15 min per session. There will be two breaks one after each sessions, when you will be able to leave the room, relax and eat some cake. After the first session you will receive instructions for sessions number 2 and 3. The following gives description of the first session. Read it carefully and if you have any questions regarding the procedure don't hesitate to ask.

Instructions Session 1

The first session consists of the presentation of a set of 3 consecutive sounds through headphones. As the sounds are played back the labels 1, 2 and 3 will appear on the box corresponding to each signal. Two of the sounds you will hear are equal and one is of similar characteristics but different. This will be presented in a random way and your task is to identify which one of the three sounds is different. After making your selection, a message saying whether your choice was correct or not will appear. Additionally, the label of box corresponding to the right answer will be displayed. The test will start with a short training session for you to get familiarized with the task.

Thank you in advance for your cooperation!

Group 1064

Instructions Session 2 and 3

In the following sessions you will listen to sounds played back through headphones and loudspeakers. You will have a break between sessions where you can relax and eat some cake. In session 2 the sounds will be played back through headphones. In session 3 the sounds will be played back through a loudspeaker and you will be asked to keep your head still. Some adjustments of the height of the chair will be done and this might take a few minutes.

In both sessions two sounds will be presented to you. As the sounds are played back the labels A and B will appear on the box corresponding to each sound. Your task is to select the sound ***YOU PREFER*** the most. There will be a button with the message “play again” which will allow you to play the sequence again in case you don’t feel sure about your preference. However, we encourage you to chose the sound you preferred at first as much as you can. There is no right or wrong answer and we expect you to give us your honest answer about what in ***YOUR OPINION*** sounds better.

Thank you for your participation!

Group 1064

APPENDIX C

MAIN TEST SETUP

This chapter describes the listening room and the test setup for the main listening test. It is shown how well the listening room fits the recommendations of ITU-R BS.1116 [7] for a monophonic loudspeaker setup in terms of its dimensions and reverberation time.

C.1 Listening Room

The main test was performed in the standard listening room of Aalborg University, which conforms to the IEC 268-13 standard for an average listening room. This room has a very low background noise level.

It is has a width of 4.1m, a length of 7.8m and a maximal hight of 2.78m. It is nearly symmetrical to it's vertical plane and has a rectangular ground floor, the volume is 85.6m^3. Figure C.1 shows a cross-section of this room. Table C.1 on the next page compares the dimensions of this room with the recommended dimensions in ITU-R BS.1116 for mono listening setups. It can be seen that it fits the recommendation nearly.

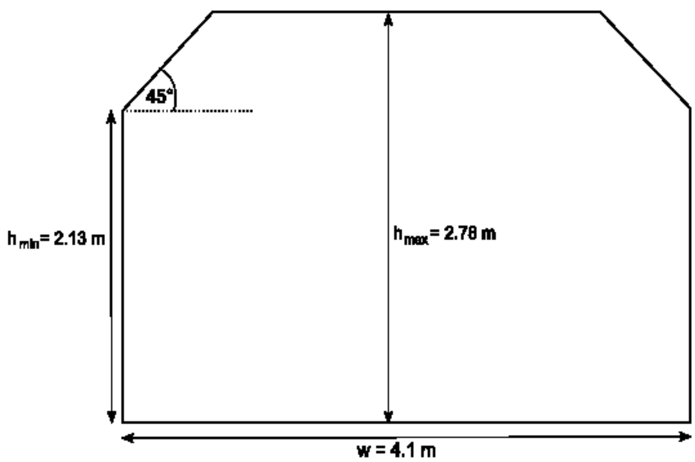

Figure C.1: *Cross-section of the listening room.*

	ITU-R BS.1116	Listening Room
Room Size	$20-60\text{m}^2$	32m^2
Dimension Ratios	$1.1 w/h \leq l/h \leq 4.5 w/h - 4$	$1.63 < 2.8$ **> 2.7**
	$l/h < 3$	$2.8 < 3$
	$w/h < 3$	$1.5 < 3$

Table C.1: *Comparison between listening room and ITU-R BS.1116; with l length, w width and h the maximum hight of the listening room.*

C.2 Reverberation Time Measurement

The reverberation time was measured in the listening room using the interrupted noise method described in ISO 3382 [1].

For the measurement an omni-directional loudspeaker was used to excite the room with pink noise. The measurement results of 3 different positions in the room were averaged. The microphone was placed always at least 1 m from reflecting surfaces and 2 m from the other microphone positions.

The measurement was performed using the 01dB software for building acoustics, calculated in one-third octave bands. The result can be found in figure C.2 on the facing page.

The equipment used can be found in the following table C.2.

Equipment	Manufacturer	Type	Serial Number
Microphone	Brüel & Kjær	4134	1253168
Calibrator	Brüel & Kjær	4230	08155
Pre-Amplifier	Brüel & Kjær	2639	81474652
Omni-Directional Loudspeaker	Brüel & Kjær	4296	2251009
Symphonie	01dB	-	-
Power Amplifier	Pioneer	A-616	9405067s

Table C.2: *Equipment used for reverberation time measurement.*

C.2.1 Results

The ITU suggests an average reverberation time in one-third octave bands from 200 Hz to 4 kHz of

$$T_m \approx 0.25 \left(\frac{V}{V_0}\right)^{\frac{1}{3}}, \tag{C.1}$$

where V is the volume of the room and $V_0 = 100\text{m}^3$ a reference volume. Therefore the suggested average reverberation time for the used room is 0.24 s. In these frequencies an average reverberation time of 0.38 s was measured in the room.

Further the ITU suggests a tolerance mask around this average reverberation time. This tolerance mask and the measured reverberation time in one-third octave bands are pictured in figure C.2. As it can be seen, the measured values are always approx. 0.1 s above the suggested tolerance mask, which could influence the test in the sense of having results not for a reference listening room as described in ITU-R BS.1116, but an average listening room as in the IEC 268-13 standard. Since this thesis is investigating the influence of phase distortion under as realistic as possible conditions, these conditions are preferred for the experiments.

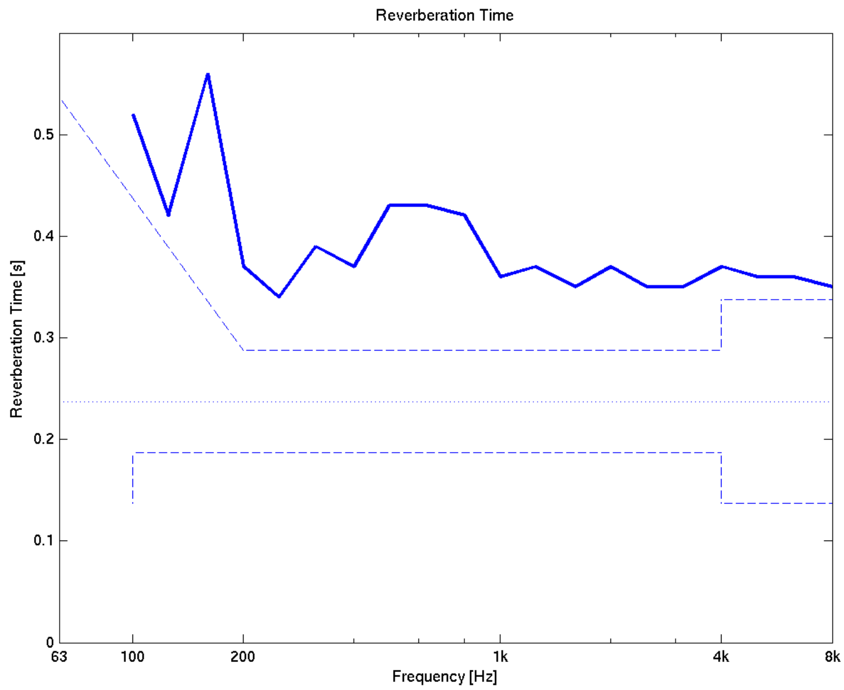

Figure C.2: *Reverberation time in the listening room and suggested reverberation time tolerance mask from ITU-R BS.1116.*

APPENDIX D

AUDIOMETRIES

An audiometry with the Madsen Electronics Orbiter 922 clinical audiometer was done for all subjects of the main test at the audiometry room of the acoustic labs of Aalborg University. The results are presented in the following graphs D.1 to D.19. Subject YMI (Figure D.19 on page 98) was rejected from the listening test because of her poor audiometry results.

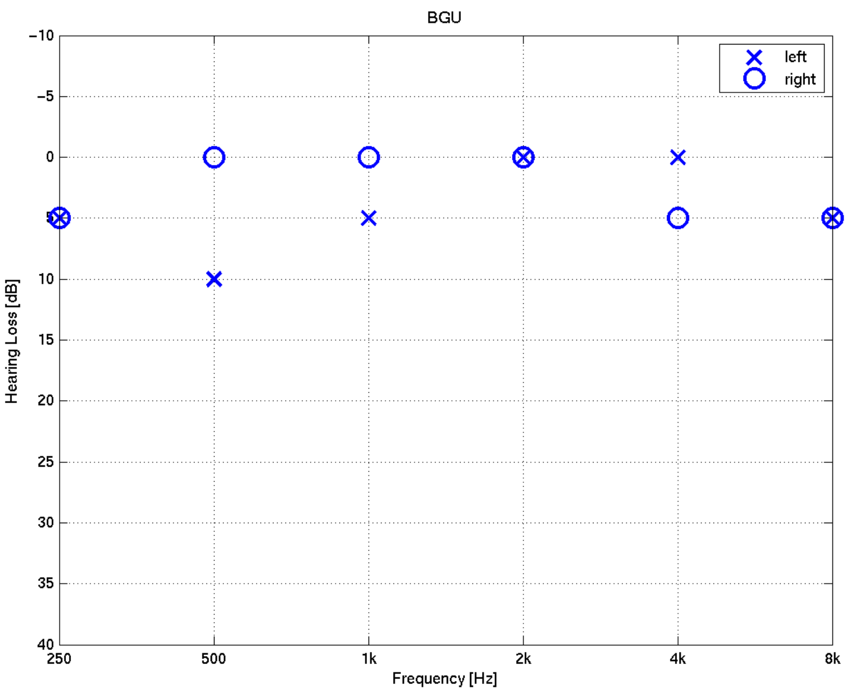

Figure D.1: *Audiometry of subject BGU.*

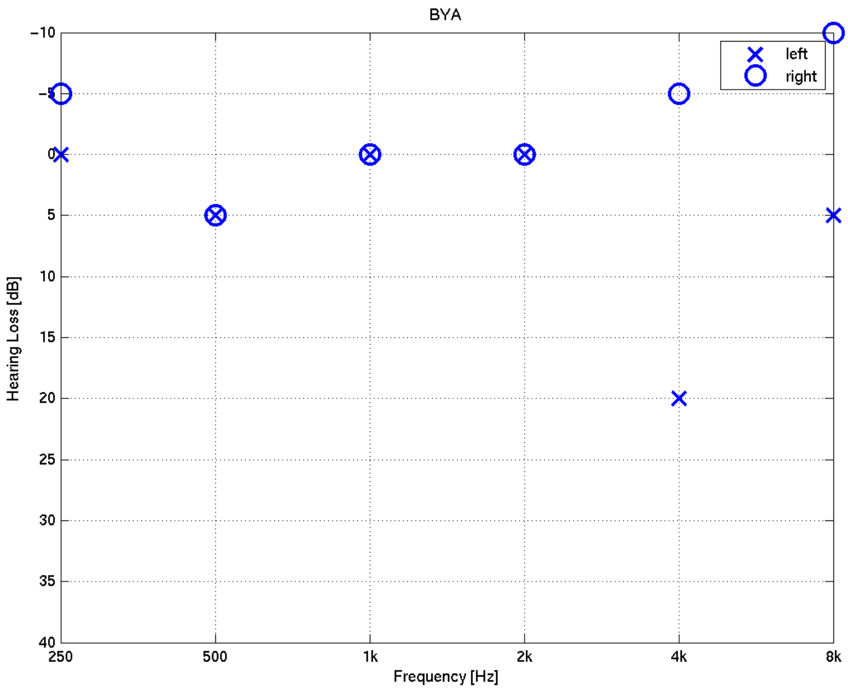

Figure D.2: *Audiometry of subject BYA.*

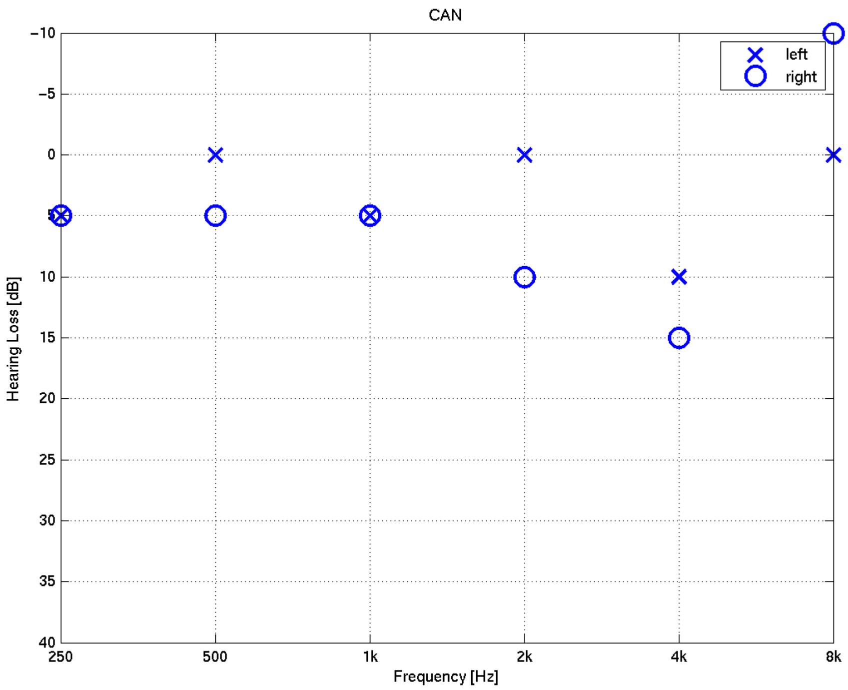

Figure D.3: *Audiometry of subject CAN.*

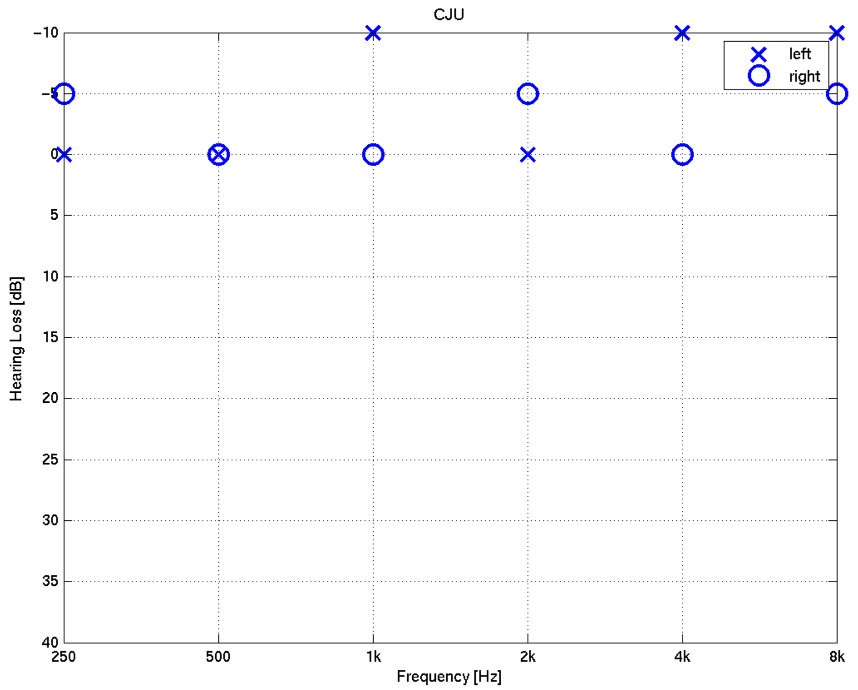

Figure D.4: *Audiometry of subject CJU.*

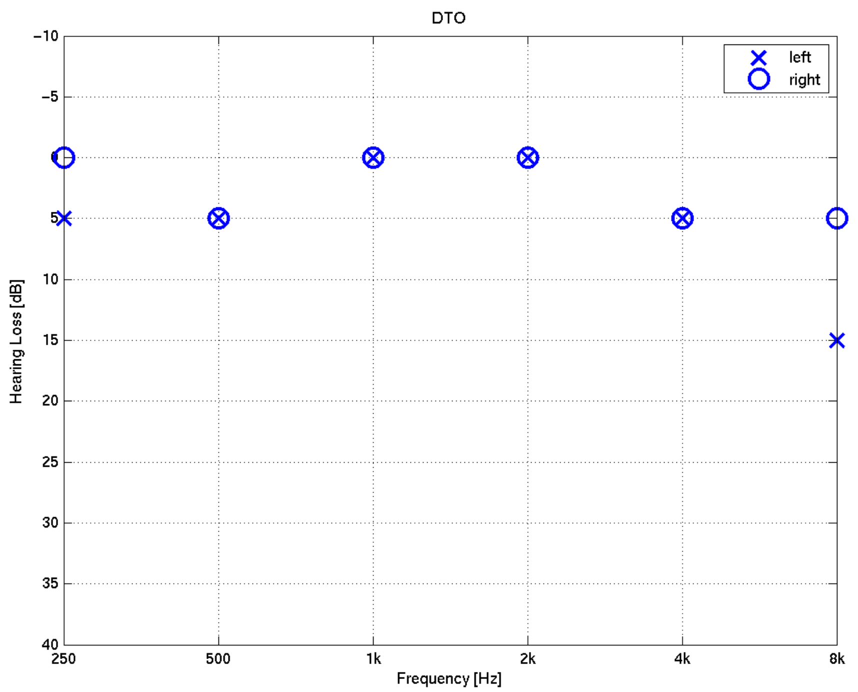

Figure D.5: *Audiometry of subject DTO.*

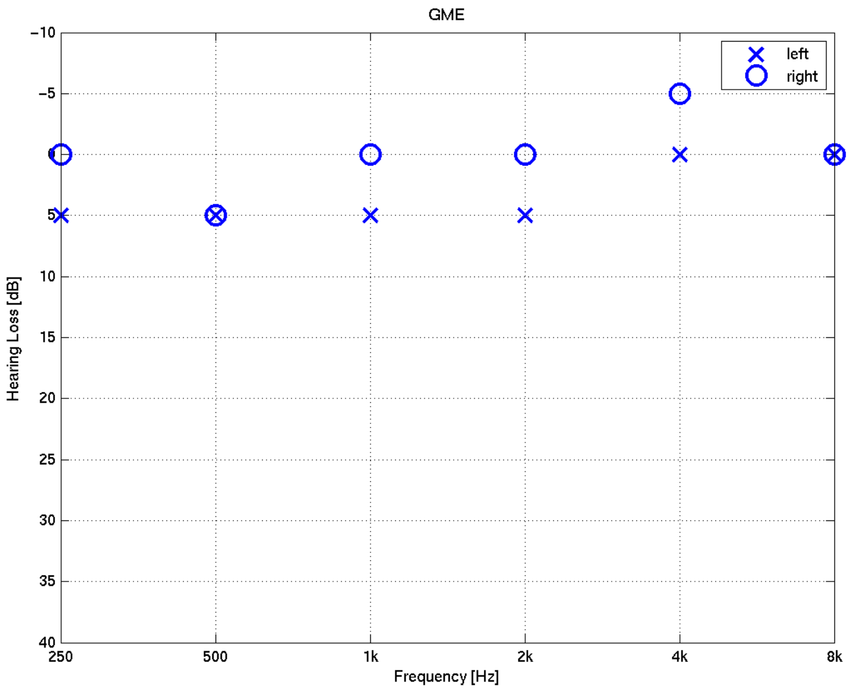

Figure D.6: *Audiometry of subject GME.*

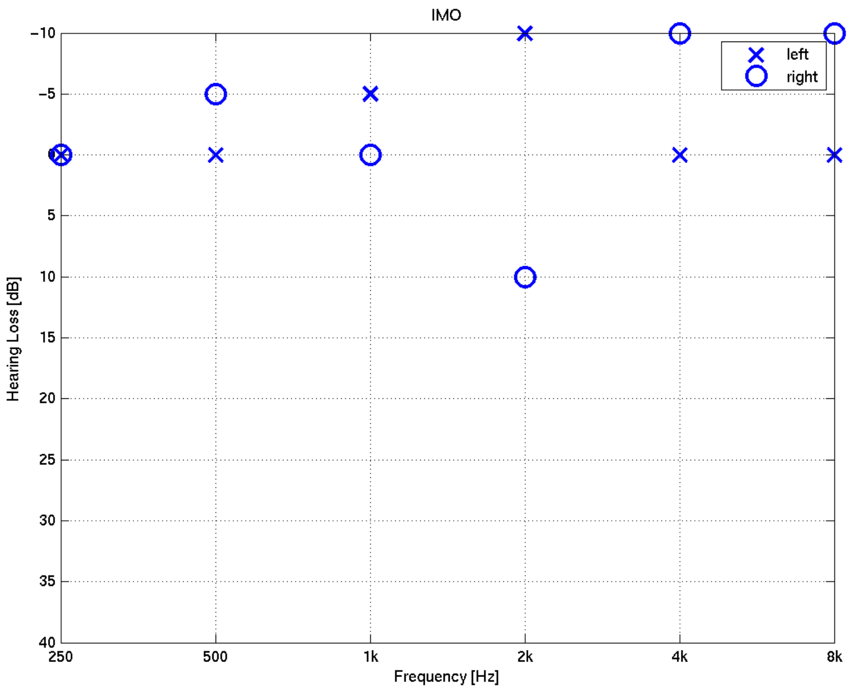

Figure D.7: *Audiometry of subject IMO.*

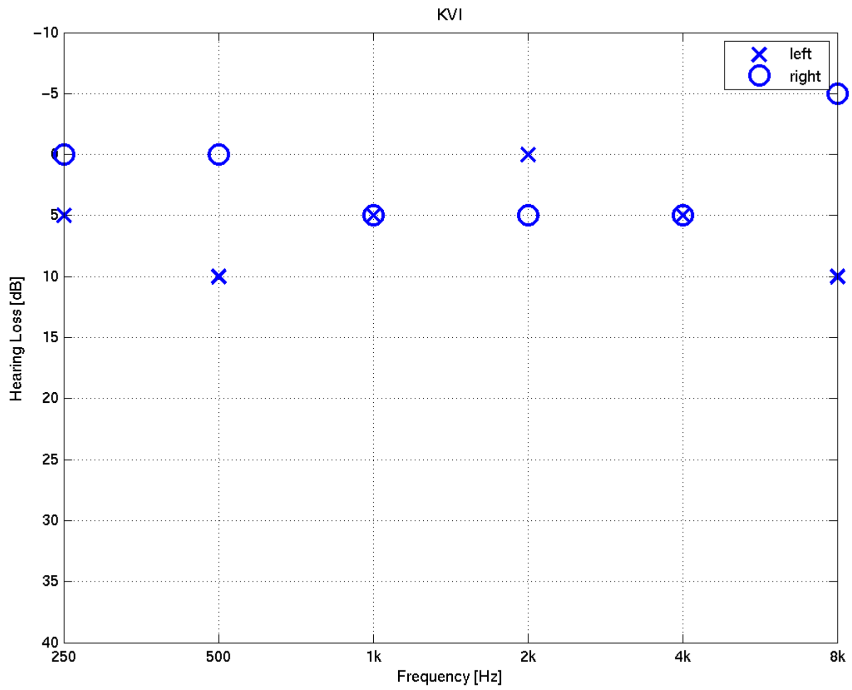

Figure D.8: *Audiometry of subject KVI.*

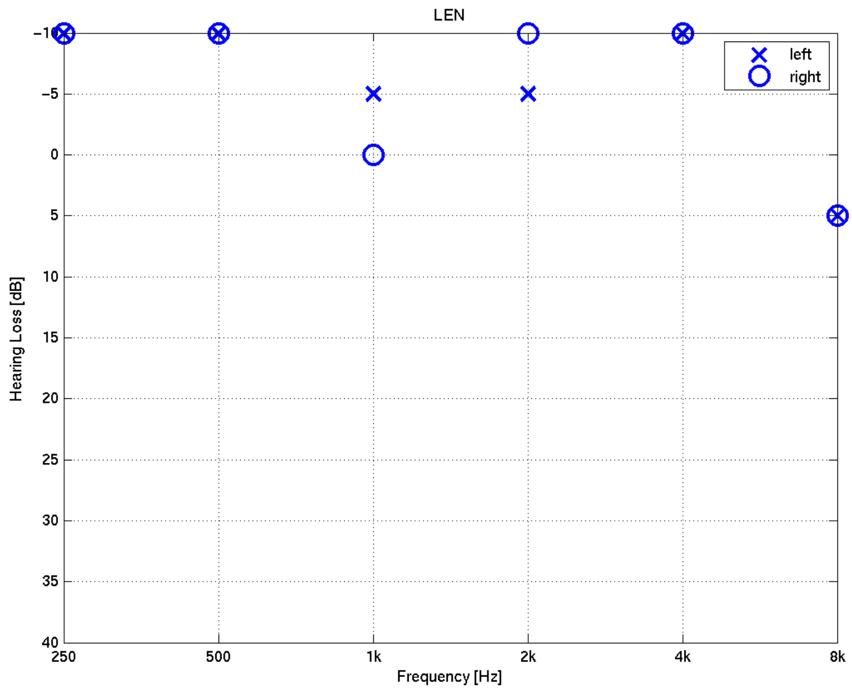

Figure D.9: *Audiometry of subject LEN.*

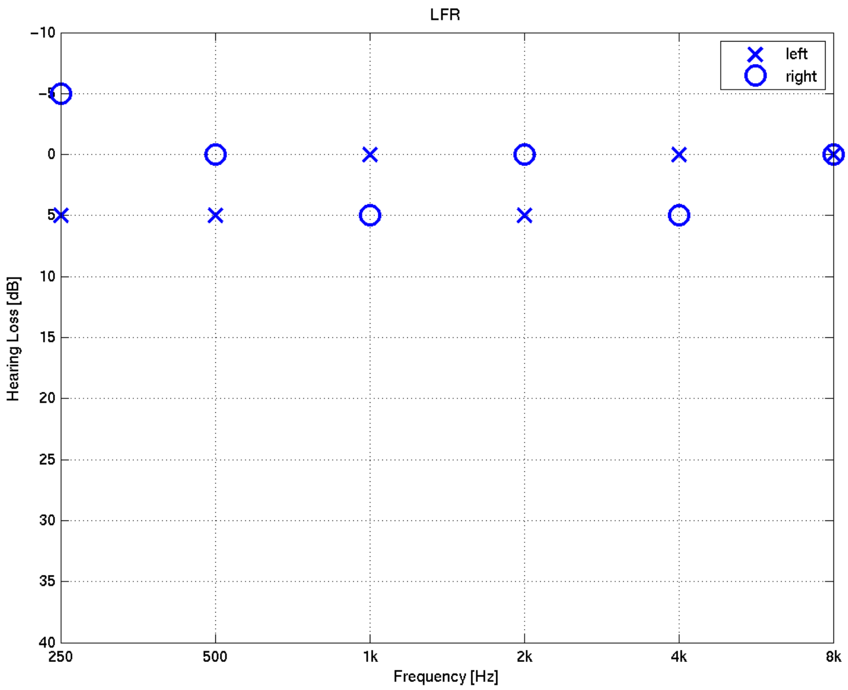

Figure D.10: *Audiometry of subject LFR.*

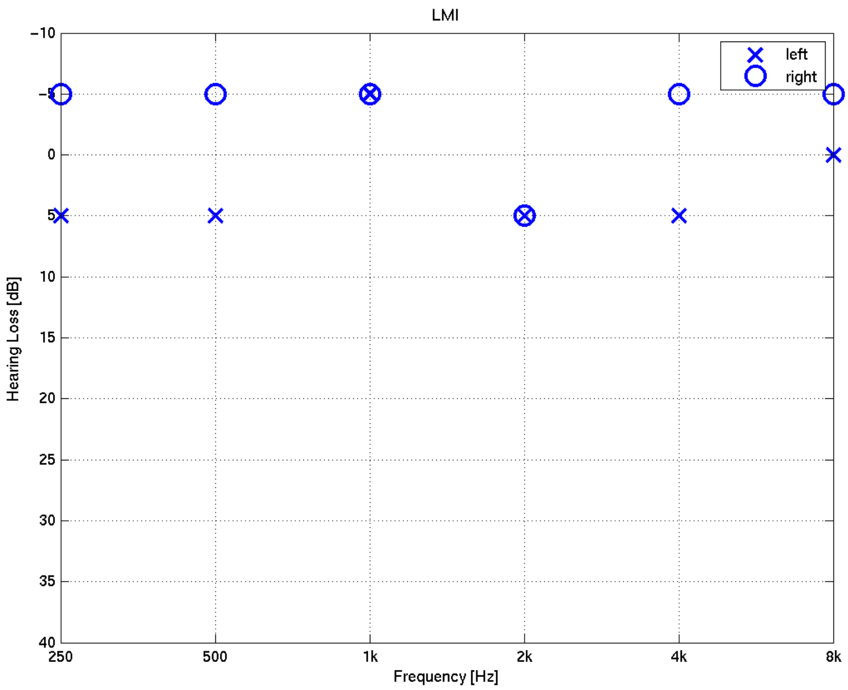

Figure D.11: *Audiometry of subject LMI.*

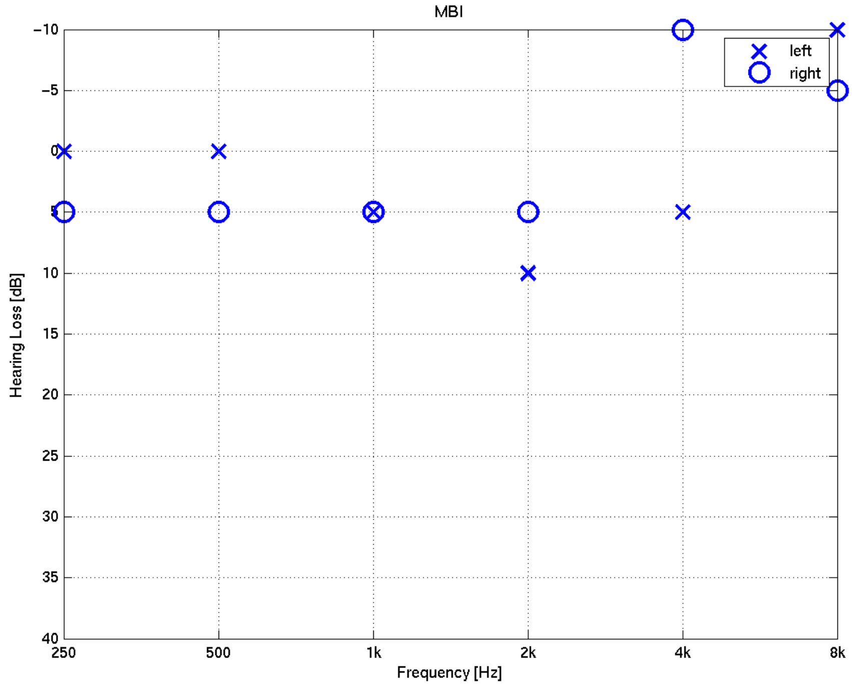

Figure D.12: *Audiometry of subject MBI.*

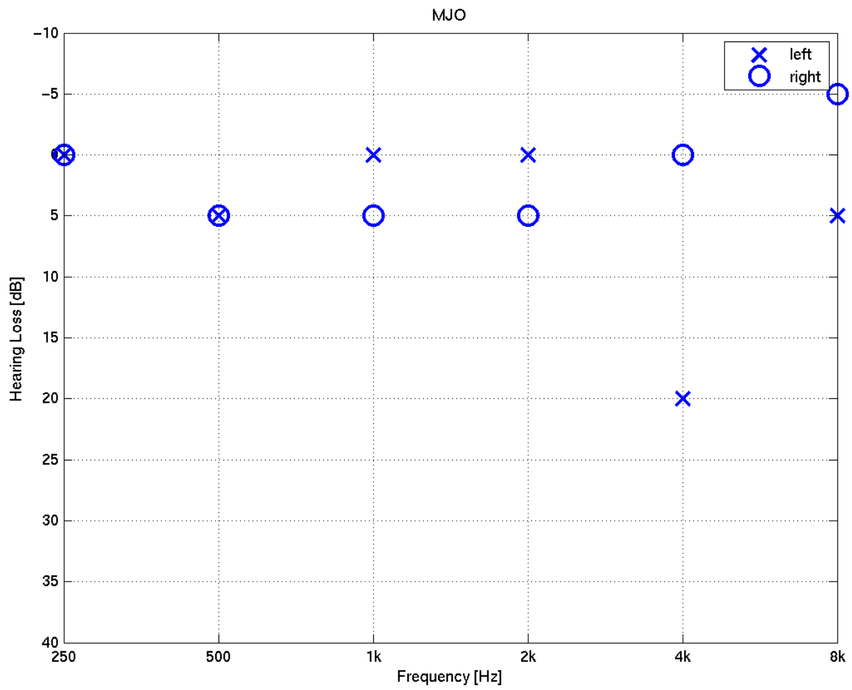

Figure D.13: *Audiometry of subject MJO.*

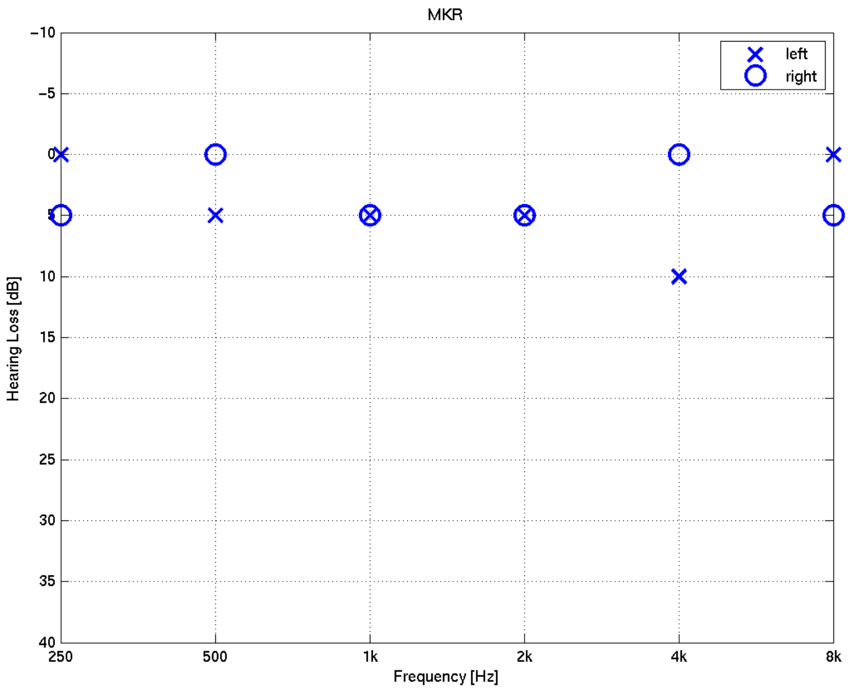

Figure D.14: *Audiometry of subject MKR.*

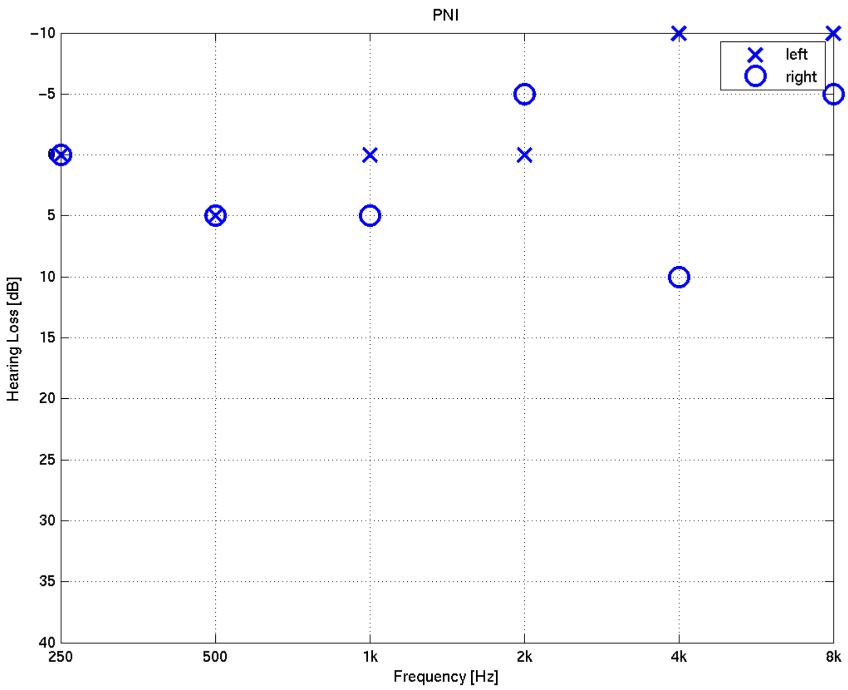

Figure D.15: *Audiometry of subject PNI.*

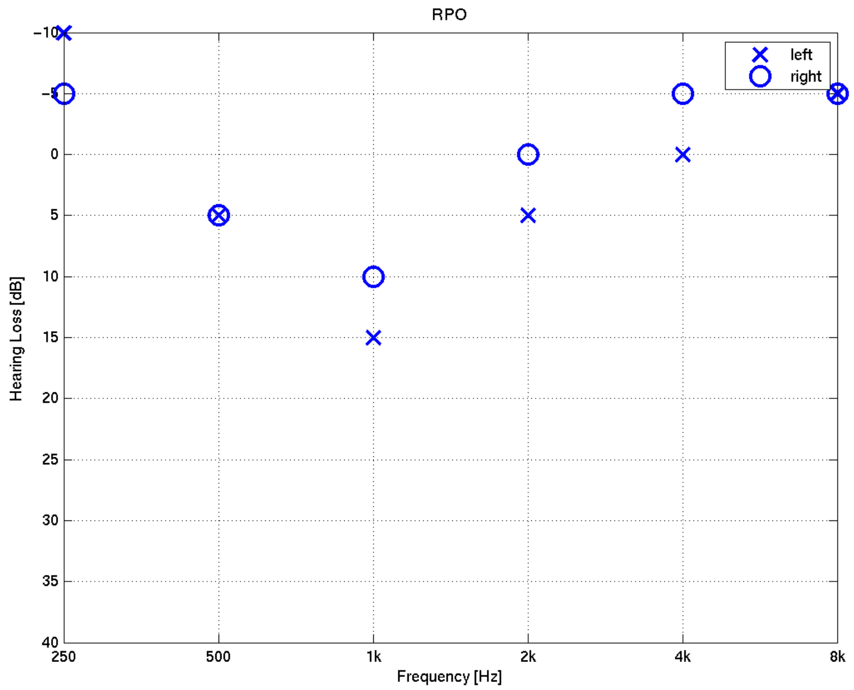

Figure D.16: *Audiometry of subject RPO.*

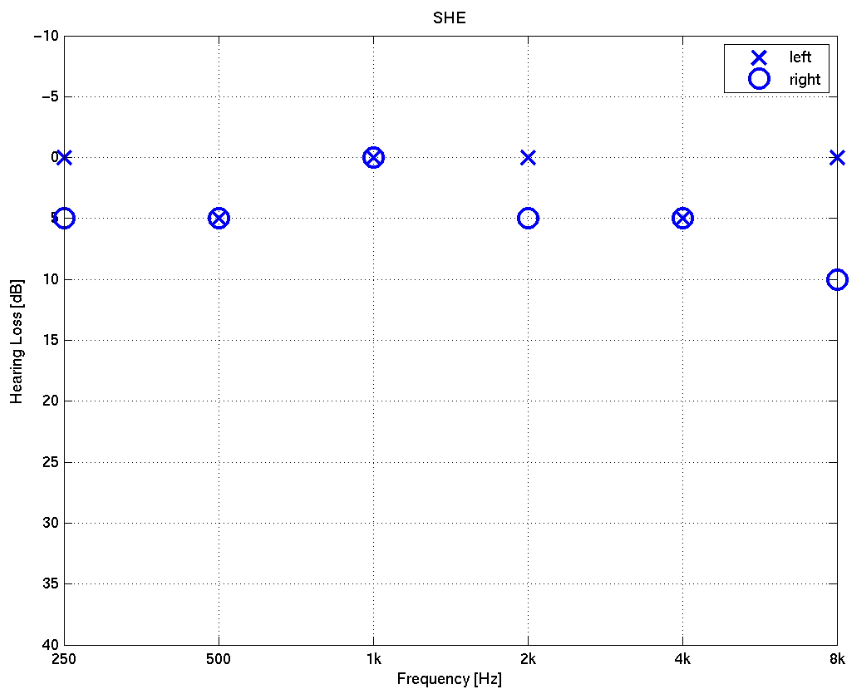

Figure D.17: *Audiometry of subject SHE.*

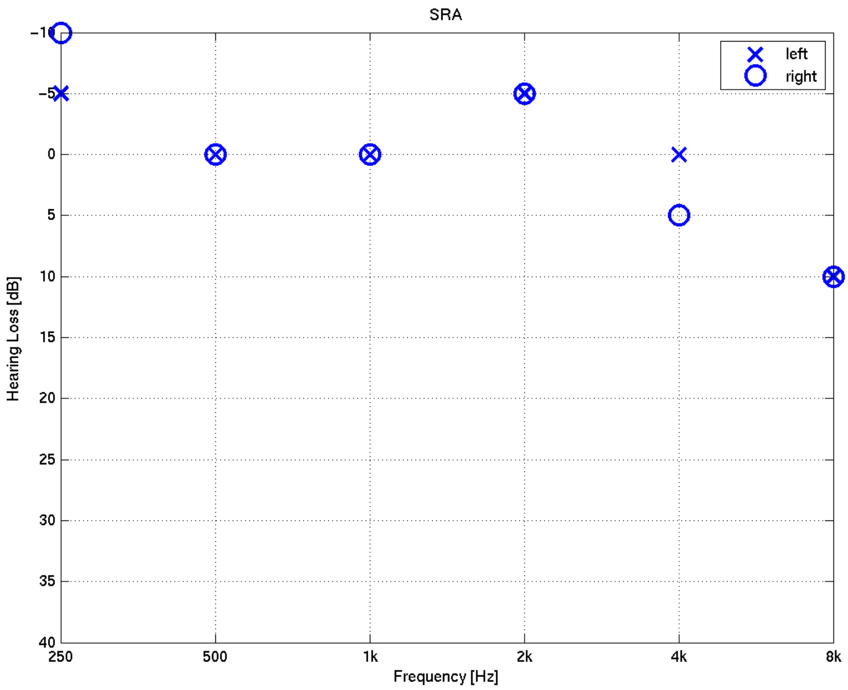

Figure D.18: *Audiometry of subject SRA.*

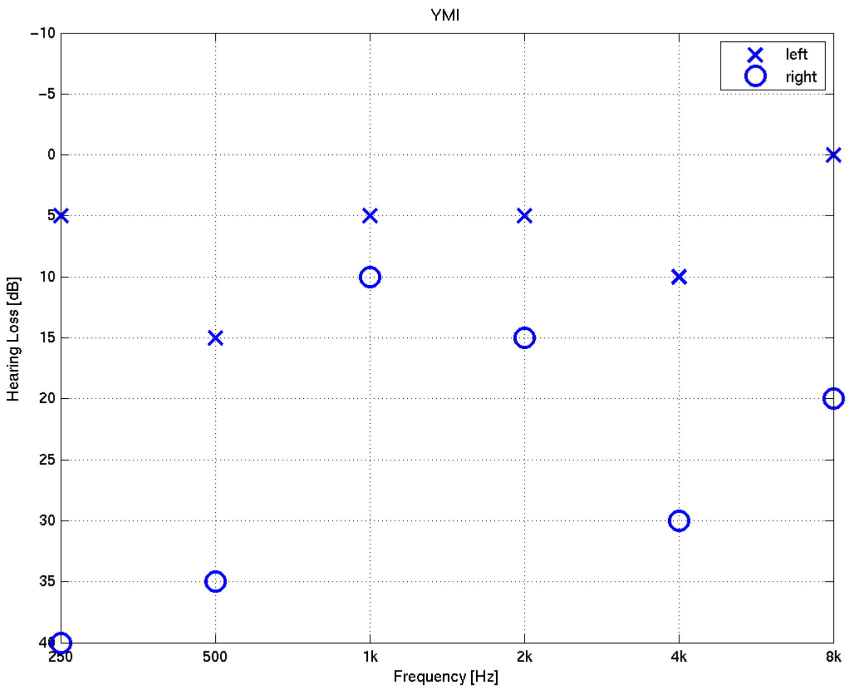

Figure D.19: *Audiometry of subject YMI.*

APPENDIX E

MAIN TEST QUESTIONNAIRE

The subjects were asked to fill out these questionnaire before performing the main test.

Listening Test Questionnaire

Name:

Age:

What is your occupation? If you are student, what are you studying?

Do you have any known hearing disorders? If yes, which?

Have you participated before in formal listening tests? If yes, which kind?

APPENDIX F

MAIN TEST RESULTS

This chapter presents the results obtained from the main test. The chapter is divided into three sections, one per signal. Each section presents the results obtained through headphones and loudspeakers. Additionally, the results were divided into groups as explained in section 5.2 on page 46 and a BTL analysis of each group was done. Detailed data obtained from the BTL model and percentage of inconsistencies are shown.

F.1 Results for Speech

The results for the speech signal were divided into two groups. The first group of 14 subjects evaluated the signal convolved with all-pass filters with Qs from 4 to 64. The second group of 18 subjects evaluated the signal down to a Q of 8. This section present the results for each group per transducer.

F.1.1 Headphones

Group 1: 14 subjects

	Original	64	32	16	8	4
Original	-	28 (27.9)	28 (27.9)	28 (27.6)	21 (21.7)	17 (13.5)
64	0 (0.08)	-	14 (12.8)	4 (4.70)	0 (0.27)	0 (0.07)
32	0 (0.09)	14 (15.1)	-	5 (5.37)	2 (0.32)	0 (0.08)
16	0 (0.39)	24 (23.2)	23 (22.6)	-	0 (1.32)	1 (0.36)
8	7 (6.23)	28 (27.7)	26 (27.6)	28 (26.6)	-	6 (5.91)
4	11 (14.4)	28 (27.9)	28 (27.9)	27 (27.6)	22 (22.0)	-

Table F.1: *Frequency table for results of the pair comparison test and predicted values in parentesis. Rows are preferred over columns.*

WST	MST	SST
0	0	5

Table F.2: *Transitivity violations.*

Goodness of fit: [$\chi^2(10) = 11.7105, p = 0.3049$]
Alternative goodness of fit: $\tilde{\chi}^2(10) = 14.4135$

Original	64	32	16	8	4
0.4204	0.0012	0.0014	0.0059	0.1205	0.4504

Table F.3: *Bradley-Terry-Luce scale values.*

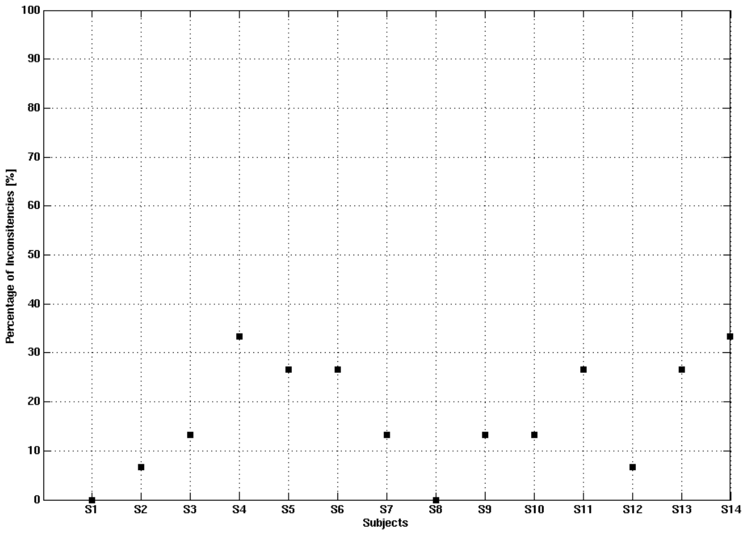

Figure F.1: *Percentage of inter-individual inconsistencies.*

Group 2: 18 subjects

	Original	64	32	16	8
Original	-	36 (35.8)	36 (35.8)	36 (35.5)	27 (27.7)
64	0 (0.10)	-	18 (17.6)	6 (5.95)	0 (0.33)
32	0 (0.10)	18 (18.3)	-	5 (6.16)	2 (0.34)
16	0 (0.49)	30 (30.0)	31 (29.8)	-	1 (1.61)
8	9 (8.29)	36 (35.6)	34 (35.6)	35 (34.3)	-

Table F.4: *Frequency table for results of the pair comparison test and predicted values in parentesis. Rows are preferred over columns.*

WST	MST	SST
0	0	3

Table F.5: *Transitivity violations.*

Goodness of fit: $[\chi^2(6) = 6.5245, p = 0.3671]$
Alternative goodness of fit: $\tilde{\chi}^2(6) = 9.6215$

Original	64	32	16	8
0.7580	0.0021	0.0022	0.0106	0.2269

Table F.6: *Bradley-Terry-Luce scale values.*

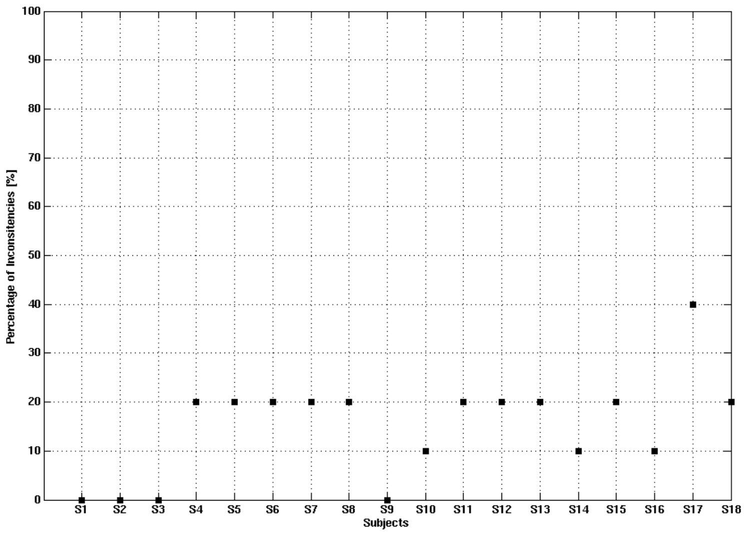

Figure F.2: *Percentage of inter-individual inconsistencies.*

F.1.2 Loudspeaker

Group 1: 14 subjects

	Original	64	32	16	8	4
Original	-	10 (14.8)	16 (15.8)	15 (15.8)	16 (13.3)	17 (14.1)
64	18 (13.1)	-	14 (15.0)	12 (15.0)	13 (12.4)	12 (13.3)
32	12 (12.1)	14 (12.9)	-	15 (13.9)	10 (11.5)	12 (12.3)
16	13 (12.1)	16 (12.9)	13 (14.0)	-	8 (11.5)	13 (12.3)
8	12 (14.6)	15 (15.5)	18 (16.4)	20 (16.4)	-	13 (14.8)
4	11 (13.8)	16 (14.6)	16 (15.6)	15 (15.6)	15 (13.1)	-

Table F.7: *Frequency table for results of the pair comparison test and predicted values in parentesis. Rows are preferred over columns.*

WST	MST	SST
13	4	9

Table F.8: *Transitivity violations.*

Goodness of fit: $[\chi^2(10) = 10.3350, p = 0.4116]$
Alternative goodness of fit: $\tilde{\chi}^2(10) = 10.2216$

Original	64	32	16	8	4
0.1817	0.1612	0.1396	0.1397	0.2001	0.1775

Table F.9: *Bradley-Terry-Luce scale values.*

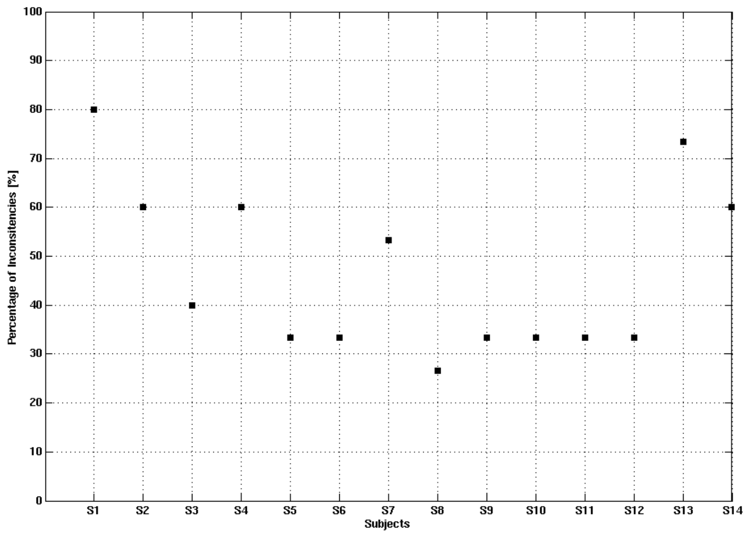

Figure F.3: *Percentage of inter-individual inconsistencies.*

Group 2: 18 subjects

	Original	64	32	16	8
Original	-	15 (18.8)	22 (20.8)	18 (19.4)	21 (16.9)
64	21 (17.1)	-	18 (20.0)	16 (18.6)	17 (16.1)
32	14 (15.1)	18 (15.9)	-	17 (16.5)	13 (14.2)
16	18 (16.5)	20 (17.3)	19 (19.4)	-	12 (15.5)
8	15 (19.0)	19 (19.8)	23 (21.7)	24 (20.4)	-

Table F.10: *Frequency table for results of the pair comparison test and predicted values in parentesis. Rows are preferred over columns.*

WST	MST	SST
9	4	5

Table F.11: *Transitivity violations.*

Goodness of fit: $[\chi^2(6) = 6.7647, p = 0.3432]$
Alternative goodness of fit: $\tilde{\chi}^2(6) = 6.7252$

Original	64	32	16	8
0.2164	0.1979	0.1581	0.1851	0.2422

Table F.12: *Bradley-Terry-Luce scale values.*

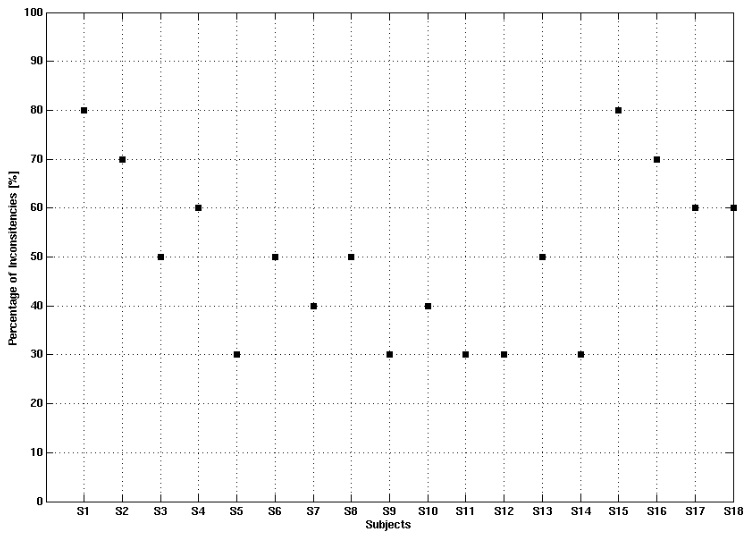

Figure F.4: *Percentage of inter-individual inconsistencies.*

F.2 Results for the Gong

The electronic gong results were divided into three groups: the first group of 5 subjects evaluated the original signal and the distorted with Qs from 8 to 64. The second group of 14 subjects evaluated the original signal and the distorted ones with Qs from 16 to 64. The third group of 18 subjects evaluated the original signal and the distorted with Qs equal to 32 and 64. The following section presents the results.

F.2.1 Headphones

Group 1: 5 subjects

	Original	64	32	16	8
Original	-	8 (9.19)	7 (8.17)	9 (6.90)	5 (4.72)
64	2 (0.80)	-	2 (2.82)	1 (1.63)	1 (0.73)
32	3 (1.82)	8 (7.17)	-	2 (3.32)	1 (1.67)
16	1 (3.09)	9 (8.36)	8 (6.67)	-	3 (2.86)
8	5 (5.27)	9 (9.26)	9 (8.32)	7 (7.13)	-

Table F.13: *Frequency table for results of the pair comparison test and predicted values in parentesis. Rows are preferred over columns.*

WST	MST	SST
0	2	3

Table F.14: *Transitivity violations.*

Goodness of fit: $[\chi^2(6) = 6.7845, p = 0.3412]$
Alternative goodness of fit: $\tilde{\chi}^2(6) = 6.7828$

Original	64	32	16	8
0.3479	0.0305	0.0777	0.1560	0.3877

Table F.15: *Bradley-Terry-Luce scale values.*

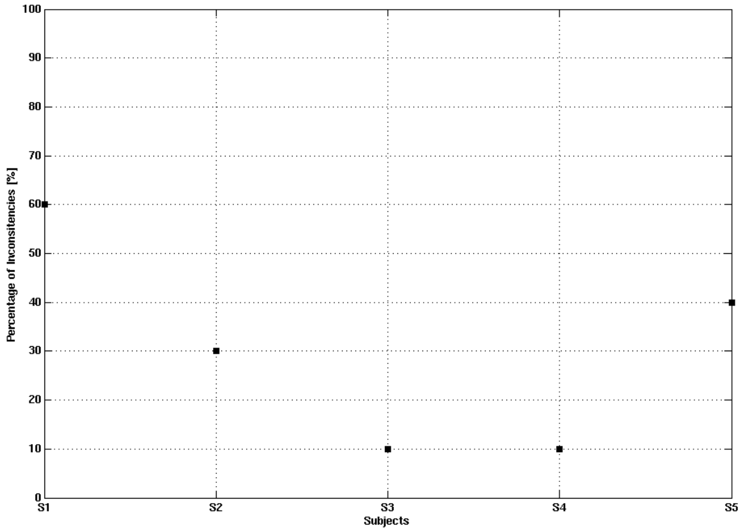

Figure F.5: *Percentage of inter-individual inconsistencies.*

Group 2: 14 subjects

	Original	64	32	16
Original	-	22 (24.0)	21 (21.0)	19 (16.9)
64	6 (3.98)	-	7 (9.35)	6 (5.66)
32	7 (6.95)	21 (18.6)	-	7 (9.40)
16	9 (11.0)	22 (22.3)	21 (18.5)	-

Table F.16: *Frequency table for results of the pair comparison test and predicted values in parentesis. Rows are preferred over columns.*

WST	MST	SST
0	0	0

Table F.17: *Transitivity violations.*

Goodness of fit: $[\chi^2(3) = 3.6458, p = 0.3023]$
Alternative goodness of fit: $\tilde{\chi}^2(3) = 3.6706$

Original	64	32	16
0.4651	0.0771	0.1536	0.3040

Table F.18: *Bradley-Terry-Luce scale values.*

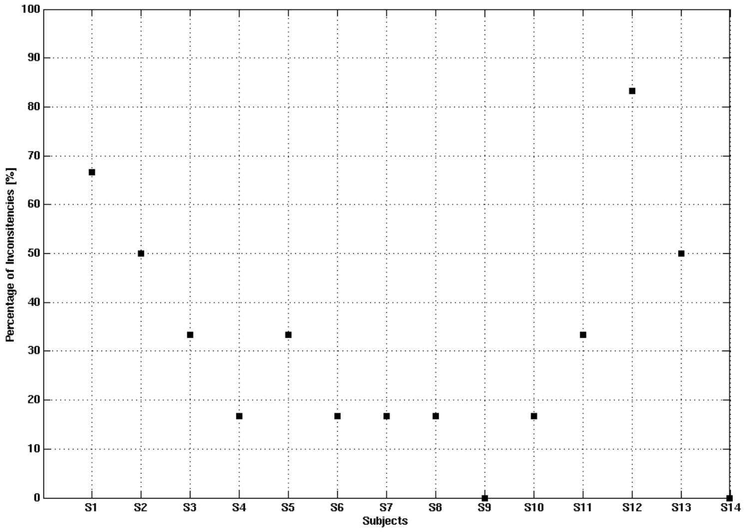

Figure F.6: *Percentage of inter-individual inconsistencies.*

Group 3: 18 subjects

	Original	64	32
Original	-	29 (30.9)	28 (26.0)
64	7 (5.09)	-	9 (10.9)
32	8 (9.90)	27 (25.0)	-

Table F.19: *Frequency table for results of the pair comparison test and predicted values in parentesis. Rows are preferred over columns.*

WST	MST	SST
0	0	0

Table F.20: *Transitivity violations.*

Goodness of fit: $[\chi^2(1) = 1.7785, p = 0.1823]$
Alternative goodness of fit: $\tilde{\chi}^2(1) = 1.8106$

Original	64	32
0.6474	0.1067	0.2457

Table F.21: *Bradley-Terry-Luce scale values.*

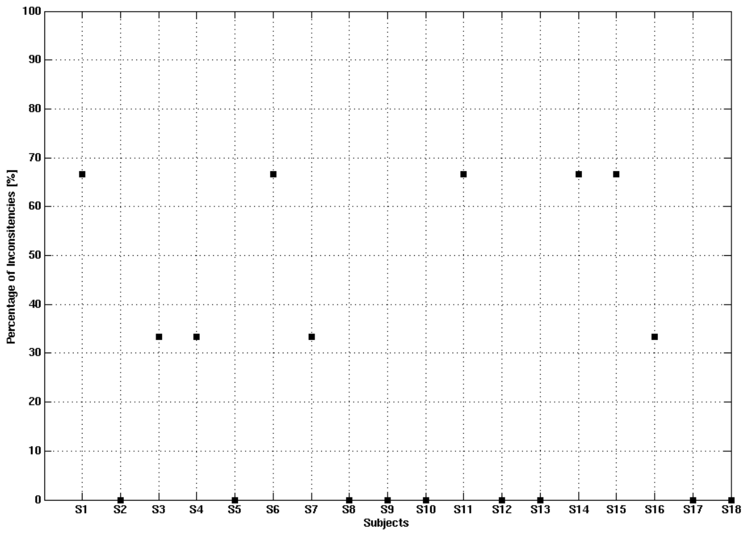

Figure F.7: *Percentage of inter-individual inconsistencies.*

F.2.2 Loudspeaker

Group 1: 5 subjects

	Original	64	32	16	8
Original	-	5 (4.79)	6 (5.40)	6 (5.40)	3 (4.39)
64	5 (5.20)	-	6 (5.60)	6 (5.60)	4 (4.59)
32	4 (4.59)	4 (4.39)	-	6 (5.00)	4 (4.00)
16	4 (4.59)	4 (4.39)	4 (4.99)	-	6 (4.00)
8	7 (5.60)	6 (5.40)	6 (5.99)	4 (5.99)	-

Table F.22: *Frequency table for results of the pair comparison test and predicted values in parentesis. Rows are preferred over columns.*

WST	MST	SST
9	0	2

Table F.23: *Transitivity violations.*

Goodness of fit: $[\chi^2(6) = 3.4207, p = 0.7545]$
Alternative goodness of fit: $\tilde{\chi}^2(6) = 3.427$

Original	64	32	16	8
0.1976	0.2141	0.1681	0.1681	0.2518

Table F.24: *Bradley-Terry-Luce scale values.*

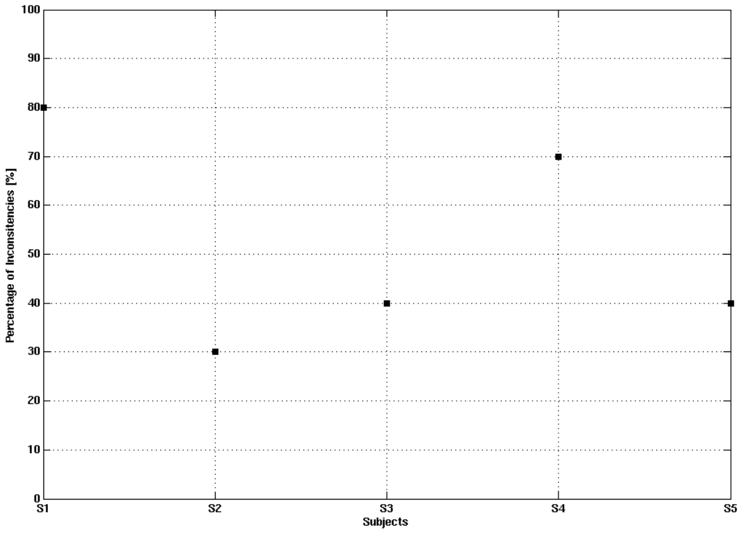

Figure F.8: *Percentage of inter-individual inconsistencies.*

Group 2: 14 subjects

	Original	64	32	16
Original	-	13 (13.9)	14 (13.9)	16 (15.0)
64	15 (14.0)	-	14 (13.9)	14 (15.0)
32	14 (14.0)	14 (14.0)	-	15 (15.0)
16	12 (12.9)	14 (12.9)	13 (12.9)	-

Table F.25: *Frequency table for results of the pair comparison test and predicted values in parentesis. Rows are preferred over columns.*

WST	MST	SST
4	1	5

Table F.26: *Transitivity violations.*

Goodness of fit: $[\chi^2(3) = 0.4304, p = 0.9339]$
Alternative goodness of fit: $\tilde{\chi}^2(3) = 0.43004$

Original	64	32	16
0.2586	0.2586	0.2586	0.2241

Table F.27: *Bradley-Terry-Luce scale values.*

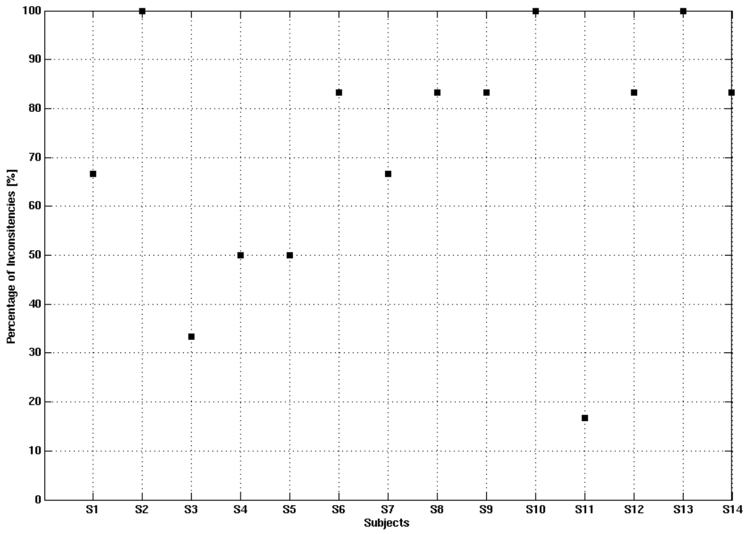

Figure F.9: *Percentage of inter-individual inconsistencies.*

Group 3: 18 subjects

	Original	64	32
Original	-	17 (17.9)	20 (18.9)
64	19 (18.0)	-	18 (19)
32	16 (17.0)	18 (17)	-

Table F.28: *Frequency table for results of the pair comparison test and predicted values in parentesis. Rows are preferred over columns.*

WST	MST	SST
2	1	0

Table F.29: *Transitivity violations.*

Goodness of fit: $[\chi^2(1) = 0.3342, p = 0.5632]$
Alternative goodness of fit: $\tilde{\chi}^2(1) = 0.33402$

Original	64	32
0.3454	0.3454	0.3091

Table F.30: Bradley-Terry-Luce scale values.

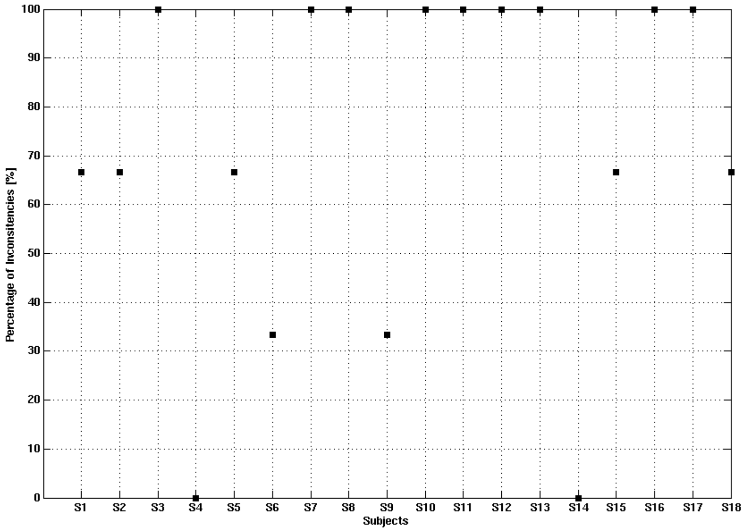

Figure F.10: Percentage of inter-individual inconsistencies.

F.3 Results for the Radiohead's Sample

Only ten subjects evaluated this signal. Four out of the ten assessed the signal distorted down to a Q equal to 16 and the rest only evaluated the signal distorted with Qs equal to 32 and 64. This section describes the results obtained per each group.

F.3.1 Headphones

Group 1: 4 subjects

	Original	64	32	16
Original	-	4 (3.74)	4 (4.76)	4 (3.49)
64	4 (4.25)	-	6 (5.00)	3 (3.74)
32	4 (3.23)	2 (2.99)	-	3 (2.76)
16	4 (4.50)	5 (4.25)	5 (5.23)	-

Table F.31: Frequency table for results of the pair comparison test and predicted values in parentesis. Rows are preferred over columns.

WST	MST	SST
3	0	7

Table F.32: *Transitivity violations.*

Goodness of fit: $[\chi^2(3) = 1.3332, p = 0.7213]$
Alternative goodness of fit: $\tilde{\chi}^2(3) = 1.3038$

Original	64	32	16
0.2435	0.2764	0.1657	0.3142

Table F.33: *Bradley-Terry-Luce scale values.*

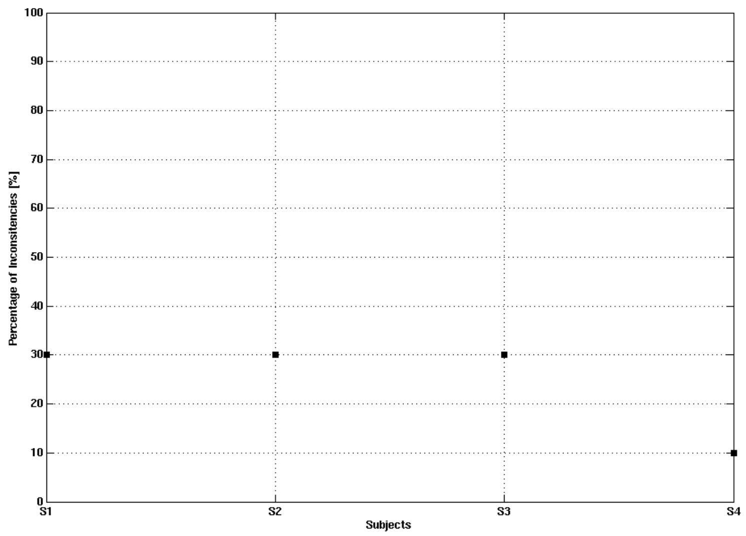

Figure F.11: *Percentage of inter-individual inconsistencies.*

Group 2: 10 subjects

	Original	64	32
Original	-	8 (9.31)	14 (12.6)
64	12 (10.6)	-	12 (13.3)
32	6 (7.31)	8 (6.68)	-

Table F.34: *Frequency table for results of the pair comparison test and predicted values in parentesis. Rows are preferred over columns.*

WST	MST	SST
0	0	1

Table F.35: *Transitivity violations.*

Goodness of fit: $[\chi^2(1) = 1.1086, p = 0.2924]$
Alternative goodness of fit: $\tilde{\chi}^2(1) = 1.1053$

Original	64	32
0.3671	0.4212	0.2116

Table F.36: *Bradley-Terry-Luce scale values.*

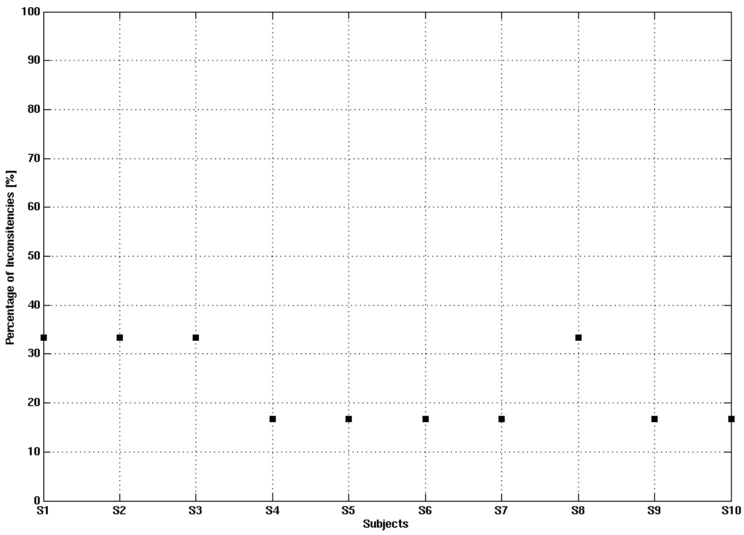

Figure F.12: *Percentage of inter-individual inconsistencies.*

F.3.2 Loudspeaker

Group 1: 4 subjects

	Original	64	32	16
Original	-	4 (5.71)	5 (4.76)	6 (4.51)
64	4 (2.28)	-	2 (2.97)	2 (2.73)
32	3 (3.23)	6 (5.02)	-	3 (3.74)
16	2 (3.48)	6 (5.26)	5 (4.25)	-

Table F.37: *Frequency table for results of the pair comparison test and predicted values in parentesis. Rows are preferred over columns.*

WST	MST	SST
4	2	1

Table F.38: *Transitivity violations.*

Goodness of fit: $[\chi^2(3) = 3.9678, p = 0.2650]$
Alternative goodness of fit: $\tilde{\chi}^2(3) = 4.0224$

Original	64	32	16
0.3511	0.1406	0.2378	0.2704

Table F.39: *Bradley-Terry-Luce scale values.*

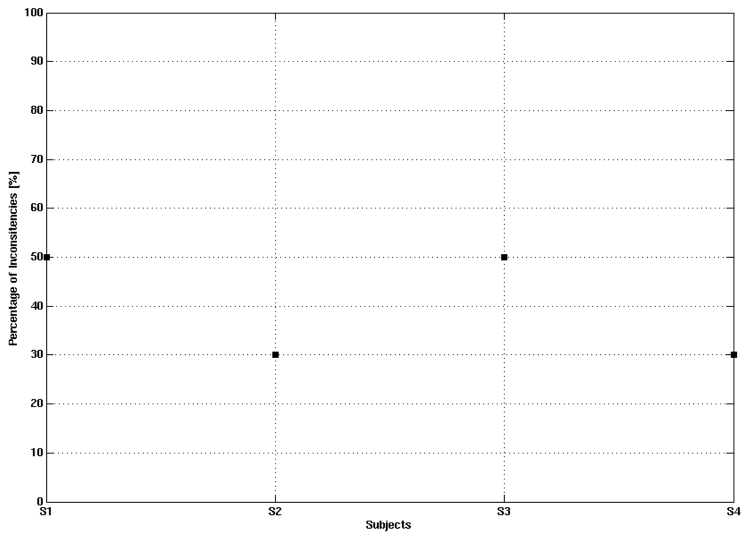

Figure F.13: *Percentage of inter-individual inconsistencies.*

Group 2: 10 subjects

	Original	64	32
Original	-	11 (10.6)	11 (11.3)
64	9 (9.33)	-	11 (10.6)
32	9 (8.66)	9 (9.33)	-

Table F.40: *Frequency table for results of the pair comparison test and predicted values in parentesis. Rows are preferred over columns.*

WST	MST	SST
0	0	0

Table F.41: *Transitivity violations.*

Goodness of fit: $[\chi^2(1) = 0.0665, p = 0.7965]$
Alternative goodness of fit: $\tilde{\chi}^2(1) = 0.066469$

Original	64	32
0.3788	0.3313	0.2898

Table F.42: *Bradley-Terry-Luce scale values.*

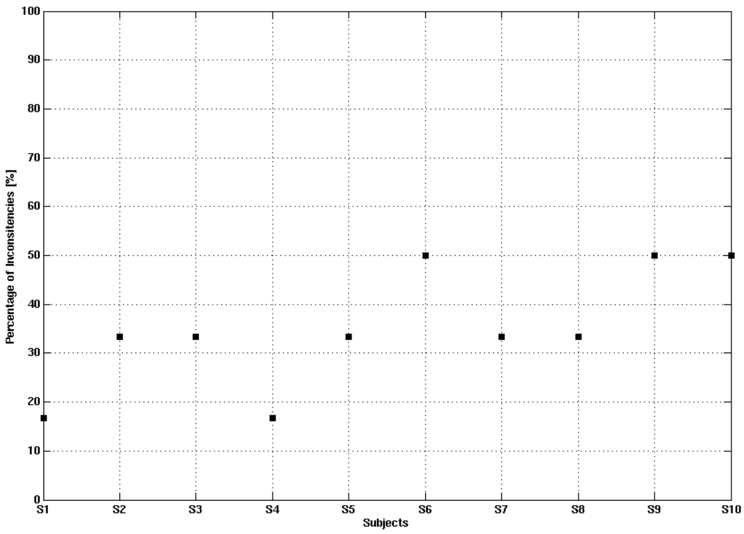

Figure F.14: *Percentage of inter-individual inconsistencies.*