HRTFs in Human Localization: Measurement, Spectral Evaluation and Practical Use in Virtual Audio Environment

Von der Fakultät für Maschinenbau, Elektrotechnik und Wirtschaftsingenieurwesen der Brandenburgischen Technischen Universität Cottbus zur Erlangung des akademischen Grades eines Doktor-Ingenieurs genehmigte Dissertation

vorgelegt von

Diplom-Ingenieur

WERSÉNYI, György

geboren am 17. Januar, 1975 in Győr, Ungarn.

Vorsitzender: Prof. Dr.-Ing. B. Falter, BTU Cottbus
Gutachter: Prof. Dr.-Ing. K.-R. Fellbaum, BTU Cottbus
Gutachter: Prof. Dr.-Ing. R. Hoffman, TU Dresden
Gutachter: Dr.-Ing. A. Illényi, TU Budapest
Tag der mündlichen Prüfung: 09.07.2002.

HRTFs in Human Localization: Measurement, Spectral Evaluation and Practical Use in Virtual Audio Environment

Wersényi György

Doctoral Thesis

Brandenburg University of Technology, Cottbus, Germany
University of Technology and Economics, Budapest, Hungary

2002

HRTFs in Human Localization: Measurement, Spectral Evaluation and Practical Use in Virtual Audio Environment

Doctoral Thesis

ABSTRACT

There are still many open questions in the field of spatial and directional hearing studies. The human hearing system works in very complex and individual ways. Many recent works have focused on hearing modeling from the outer ears to the central nervous system. The traditional models of a hierarchical evaluation have been revised and expanded. Nowadays, the results in the physiology of hearing suggest the relevant influence of higher processing in the nervous system and the brain.

The directional information is encoded inside the physical sound waves and it is decoded and evaluated by the hearing system both in the time-domain and frequency-domain from the signals reaching the eardrums. Due to its spectral filtering the outer ears and the shape of the human body play a significant role.

This work analyses the transmission of the directional information and contributes to the suggestions of the physiological hypotheses. It will be proved that higher processing algorithms in the brain do have significant effects and evaluation mechanisms even at the level of the outer ears.

We focus on the role and effect of the transfer function of the outer ear. The microphones are placed at the eardrums of a dummy-head, and accurate measurements will be performed.

The transmission from a point in the free-field to the eardrums is described by the complex Head-Related Transfer Functions (HRTFs) or their time-domain variant:

Head-Related Impulse Responses (HRIRs). In everyday life environments humans use their individual HRTFs for the localization, but in virtual audio environments the HRTFs have to be reproduced “exactly” through headphones. We will demonstrate that little modifications of the acoustical environment near to the listeners’ head do have significant effects on the transfer functions.

This indicates and supports the hypothesis that the brain has significant influence at the level of the external ears and that virtual audio simulation does not fail on the artificial HRTF filtering, accuracy of the transfer functions or signal processing algorithms.

First, a good quality binaural playback system is tested in a listening test using equalized headphones and HRTF filtering in order to find well-known errors and the localization blur it can be achieved in a virtual audio simulation. Our results are comparable with former results obtained by others, supporting the fact that virtual audio synthesis is inferior to loudspeaker playback. The second part introduces an accurate dummy-head measurement system with increased precision and new methods to increase the signal-to-noise (SNR) ratio and reproducibility. Then, finally, the HRTFs in the frequency domain and the effect of the acoustical environment near to the head will be evaluated. With the help of simple mathematical terms and definitions we will introduce a novel 2D representation of the HRTFs (so-called polar histograms), the monaural and binaural sensitivity domains, the effect of a moving sound source in the horizontal plane and the statistical (averaged) effects of the “everyday objects” near to the head. In the discussion section we try to specify the quality levels of existing binaural systems and the role of the HRTFs in virtual and non-virtual environments during the decoding of the acoustical information.

HRTFs in der menschlichen Lokalisation: Messungen, spektrale Auswertung und praktische Anwendungen in virtueller akustischer Umgebung

Doktorarbeit

AUSZUG

Auf dem Gebiet des räumlichen Hörens sind immer noch Fragen zu beantworten. Das menschliche Gehör arbeitet sehr komplex und individuell. Heutzutage stehen die binaurale Technik und die Gehörmodelle vom Außenohr bis zum zentralen Nervensystem wieder im Vordergrund. Die traditionellen Modelle des hierarchischen Aufbaus der Auswertung von akustischen Informationen sind revidiert und ergänzt worden. Die Ergebnisse der Hörphysiologie deuten einen relevanten Einfluss der Bearbeitung im Gehirn schon bei der Lokalisation an.

Die Richtungsinformation ist in den Schallwellen kodiert und wird vom Gehör aus dem Ohrsignal am Trommelfell, sowohl im Zeitbereich als auch im Frequenzbereich dekodiert. Die Ohrmuscheln, der Kopf und der Körper spielen dabei eine wichtige Filterrolle.

Die Übertragung von einem Punkt im Freifeld bis zum Trommelfell ist durch die komplexen Head-Related Transfer Functions (HRTFs) definiert. Im alltäglichen Leben nutzt der Mensch seine eigenen individuellen HRTFs während der Lokalisation. In virtuellen Umgebungen müssen diese Übertragungsfunktionen elektronisch durch einen Kopfhörer „exakt“ nachgebildet werden.

Diese Doktorarbeit trägt einiges zu den Hypothesen der Hörphysiologie bei. Es wird gezeigt, daß höhere Bearbeitungsstufen des Gehirns eine wichtige Rolle und einen großen Einfluss schon an der Ebene der Außenohren haben.

Wir konzentrieren uns auf die Übertragungsfunktionen und deren Feinstruktur. Mikrofone sind am Trommelfell eines Kunstkopfes fixiert und präzise Messungen werden durchgeführt. Wir demonstrieren, dass kleine Änderungen in der akustischen Umgebung in der Nähe des Kopfes einen signifikanten Einfluss auf die HRTFs haben.

Vor der Messung werden ein binaurales Abspielsystem und die Anwendungsmöglichkeiten präsentiert. Die Ergebnisse sind im Vergleich zu früheren Resultaten dargestellt. Dies bestätigt die Tatsache, dass virtuelle Simulationen immer noch eine schlechtere Qualität haben als die Lautsprecherwiedergabe.

Der zweite Teil stellt eine Messeinrichtung für präzise, automatische Kunstkopfaufnahmen mit erhöhtem Signal-Rausch Abstand vor.

Schließlich werden die gemessenen HRTFs und die Umgebungseffekte in der Nähe des Kopfes analysiert. Mit einfachen mathematischen Definitionen zeigen wir, wie alltägliche Objekte (Haare, Brille usw.) auf die HRTFs wirken, ohne dabei die Lokalisation und die Aufarbeitung der akustischen Information zu verändern. Es wird eine neuartige 2D-Darstellung gezeigt, in der HRTFs als Funktion der Frequenz und des Azimuts gleichzeitig abgebildet werden können. Es werden auch typische Grenzfrequenzen vorgestellt, die in der Auswertung von räumlichen Informationen eine wichtige Rolle spielen. Dabei wurde der Effekt des Kopfschattens näher analysiert.

All das führt uns zu der Bestätigung des Einfluss von höheren Mechanismen im Gehirn. Simulierte Schallfelder scheitern nicht an der Qualität der angewendeten HRTFs oder der Signalbearbeitung, sondern vielmehr an der Qualität der Kopfhörer und der nicht simulierten Schallfeldkomponenten (Reflexionen, Kopfbewegungen usw.) Aussagen der binauralen Technik stehen dabei im Mittelpunkt, denn nicht nur der Schalldruckpegel und die Ohrsignale am Trommelfell sind bei der Lokalisation von Bedeutung.

Contents

1 Introduction 9
2 Background and overview 16
2.1 Physiology of hearing 16
2.2 Psychoacoustic localization cues 19
2.2.1 Localization in free-field and virtual environments 22
2.3 The Head-Related Transfer Functions 25
2.3.1 Measurement of HRTFs 28
2.3.2 Dummy-heads and modeling in the measurement technique 31
2.3.3 Measurement signals and signal-to-noise ratio 31
2.4 Experimental results 33
2.4.1 The binaural technique 33
2.4.2 Subjective listening tests in virtual audio synthesis 33
2.4.3 Localization results of listening tests 34
2.4.4 Headphone playback errors 39
2.4.5 Quality of HRTFs in the virtual audio simulation 40
2.4.6 Virtual Acoustic Displays 42
2.5 Problems 45
2.6 Goals 46
3 HRTFs in listening tests: localization blur in a 2D Virtual Audio system 47
3.1 Introduction 47
3.2 Measurement method 48
3.3 Results 53
3.3.1 Capability and errors in headphone playback 53
3.3.2 Localisation blur and discrimination skills 56
3.3.3 Localization judgements 64
3.4 Summary 66
4 Measurement of dummy-head HRTFs 69
4.1 Introduction 69
4.2 The measurement setup 70
4.2.1 General parameters 70
4.2.2 Setting of the elevation 73
4.2.3 Setting of the azimuth 75
4.2.4 The pseudo-random noise excitation 76
4.2.5 Effect of the high voltage periodicity 82
4.2.6 Environmental reflections and room impulse response 84

4.3 Testing....86
4.4 Summary....91
5 Evaluation of differences in dummy-head HRTFs caused by the acoustical environment near to the head....93
5.1 Introduction....93
5.2 Terms of use....94
5.3 Evaluation of dummy-head HRTFs in the horizontal plane based on deviations in the peak-valley structure of the bare torso....96
5.4 Effect of the acoustical environment near to the head....106
5.4.1 Hair....119
5.4.2 Glasses....119
5.4.3 Baseball cap....119
5.4.4 Clothing....120
6 Discussion....126
6.1.1 The sensitivity domains....126
6.1.2 Frequency limits in the lateral-contralateral evaluation....127
6.2 Binaural evaluation....129
6.2.1 Localisation performance in binaural playback systems....130
7 Results....134
8 Conclusions....138
8.1 Future works and application notes....140
9 References....143
10 German Abstract....155
11 Appendix A References alphabetical order....164
12 Appendix B Localization results....170
13 Appendix C Photos....174
14 Appendix E Abbreviations....177
15 Appendix F Detailed results of the listening test....178

1 Introduction

„Hearing research at the moment is a complicated interaction between physics, anatomy, physiology, and psychology. We cannot separate certain variables to the degree that is possible in physics. Furthermore, our measurements are not so precise, and the range of validity is not so well defined. Therefore, we often have to modify our earlier findings in light of the new, at least in the range of validity.

If we have a speaker in a normal living room and we listen to him monaurally or binaurally first from a distance of one meter and then from three meters, *we notice hardly any difference* except for a small drop in loudness at the greater distance. But if we have two identical microphones, one placed one meter away from the speaker and the second three meters away, then *the recordings show two different sound pressure patterns over time*. There is a small time delay for the lower trace, which was recorded from the more distant microphone. *It is difficult to understand how such different stimuli as the sound patterns in the upper and lower traces can product the same sensations. Much research has to be done to find the reasons why this is possible.*" [1].

The Hungarian Nobel-prize winner *Gy. Békésy* described this phenomenon in the early sixties [2, 3]. He realized that the sensation, the auditory image in the auditory system, does not depend strongly on the electro-acoustical transmission. Sometimes, a person is able to get a correct sound-field image in the brain and extract the acoustical information properly even in a disturbed environment. On the other hand, we could generate or transmit almost the same signal to the eardrums, but the transmission of directional and spatial information may fail.

There are still many questions open in the field of the spatial and directional hearing. The human hearing system works in very complex, individual and non-linear ways. Recently, many works focus on hearing modeling from the outer ears up to the central nervous system. The traditional models of a hierarchical (sequential) evaluation are not valid any longer, thus, it is still in our interest to expand these models. The former auditory models cited the outer ears as the one and only responsible part for the localization. Nowadays, the new results in the physiology of hearing suggest stronger influence of higher processing in the nervous system and the brain (cortex).

Although, the *acoustical information* can be seen as more complex, in this work it is defined only as the pure directional information. This kind of *directional*

information is encoded inside the physical sound waves and it is decoded and evaluated by the hearing system both in the time-domain and frequency-domain from the signals reaching the eardrums. The outer ear and the shape of the human body play the most significant role due to its spectral filtering and production of the interaural differences. Most of these phenomena can be easily modeled, handled physically and measured exactly.

Fig.1. shows information elements of the acoustical information: localization means pure directional information and source distance. Other parameters like size, type of the sound source and the environment, quality of the signals and the transmission path etc. are also transmitted and subjectively evaluated by the auditory system. Fig.2. focuses only on localization. ITD stays for Interaural Time Differences, ILD for Interaural Level Differences between the ears and HRTF for the Head-Related Transfer Function. Physiology describes the anatomy, and auditory models try to explain the localization. The traditional models are revised and replaced by more sophisticated models of parallel-distributed evaluation. Generally, the many parameters of localization are measured separated.

Localization measurements include binaural recordings and playback (Fig.3.). The main methods and environments during playback will be discussed later in Section 2. Many parameters have to be taken into account during headphone playback for a correct localization performance. Recordings can be made with different stimuli and equipment. Measurements and the quality of recordings made on "real humans" differ from those made on dummy-heads.

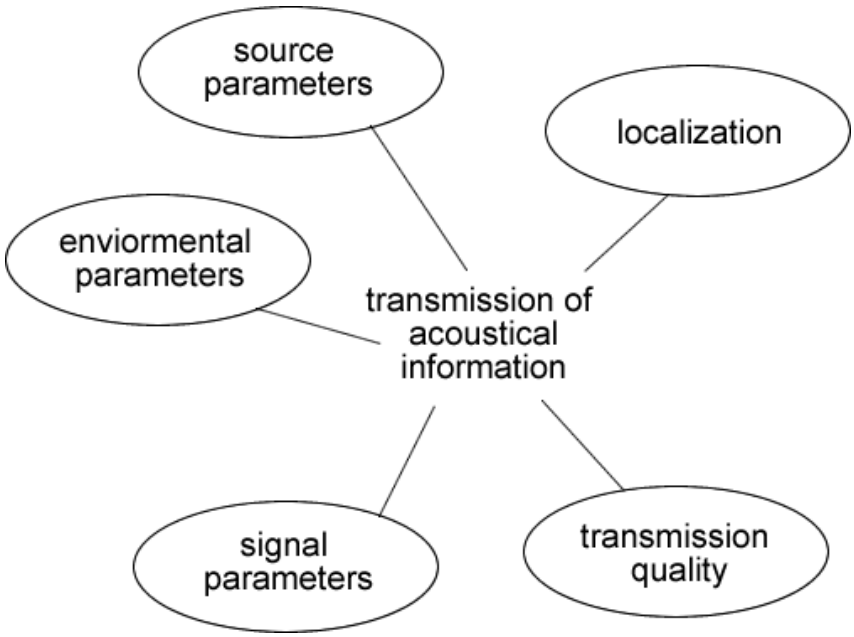


Fig.1. Information elements transmitted by the sound waves.

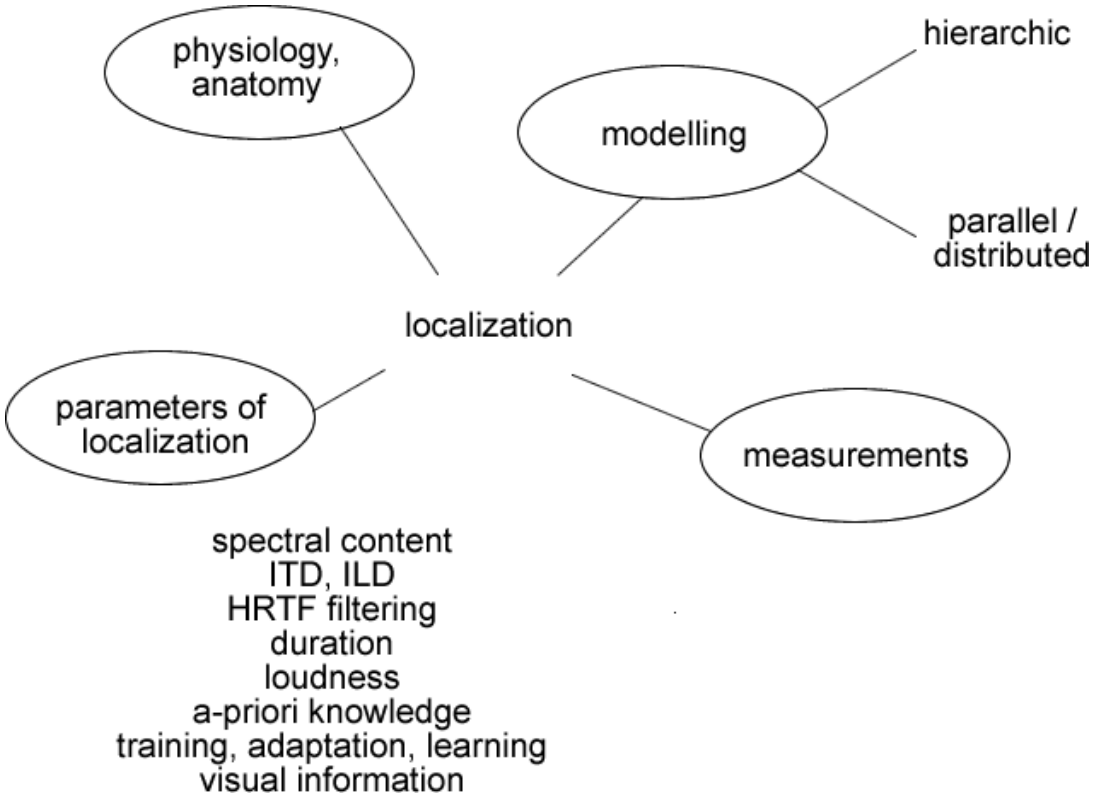


Fig.2. Disciplines related to localization.

This work analyses and contributes to the suggestions of the physiological hypotheses. It will be proved that although the sound pressure levels at the eardrums should contain almost all the directional information, higher processing algorithms in the brain do have significant effects and evaluation mechanisms even at the level of the outer ears.

Based on *Békésy's* early observation we focus on the role and effect of the transfer function of the outer ear. The microphones are placed at the eardrums of a dummy-head, and accurate measurements will be performed. We will demonstrate that little modifications of the acoustical environment near to the listeners' head do have significant effects on the transfer functions and thus, on the signal pressure at the eardrums. Nevertheless, in real life and free-field playback situation these "disturbances" of the transmission do not have any significant effect in localization performance.

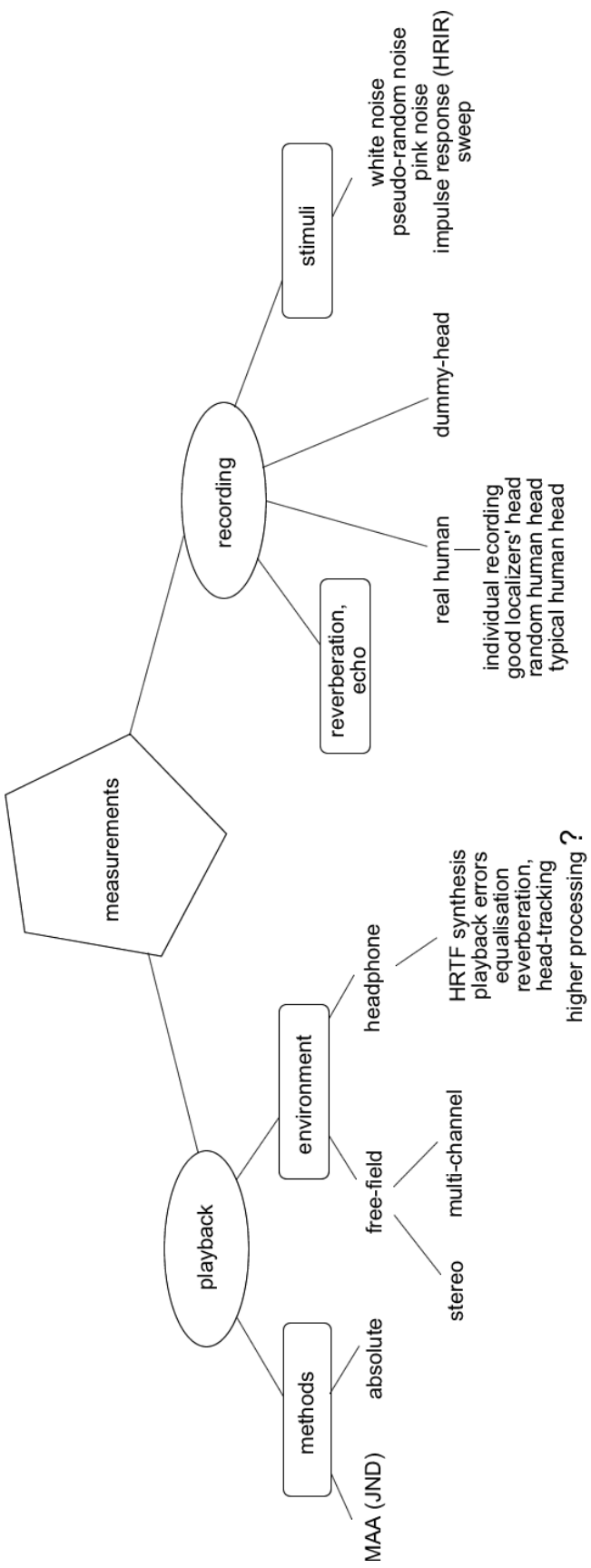


Fig.3. Classification and parameters of localization measurements.

Fig.4. shows the table of contents in a block diagram. After a short introduction the basics and the background theory together with former results and the stand of the technological level are presented. Physiology and anatomy of hearing, psychoacoustics, localization cues, definition and measurement methods of HRTFs in different experimental designs are listed in chapter 2. Open questions, problems and the goals of this work are also highlighted.
Before the measurement, a listening test was performed to test a good quality binaural headphone playback system. We were searching for typical headphone playback errors and the localization blur it can be achieved in this virtual audio simulation. Our results are comparable with former results by others, supporting the fact that virtual audio synthesis is inferior to loudspeaker playback. We will see that in this case only smaller disturbances are allowed and these could lead to decreased localization performance by losing the natural "air coupling" from the sound source to the eardrum. These results query the binaural techniques' statements, indicates and supports the hypothesis for the brain to have significant influence at the level of the external ears and that virtual audio simulation does not fail on the artificial filtering, accuracy of the transfer functions or signal processing algorithms. Detailed results of the listening tests can be found in chapter 3.
For measuring the HRTFs of a dummy-head a computer controlled, full automatic measurement system was installed in the anechoic room with increased spatial resolution, accuracy, reproducibility and signal-to-noise ratio based on our former system (chapter 4). After measuring the "bare" torsos' HRTFs the data of the "dressed" torso will be evaluated in the frequency domain in chapter 5 to 6. Evaluation of the peak-valley structure of HRTFs, head-shadow analysis and typical frequency limits are presented. Finally, the new achievements and results of this work are summarized in the results section in chapter 7 to 8.

The author would like to thank the following people:
Dr. Illényi András, as my Ph.D. consultant in Budapest, for the guiding and helpful comments given for a period of five years. Prof. Dr.-Ing. Klaus Fellbaum in Cottbus, for the opportunity to stay one year in Germany and for the possibility of producing the doctoral thesis.
All the colleagues at the Department of Telecommunications and Telematics in Budapest and the Lehrstuhl Kommunkationstechnik in Cottbus.
Steven C. Scheer and Dipl-Ing. Jörn Klimpel, for all the grammatical corrections and linguistic help in my papers, presentations and articles.

Dr. Borbély Gábor, Dr. Jámbor Attila and Szily István at the foundation "UNIVERSITAS" for the scholarship at the Széchenyi István University, Győr, for the financial support and lot of free time to work.
Special thanks go to Berényi Péter and Tátrai Richárd for their ideas, useful comments, and help during the setup and programming.
And last but not least for the 40 participants in Budapest and Cottbus, who contributed in the listening tests.

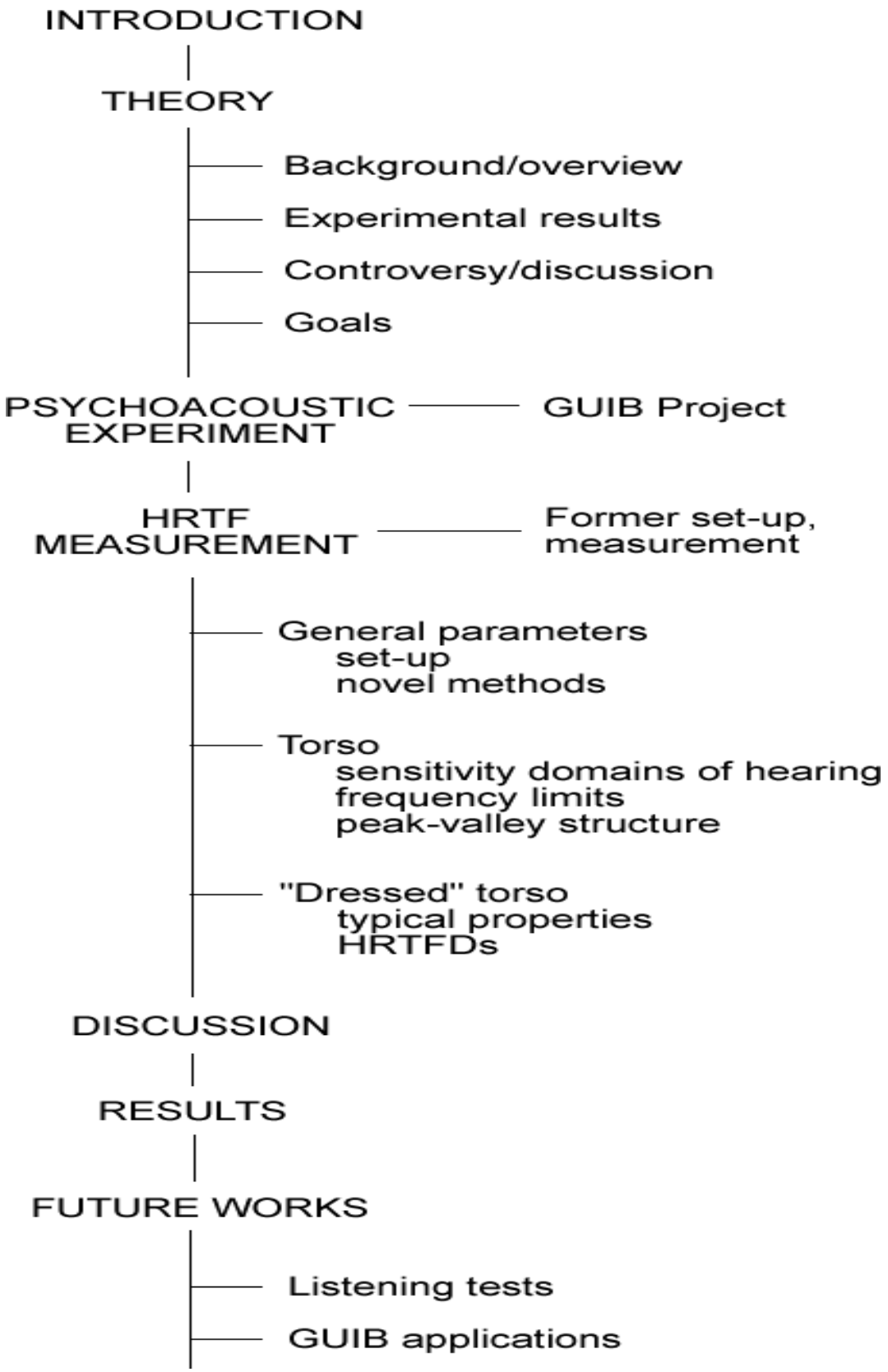


Fig.4. Block diagram and build-up of the doctoral thesis.

2 Background and overview

2.1 Physiology of hearing

This section is a brief overview of the physiology of the full auditory system from the outer ears to the cortex [1-6].
Physiological acoustics describes the structure and function of the auditory system: all of the acoustical and neural events that occur when sound waves evoke the spatial and temporal patterns of neuronal activity in the brain.
The auditory system provides an acoustical image of the external world by detecting, localizing and separating external sounds, creating coherent representation of events (fusion) and performing a frequency analysis of the sounds. Recently, the traditional, linear and hierarchical perspective of the auditory periphery has been revised as a more complex system. It has long been thought to function as a frequency analyser while more sophisticated functions like acoustical recognition has been considered the central pathway and higher processing. The traditional view requires serious revision in order to account for the auditory system's capability of encoding information under a wide range of environmental conditions. It is likely that the system uses special strategies to focus on the elements that contain meaningful components. The auditory system creates robust representations of significant information through the use of distributed, multiple representation and coding strategies.
The "place principle" of frequency coding describes the neural representation of spectral content. It is becoming clear that this cannot account for all the ways in which auditory information is processed and assembled into a complex representation of the external environment.
The peripheral auditory system can be divided into three parts: outer, middle and inner ear (Fig.5.). The outer ears and the outer shape of the human body are a complex antenna system that couples the eardrums to the sound field. Transfer function and interaural differences characterize the function and the way of localization. Physical models exist and describe the phenomenon. The pinnae has its individual form and the function to collect, reflect and shadow the incoming sound waves. The size of the ear canal entrance specifies the physical properties of the sound waves travelling inside. Approximately, up to 17 kHz the ear canal entrance functions as a "point source" and the *directional information* remains unchanged along the cavity of the ear canal. However, the

ear canal resonance together with the eardrum impedance is a special acoustic element.

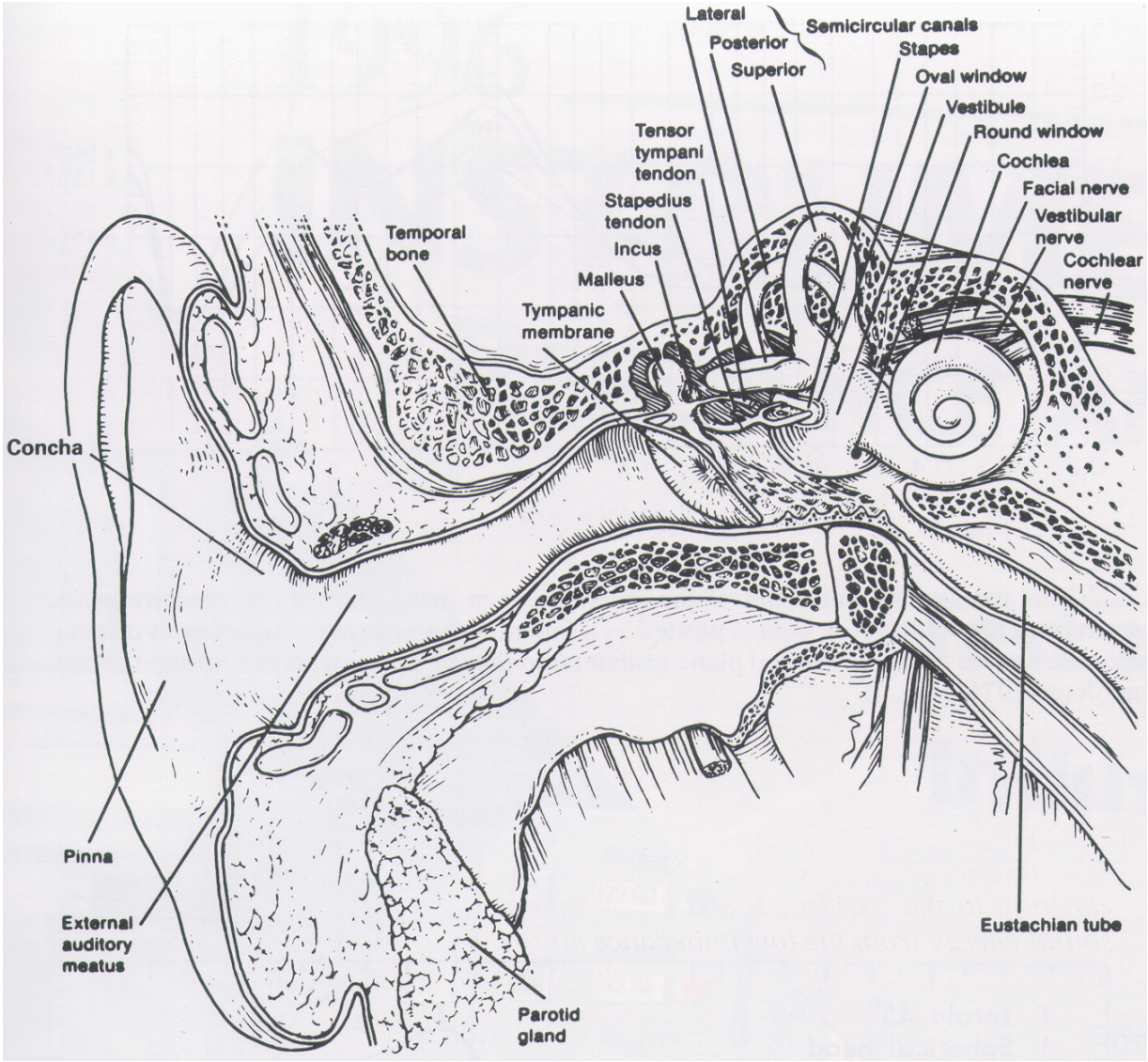


Fig.5. A cross section of the ear (adapted from *Blauert* and *Pickles* [5, 181])

The middle ear couples acoustic signals from the ear canal to the inner ear. The main parameters are: input impedance and transfer ratios (output volume velocity-input pressure, input-output pressure). The transmission can be regarded almost linear, with a pressure gain of 20-30 dB in the midband. The connection between the eardrum and the oval window is the linkage of three bones called hammer, anvil and stirrup in the air filled middle ear cavity. The function is to transmit the signals (the motion of the eardrum) to the cochlea.

The sound pressure is approximately constant, and a linear two-port model can be used with a power transmission of only about 50%. The most significant function of the middle ear is to transfer the incoming vibration from the large, low impedance eardrum to the much smaller, higher impedance oval window.
The main part of the inner ear is the cochlea. It contains sensory and supporting cells and the nerve fibres that are responsible for detection of sound. The vibrations of the eardrums are conducted through the middle ear by the bones into the fluid-filled cochlea by the oval window. The vibrations are distributed in a frequency-specific way along the length of the sensory organ (called the organ of Corti) to stimulate hair cells and nerve fibres. Mechanical vibrations are converted into electrical responses and neural impulses. The cochlea is a cavity having the shape of a coiled tube. In the organ of Corti, the sensory and supporting hair cells of about 20000 pieces rest on the basilar membrane. The movement of the basilar membrane is transmitted through the supporting cells so that the sensory cells vibrate at the same frequency as the stimulus itself. Sensory hair cells are designed for maximal sensitivity to small motions (non linearity). They transmit the mechanical information to nerve fibres that connect with the central nervous system. The organ of Corti rests on the basilar membrane. Sounds arriving in the inner ear create pressure fluctuations in the cochlear fluids, causing a displacement wave to propagate along the basilar membrane. This wave stimulates the hair cells, like tiny microphones. The pressure wave propagates very rapidly throughout the cochlear fluids but also causes a secondary and much slower transverse wave that travels on the membrane. This wave is called Békésy's travelling wave. The frequency of the stimulus will cause a distance dependent movement of the membrane. Thus, a "frequency map" is laid out along the cochlea, in which each longitudinal location of the basilar membrane vibrates at its characteristic frequency. Travelling waves produced by high frequency signals do not travel far up the cochlea in comparison with low frequency waves. This frequency mapping and spectral analysis is a non linear effect along the membrane.
Finally, the auditory nerve, which includes thousands of fibres, conveys information of action potentials from the cochlea to the brain. The fibres have different thresholds to the tones of some frequencies than of others. Neurons may have a role, not representing the direction of a sound, but in discriminating between directions of sources. The central auditory system is the neural system devoted to processing information about acoustic stimuli. The sensation of loudness seems to depend on the total quantity of activity in the auditory nerve. The ascending and descending auditory pathways create feedback between the brain and the peripheral parts.
The auditory system converts incoming sounds to neuronal signals, which are then processed in a very sophisticated way. Autocorrelation of the signals from

each ear as well as cross-correlation of the signals from both ears are performed. Specific inhibition and excitation effects are also present. Hearing models are based on physiological results but they try to model psychoacoustic findings through the simulation of the outer and inner ears and by creating set of algorithms for final evaluation [5].

2.2 Psychoacoustic localization cues

Psychoacoustics handles parameters like distance and location of sources, loudness, pitch, timbre or overall sound quality. Localization is one of the basic functions of the auditory system.
The outer ear modifies the sound wave in transferring the acoustic vibrations to the eardrum. Firstly, the resonances of the external ear increase the sound pressure at the eardrum, particularly in the range of 2-7 kHz. Secondly, the change in pressure depends on the direction of sound. This is an important cue for localization, first of all it enables us to distinguish above from below and front from behind.
The outer ear consists of the pinnae, which includes a resonant cavity called the concha. The ear canal leads then to the eardrum. The effects of the outer ear can be handled separately: one is the influence of the resonances on the sound pressure on the eardrum. The pressure increase is not due to power amplification because only passive elements are present. The other is the extent to which the outer ear provides directional filtering for help in localization [6, 23].
The different cues for the localization can be handled often separately. The *monaural* cues are responsible for the perception of elevation in the median plane, front-back directions and distance. The monaural parameters can be evaluated with one ear as well. Sources in the median plane create the same signal at the eardrums in case of perfect symmetry. The only tool for the hearing system is the directional filtering of the ears and this leads to a poor localization performance [5, 16-23]. The information for median plane localization comes from the concha and pinnae filtering. First of all, for short wavelengths the pinnae will show a directional selectivity. The external ear produces a spectral modulation of the incoming sound. Low frequency elevation cues do contribute in the vertical localization and the localization performance could be much better away from the median plane [24]. Timbre and loudness are also monaural properties that vary with elevation.

Because in the case of lateral movements the sound sources outside the median plane create slightly different signals on the eardrums, we get *interaural* differences. The closer (lateral) ears' signal level is in general higher and the signals reach the eardrum earlier than the other (contralateral) ear. The Interaural Time Delays (ITD) and the Interaural Level Differences (ILD) are the basic cues for the localization, and together with the spectral filtering this results in a much better localization performance in the horizontal plane [25-33]. If the signal contains significant components below 1600-2000 Hz lateral movements will be determined by the ITD else by the ILD. The hearing system is sensitive even to brief changes of interaural differences [42]. Based on the size of the head is the maximal ITD about 0,63 ms (Fig.6.). The minimal time difference to be perceived is only 0,03 ms corresponding to a location difference of 3°-5° [171].

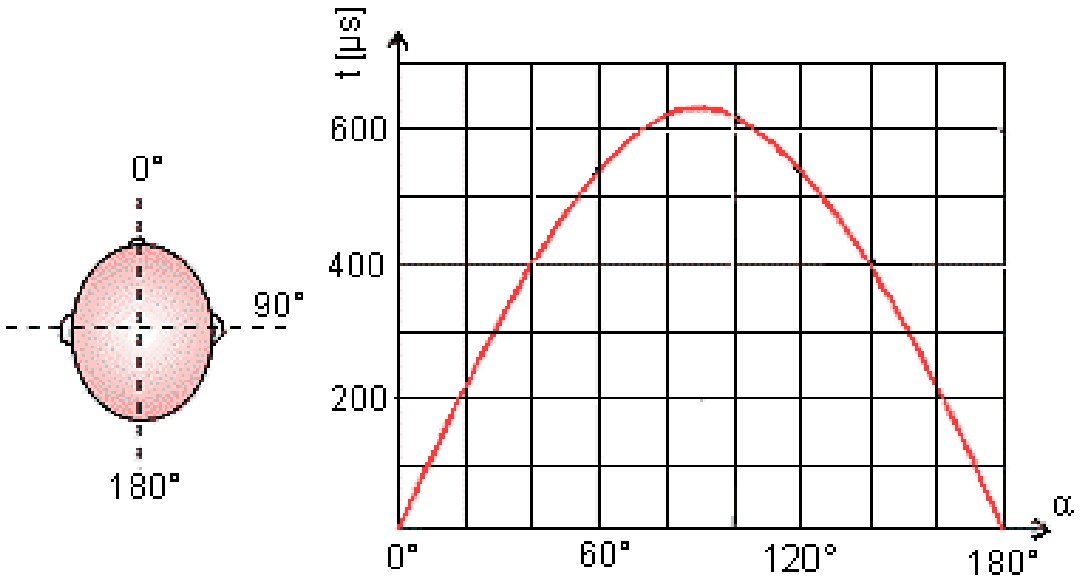


Fig.6. ITD as the function of azimuth [171].

Fig.7. shows ITD estimation based on geometrical calculations of head size data [51]. Head shadow area, time and path length difference between the ears can be calculated easily, but estimations like this model only basic geometrical properties.

The time-domain analysis of the monaural evaluation tries to calculate with secondary sound paths and primary reflections (in case of impulse excitations) [53-55]. Pinnae reflections make it possible to detect very short time delays, they cause spectral changes and these could be dominant for localization [15, 56, 57, 58]. On the other hand, it is unlikely that the hearing system contains time-domain analysis of the monaural cues, because structures limited under

1ms will not be evaluated and the accurate phase-spectra of the monaural parameters are not significant for the directional hearing [59]. The impulse response analysis of a model pinnae showed that the high frequency components are affected by the pinnae and secondary peaks can be suppressed by filling the cavities [54].

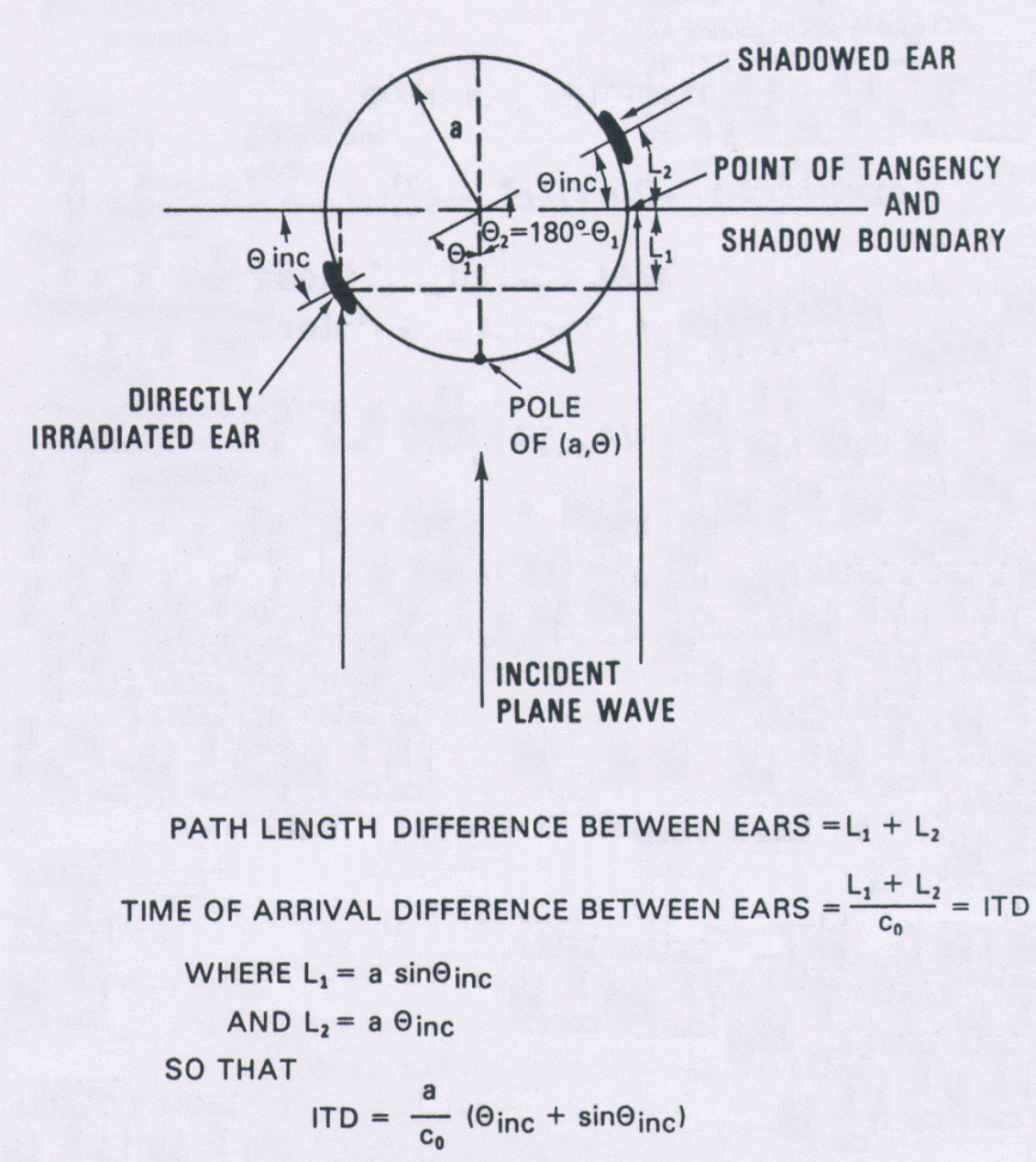


Fig.7. Estimated ITD values based on geometrical calculations of head data given by *Kuhn* [51].

2.2.1 Localization in free-field and virtual environments

The localization performance depends on many parameters, first of all on the monaural and interaural parameters. In the horizontal plane interaural differences are the basic cues. In the median plane it is the filtering effect of the pinnae, head and torso. Other basic psychoacoustic parameters for the localization are:

- spectral content
- bandwidth
- volume
- duration
- adaptation and learning, a-priori knowledge
- visual information.

The spectral shape, bandwidth and cut-off frequency of the signal also influences the localization. Broadband signals and high frequency components are better to localize. Broadband signals contain more information about the location of the source than narrow-band signals. White noise and filtered versions of white noise excitation are well suited for listening tests. Localization errors occur seldom using long-run, broadband and replayed signals. Auditory events of high pitch called "high tones" are tend to be localized at a higher elevation than events whose pitch is low, regardless of the direction of incidence.

Increase of the volume and the duration also increase the localization performance: sound sources between 40-80 dB SPL and signals over 250 ms are localized the best [5].

The localization of a human is time variant. It needs adaptation, learning phases and it is influenced by a-priori expectation and fatigue as well. It can be helpful if the subject is trained, familiar with the signal and the measurement procedure. The adaptation needs ca. 3-5 minutes. It has to be mentioned again that the directional information added by the filtering effects of the outer ears are complete at the entrance of the ear canal and this information does not vary along the cavity of the ear canal nor will it be affected by the eardrum impedance [8, 10, 11, 60].

The so-called *duplex theory* explains only the left-right displacements of sound sources but this is not sufficient if sources have the same ITD, placed on the "cone of confusion" (Fig.8.). *Shinn-Cunningham et al.* showed a "tori of confusion" with a single geometric approximation of a head (rigid perfect sphere). The ILD is constant in the head shadow zone along a cone of confusion [50]. Peaks in the ILD between 2-5 kHz are due to torso reflections and ILD vary with both distance and direction for sources within 1-2 meters to the

interaural axis. By relative distant sources only the head shadow contributed to the ILD. Measured ITDs correspond reasonably accurately at low and high frequencies to the computed theoretical values for a rigid sphere [51].

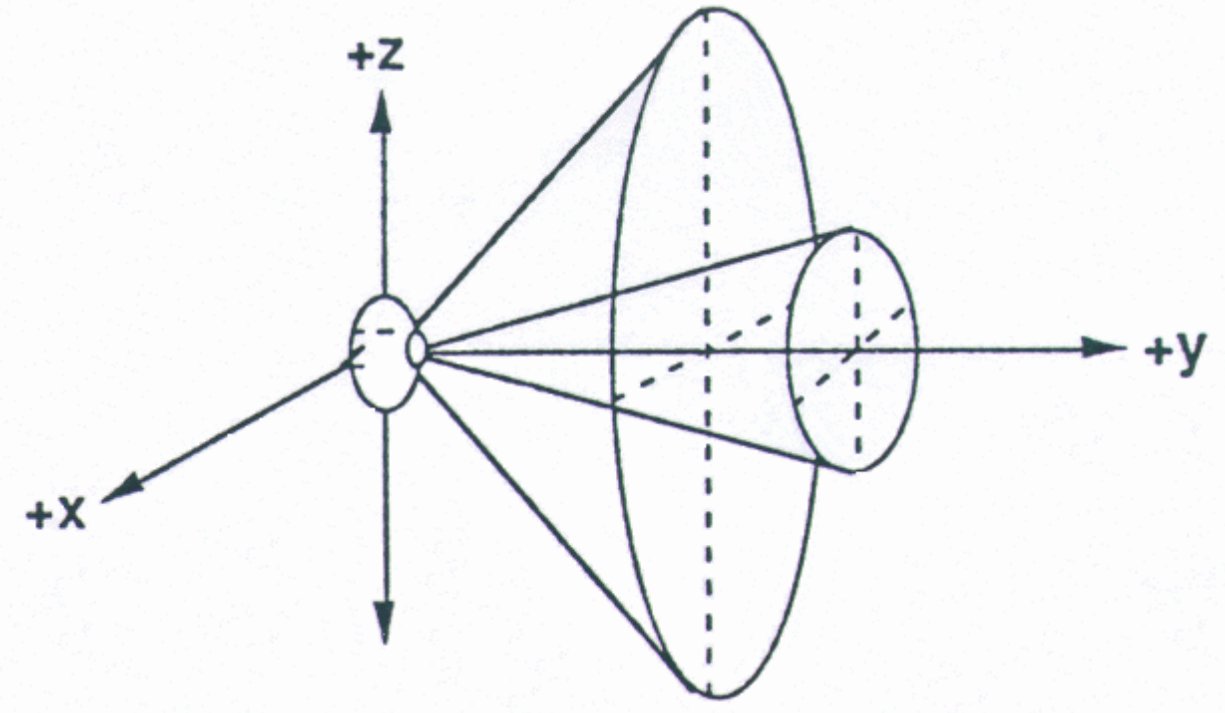


Fig.8. The “cone of confusion”. Sources having the same ITD are placed on a cone. Two different cones of confusion are presented for different ITDs [25,46].

Brungart analyzed the effect of sources near to the listener (less than 1 meter) with the rigid sphere model calculation and dummy-head data [52]. He found the pinnae effect is the same after some centimeters.
The lot of independent parameters indicate various methods and excitation signals for listening test related to the localization performance. Although the mentioned parameters and cues are the most relevant in every case, real life listening situations and virtual audio syntheses basically differ. The experimental apparatus on Fig.9 is adapted from *Roffler and Butler* [21]. When sinusoidal pulses at different frequencies are presented, the auditory event appears in the directions shown, regardless of the direction of the sound incidence. This indicates that signals with high frequency content tend to be localized at higher elevations.
Fig.10. and Fig.11. show the dependence of localization blur on the signal duration and sound pressure level [5]. Signals over 250 ms are localized more accurate than shorter impulses. A SPL of 50-60 dB is required for the best

localization performance. With other words, by signals longer than 250 ms and louder than 50 dB SPL is the localization independent of the duration and loudness.

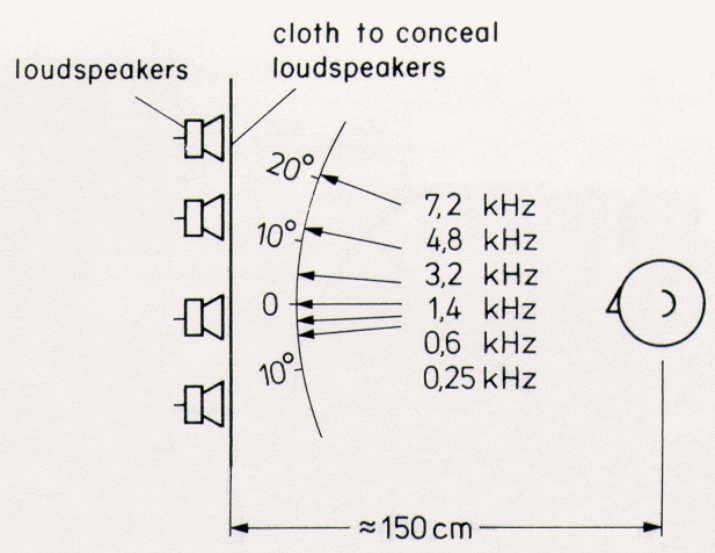


Fig.9. Experimental apparatus of *Roffler and Butler* [21].

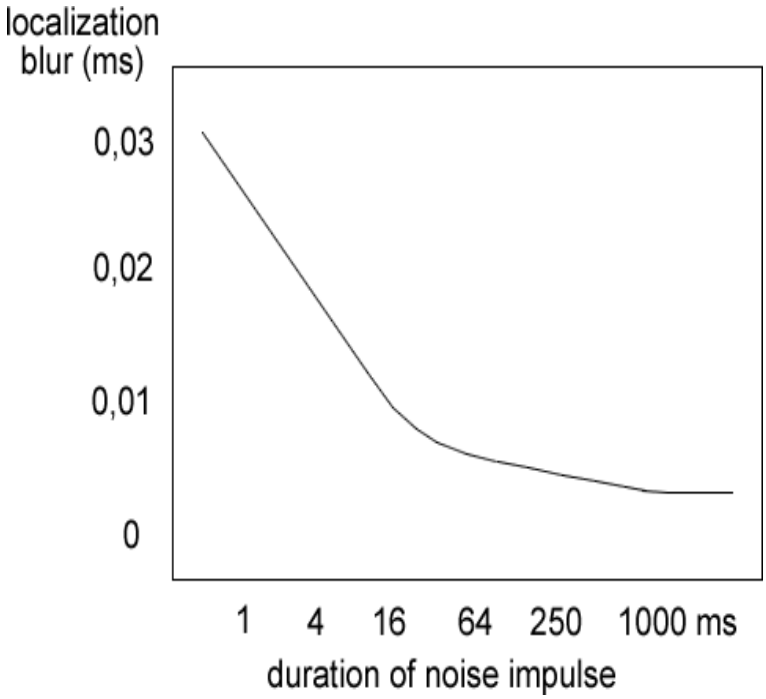


Fig.10. Dependence of localization blur on the noise signal duration [5].

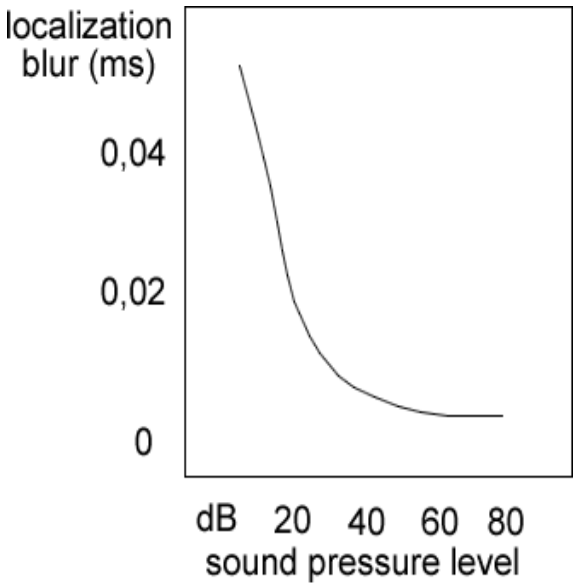


Fig.11. Dependence of localization blur on the SPL of a 500 Hz sinus signal [5].

2.3 The Head-Related Transfer Functions

Localization means finding the sound sources in the 3-dimensional space. During the localization the outer ears play a significant role. Sound waves reaching the eardrums are affected by directional filtering of the pinnae, head and the torso. This binaural filtering effect determines basically the perception of the direction of sound sources depending on the angle of incidence [5-14].
The transmission from a point in the free-field to the eardrums (or any other point within the ear canal) is described by the complex Head-Related Transfer Functions (HRTFs) or their time-domain variant: Head-Related Impulse Responses (HRIRs). The complex HRTFs contain information in the magnitude response and in the phase spectrum respectively. In everyday life environments humans use their individual HRTFs for the localization, but in virtual audio environments the HRTFs have to be reproduced "exactly" through headphones.
The HRTFs are defined in the Head-Related Coordinate System as seen in Fig.12. The two main parameters are: azimuth (φ) and elevation (δ). The shape of the human body and pinnae determine the directional dependent HRTFs. Therefore, they are individual and from the same direction a large deviation between subjects is natural [10, 12, 15].

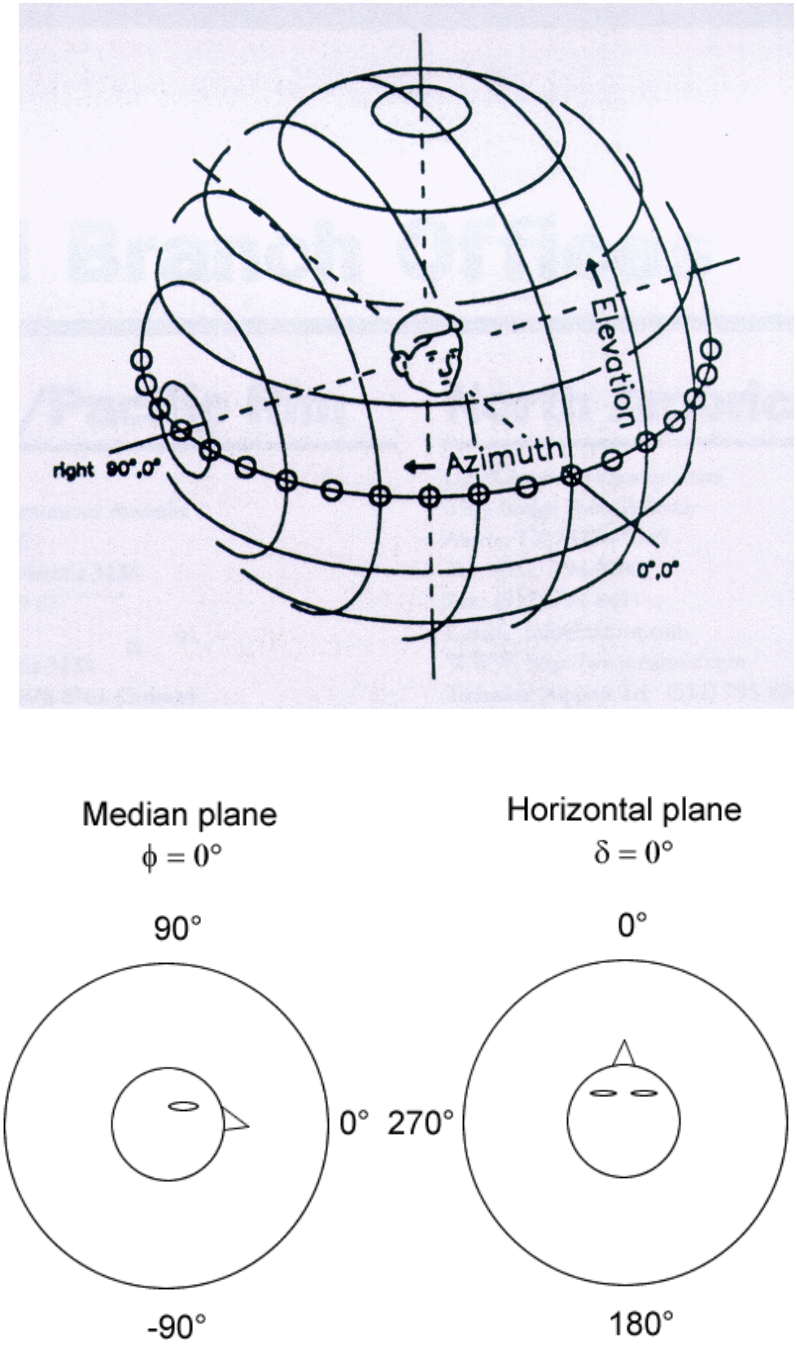


Fig.12. The Head-Related Coordinate System. Lateral movements are identified by $0° \leq \varphi \leq 359°$, elevational by $-90° \leq \delta \leq 90°$ in the horizontal plane and median plane respectively [5].

Eq.1. defines the free-field complex HRTF:

$$HRTF = \frac{P_1(j\omega)}{P_2(j\omega)} \tag{1}$$

where P_1 is the sound pressure at the eardrum and P_2 is the sound pressure in the origin of the head-related coordinate system at the same signal and sound source, but recorded with a unidirectional microphone (assuming that the dividing is mathematically correct) [5]. We plot the magnitude response for further investigation on a log-log axe: 20log/HRTF/.
HRIRs for the left and the right ear are defined with the convolution-integral:

$$s_L(t) = \int h_L(\tau)x(t-\tau)d\tau \tag{2}$$

$$s_R(t) = \int h_R(\tau)x(t-\tau)d\tau \tag{3}$$

where s(t) is the signal from the sound source, $h_L(t)$ and $h_R(t)$ are the impulse response functions for the left and right ear respectively. HRTFs in this case can be calculated by the Fourier transforms as usual.
Interaural HRTFs can be defined and calculated from monaural HRTFs [5]. Because the head and body of a human is never symmetric (in contrast to a dummy-head), interaural cues may take part in the median plane localization [34]. Monaural and interaural parameters occur in the real life together. First of all the ILD and ITD cannot be evaluated separated in normal listening situations.
The auditory system can utilize complex waveforms using ITD information at high frequencies just as accurately as time-delay low-frequency waveforms based on the fluctuating envelope [43]. In an accurate spatial simulation the phase-information can probably not be neglected. The minimum-phase-filter assumption of the HRTFs allows the specification of their phase by its magnitude response alone, so HRIR specifications and ITD information can be handled separately. The outer ears show minimum-phase properties up to 10 kHz [7]. The minimum-phase assumption does not restrict the performance significantly [44-48]. *Hammershøi* showed that noise stimuli spatialized with FIR filters of order 72 per channel is not significantly different from those generated from a measured 256 tap version, and rectangular windowing results

in an optimal non-minimum-phase FIR filter approximation of the HRIRs [49]. HRIRs contain a minimum-phase function part and non-minimum-phase zeros at high frequencies. They come from delayed reflections that are more energetic that the direct response caused by the pinnae. This could be important for the modelling, where we would like to have a parametrized model that can be customized for all listeners (e.g. with head geometry data) [179].

2.3.1 Measurement of HRTFs

There are a lot of methods to measure the HRTFs [5, 7, 10]. HRTF measurements and binaural recordings can be made on real human subjects or using dummy-heads [9, 69-72]. Head and torso simulators (HATs) are created to model the median human adult body geometrically: the reflections of the torso, shoulder, head and the pinnae. Microphones are placed at the eardrum or at the entrance of the ear canal. A HAT is more reflective than absorbing and it suits the reproducibility criteria [69, 73]. The torso itself influences the transmission below 2 kHz at frontal incidence and at φ=90° appear shoulder and pinnae effects. HATs are suitable for long-time objective measurements. Some information can be found about the effect of the torso with and without clothing in [74, 75]. Undressed torso causes sound pressure level increase at the head between 2-5 kHz. In a diffuse-field the fine structure of the torso and the head below 10 kHz is not significant. The SPL is at the head maximal. Clothing can also cause amplifying effects. If our goal is to get precise data from long-time measurements, dealing with human subjects is not preferred. Of course, measurements using dummy-heads have their problems and disadvantages [69, 70, 72]. Binaural recordings result in worse localization performance using dummy-head HRTFs than individuals [13, 70, 71].
Conventional acoustic measurements do not have the A and B-weighted subjective, psychoacoustic properties by the evaluation. For a dummy-head application the torso and the transmission chain has to be calibrated carefully by keeping the compatibility between the traditionally stereophonic recordings [5, 76, 77] or by using a 4-channel loudspeaker playback [78, 79]. To find the necessary elements of a head and torso simulator the structural analysis and modeling of the HRTFs is necessary [34, 80]. The SNR of a system and of a HAT can be treated separated. The latter can be influenced e.g. by the inner geometry of the head and body, the placement and properties of the microphone [73, 77, 81]. The goal is to get the maximal signal level at the microphones. The torso must have microphones with the *Zwislocki*-coupler simulating the average

eardrum impedance for a correct electroacoustical transmission (which is defined to be the ratio of the acoustic pressure to the volume velocity in the ear canal) [75]. Fig.13. shows the applied “ear simulator” of the Brüel&Kjær 4128 dummy-head.

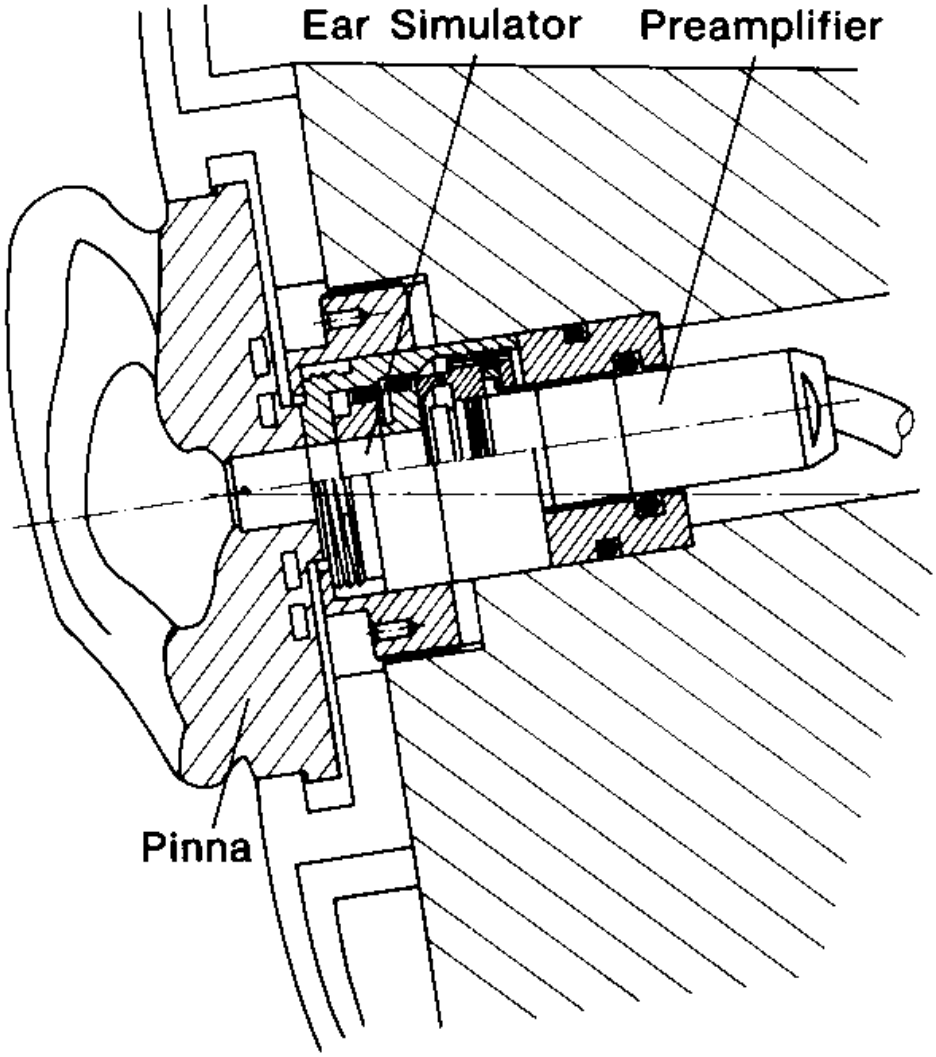


Fig.13. Cross section of the inner geometry of the used BK 4128 dummy-head according to the users manual. The ear simulator simulates the eardrum and ear canal impedance.

Fig.14. shows the electroacoustic *Thevinin* equivalent of the outer ears including pinnae, ear canal and eardrum impedance. The proper equalization for a FEC headphone (free ear coupled to the ears) can be found in [10, 62]. The sound transmission in the ear canal is direction independent [8-11, 60, 77, 82]. The ear canal is handled as a passive acoustic amplifier [83]. In real life the

impedance of the pinnae, ear canal and eardrum constitute a broadband (almost frequency independent) amplifier [77].

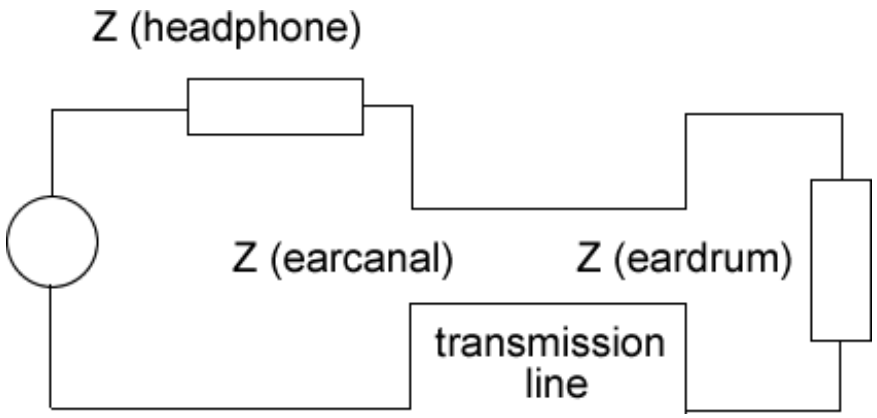


Fig.14. Thevenin –and electroacoustic equivalent of the outer ears and the headphone given by *Møller* [8-11,62].

For the major part of the audio frequency range the transmission to the eardrum proved to be independent of direction from points at the centerline of the ear canal. The entrance can be open or blocked and recording can be made at the entrance of the ear canal. The transmission from the free-field to the eardrum was divided into a directional-dependent part and two independent parts: (1) free-field to the blocked entrance (2) a pressure division between the radiation impedance and the ear canal input impedance and the (3) transmission along the ear canal. The blocked ear canal entrance is the most suitable for HRTF and binaural recordings, since sound at this point includes full spatial information and the minimum amount of individual information and deviations. The sound transmission along the ear canal is independent of direction, since only one mode of propagation (the longitudinal) is present. The sound field outside the ear canal is complicated due to diffractions around the pinnae, head and torso, and no simple prediction of the sound field can be made [11]. Blocked-ear-canal entrance seems to be the most suitable point for data collection, because the dynamic range of the HRIRs is optimal [84].

2.3.2 Dummy-heads and modeling in the measurement technique

The good torsos monaural and interaural functions are independent of the ear canal, only the outer geometry has to be modeled precisely [77, 79, 82, 85, 86]. Proper equalization of the measurement chain is needed [34, 76, 87-91]. *Genuit* discusses the problems of the calibrations of a dummy-head measurement chain and the structural averaged HRTFs from simple geometrical forms [83]. The following are responsible for shaping the outer geometry: torso and shoulders in the median plane up to 2 kHz, the head only in the horizontal plane, pinnae in the median and in the horizontal plane above 3 kHz. This simplifying allows easy calculations and simulations using only adding, filtering and all-pass elements and only a few parameters for setting the HRTFs [82, 92].
In [44, 93] we found a system-theoretic mathematical description of the variations of the sound pressure, reflections and declinations at a sphere and cylinder using 28-36 parameters to set the HRTFs for all directions. These HRTFs are within a 5-7 dB range, which is as good as the reproducibility of the HRTF measurement with humans.
For modeling the HRTFs the common properties of the human HRTFs have to be extracted. Averaging can only be made without smoothing of the significant peaks [94, 95]. This is called structural or parametric averaging. Typical asymmetry exists between the left and right ear, and different HRTFs have to be determined for each ear even in case of a monaural source [34]. Another smoothing method is the ARMA modeling [96].

2.3.3 Measurement signals and signal-to-noise ratio

HRTFs usually are measured with impulse excitations or with noise stimuli. White noise is a wide band of random noise (i. e. a signal containing all the frequencies of the spectrum with a random amplitude distribution) with a constant level per Hertz over the entire spectrum. In the time-domain it can be described as

$$p(t) = \sum_x \hat{p}_x \cos(\omega_x t + \varphi_x) \qquad (4)$$

where $\widehat{p}_x$ are almost identical, ω_x goes from 20 Hz to 20 kHz and φ_x is a random variable with uniform distribution between [0, 2π]. Pink noise is similar, but with a level decreasing by 3 dB per octave.
The signal-to-noise ratio is defined in an analog system as follows:

$$SNR = \left(\frac{S}{N}\right)[dB] = 10\log\left(\frac{S}{N}\right) \tag{5}$$

where S is the signal power and N is the noise power. In case of amplitudes, the multiplier constant is 20. The SNR is a good measure to estimate the quality of the transmission.
The signal-to-noise ratio in a digital system is defined as the ratio of the (incoming) signal power and the power of the quantisation noise. In practice, we use for estimation the following equation:

$$SNR \approx 1{,}74 + 6{,}02n \tag{6}$$

where *n* is the number of used bits (word-length) [155, 156]. If the signal to be converted from analog to digital does not use the entire dynamic range presented by *n* bits, the SNR will decrease, e.g., if only 15 bits are used in a 16 bit system, the SNR decreases with 6 dB. The SNR is calculated from the effective number of bits, which means, the SNR depends on the signal. A quadratic signal between ±32768 could be converted noiseless with 16 bits.
In a hybrid system where digital and analog components function parallel, the SNR are basically determined by the digital part based on the signal processing. Usually the analog circuits, amplifiers etc. have a better SNR than the computed digital signal-to-quantisation noise ratio. However, it can be increased by common used SNR improvement techniques. E.g. averaging helps against the random uncorrelated measurement noise.

2.4 Experimental results

2.4.1 The binaural technique

According to the statements of the *binaural technique*, if we reproduce the sound pressures at the eardrums exactly, the reproduced signal will have the full spatial information about the environment it was recorded in. For the reproduction a proper and individual headphone-equalization is required, as far as possible [60-63]. Models of the hearing system based on the binaural technique or neural networks try to explain the localization in free-field environments and in rooms with small reverberation time. This technique may contain errors as well, like front-back confusion and in-the-head localization due to headphone playback [64-68].
Binaural fusion is when the ear signals are correlated and a well-defined and localized sensation arises [94]. This is valid even in reverberant rooms up to an ITD of 1-80 ms. Spectral analysis is made in the inner ear by a "filter bank" represented by the hair cells and the basilar membrane. These can be modeled in the inner ear, using neural networks or alternative modeling methods and data reduction methods [67, 68, 94, 98, 100, 176].
Localization in binaural recordings made with artificial heads is inferior to real-life localization as well as to localization in binaural recordings made in the ears of selected humans, according to a paper by *Minnar et al.* [4]. In a series of experiments it was shown that artificial heads are still not as good for binaural recording as a well-selected human head, although some of the new artificial heads approach the performance of human heads.

2.4.2 Subjective listening tests in virtual audio synthesis

Localization means finding the absolute position of the sound source. *Localization blur* is the smallest change in the direction of the sound source, which can be perceived. There are two basic methods to measure the localization performance.

The first is the so-called *absolute* measurement, where the subjects have to localize and tell where to source is by pointing to it [99]. This is for loudspeaker playback. The question relating to *localization* is: where is the sound source?
The other solution is to search for the Minimum Audible Angle (MAA) or the Just Noticeable Difference (JND), where the subjects only have to compare two sound sources and identify only the change of the source direction [100-105]. The question relating to *localization blur* is: what is the smallest change of the position of the sound source that produces a noticeable change of the auditory event? The MAA is usually measured in degrees, but it can be determined in interaural parameters (dB or μs.) and these results can be calculated from each other [106, 107]. The MAA method is easier and delivers better results.
The Minimum Audible Movement Angle (MAMA) differs from the MAA. The former can be determined as the function of velocity, but the performance seems to be independent from the MAA [103, 108, 109]. It was found that both MAA and MAMA are optimal for signals that either are below 1000 Hz or above 3000-4000 Hz. The auditory system seems not to be able to perceive velocity, subjects make discriminations based on distance and space [110]. MAA has a minimum between 250 and 1000 Hz, and above increases to a maximum. Between 3 and 6 kHz there is another minimum. Usually, the MAA is smaller at large azimuths in the median plane [51, 110].

2.4.3 Localization results of listening tests

Free-field or real-life situations are easy to investigate. We only need a suitable anechoic room with a set of loudspeakers. Subjects are sitting in the room mostly with a fixed head and declare in an absolute or MAA measurement.
Virtual audio simulation needs good quality headphones and the artificial reproduction of the transfer function between the ear(drum) and the sound source. This means, the digital representation of the HRTF filtering. This introduced the techniques and problems of the HRTF measurements as well. Of course, a proper equalization of the headphones' transfer function is also necessary. In general the results from free-field measurements tend to be better than when using headphone playback [122, 123].
Results on this field are difficult to compare, because experimental designs and methods differ. For a direct comparison of results, similar conditions are needed. Furthermore, better results can be obtained in a MAA measurement in contrast to an absolute measurement. To show the wide range of measurement methods and results, a bibliographic overview can be found in Appendix B.

Tables 15 and 16 show the huge range of measured data depending on stimuli and method.
Here we focus only on results from others using similar conditions, as we will work with: headphone playback and noise stimuli (see also summarized in Table 8 and 9).

Free-field listening

An overview of experimental results before 1970 was given by *Blauert* [5]. Research has shown that the region of most precise spatial hearing lies in or close to the forward direction. The absolute lower limit for the localization blur is about 1 degree using broadband signals in free-field listening. In the horizontal plane MAA results are 1°-2° better for broadband signals than for absolute measurements [177].
R.S. Heffner and H.E. Heffner measured the MAA of noise signals. The minimal MAA of 1,3°-1,8° are reported frontal and about 9°-10° at φ=90° in the horizontal plane [169].
Haustein and Schirmer used 100 ms long white noise impulses. The average values of 900 subjects are ±3,6° frontal and ±10° at the sides [168].
The same was observed by *Litovsky and Ashmed* in the investigation of the development of the auditory system by children and young adults [170]. After five years of age a person reaches a MAA of about 1°.
Discrimination error in the horizontal plane was found to be the best in the forward direction both by *Hartmann* (1°) and *Cohen and Wenzel* (1°-5°) [12, 25]. The localization blur increases depending on the stimuli as well, e.g. for a speech signal of unknown speaker it is 17°, for a known speaker 9° and ca. 4° for white noise in the front direction [178].
The localization blur measured by *Wettschurek* in the median plane for white noise is ca. ±4° frontal and reaches ±10° above [177]. For a low-frequency noise with a cut-off frequency of 4 kHz the results are two times greater. Results obtained with speech signals are ±9° frontal, ±10° at δ=36°, and ±13° to ±22° above [5].
Wenzel and Fosters' free-field measurement with 16 subjects showed average error values at low elevations of about 25° and ca. 22° at the sides [165].
Middlebrooks obtained results of subjects between 19 and 36 years old using synthesised broadband noise stimuli of 150 ms [173]. Results showed the best performance for noise in the front: 5,8° mean error values horizontal and 5,7° vertical.

Makous and Middlebrooks used 150 ms broadband noise bursts in an absolute free-field listening test [117]. The minimal average error was found frontal: 2° horizontal and 3,5° vertical. From the sides it is 20°. The standard deviation showed that 94% of the tests are less than 10 degrees away from the mean value. The median plane MAA results of 4° of *Wettschurek* showed the same standard deviations [177].

Headphone

Under optimal conditions (individual HRTFs, forward direction and broadband noise stimuli) by *Begault* is the MAA 1° and it increases with location and signal type [133].
Oldfield and Parker reported in an absolute measurement ca. 9° azimuthal mean value and 12° elevational mean value [135, 172]. The azimuth error values were between 4°-6° (φ=0° to 80°) and the elevational errors between 6°-8°. Without HRTF filtering the values increase up to 11,9° and 21,9° respectively.
*Wenzel and Foster*s' virtual measurement with 16 subjects showed average error values using non-individualized HRTFs ca. 24° at low elevations and 23° at the sides [46, 165]. Broadband noise bursts of 250 ms with 300 ms pause between were presented through the HRTFs of a good localizer. They found a good localization performance only if the listener was also a good localizer. No elevation shift was observed with noise signals. On Fig.15. the median plane results of *Wightman and Kistler* are also presented: average error at lower elevations of 21 degrees and ca. 20 degrees at the sides [120].
An interesting investigation of *McKinley and Ericson* about the average azimuth localization error was both in absolute measurement and MAA 5° [174]. The error values using octave band noise are between 4,41° to 5,87° depending on the center frequency as shown in Figure 16. The use of pink noise stimulus increases the values up to 6°-7°. The MAA in vertical direction using KEMAR HRTFs is 30°-35°. These results are twice as great as the MAA results of *Hartmann and Rakerd* [100].
Further results of localization tests are listed in Appendix B.

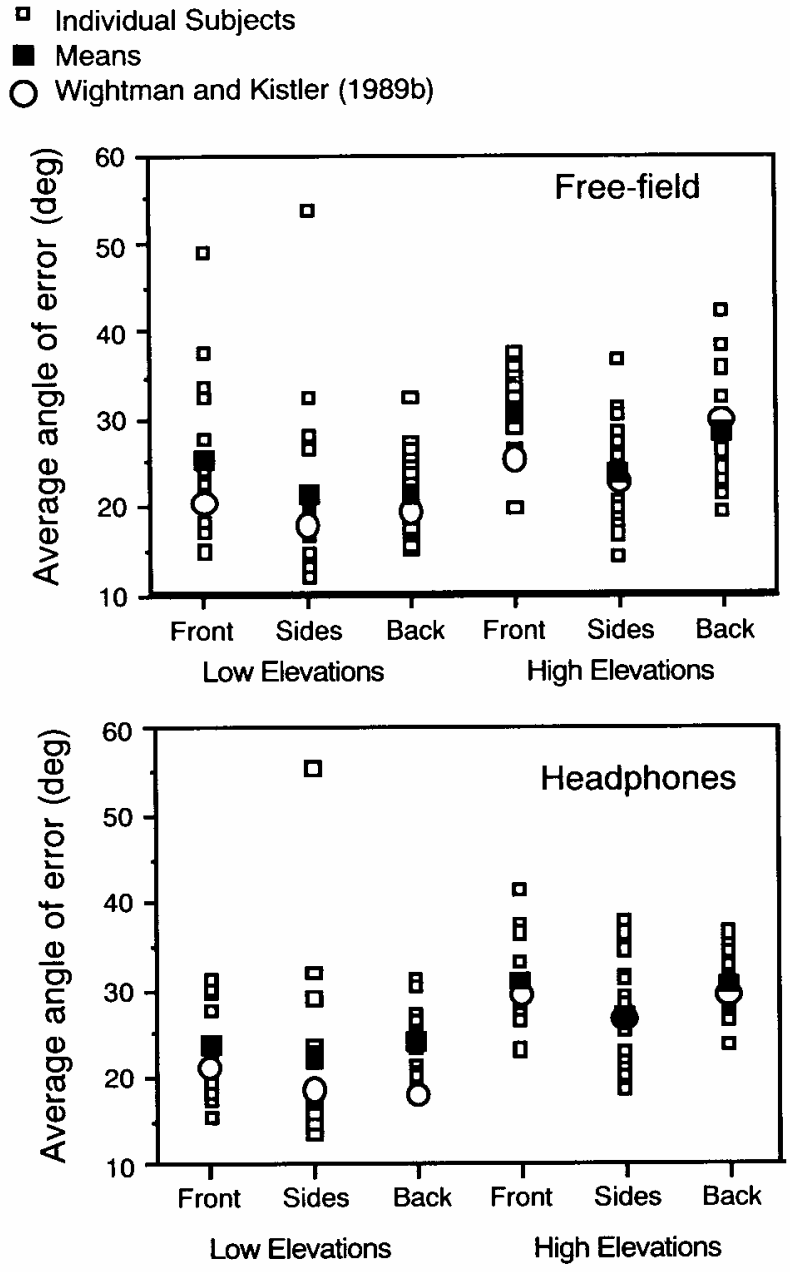
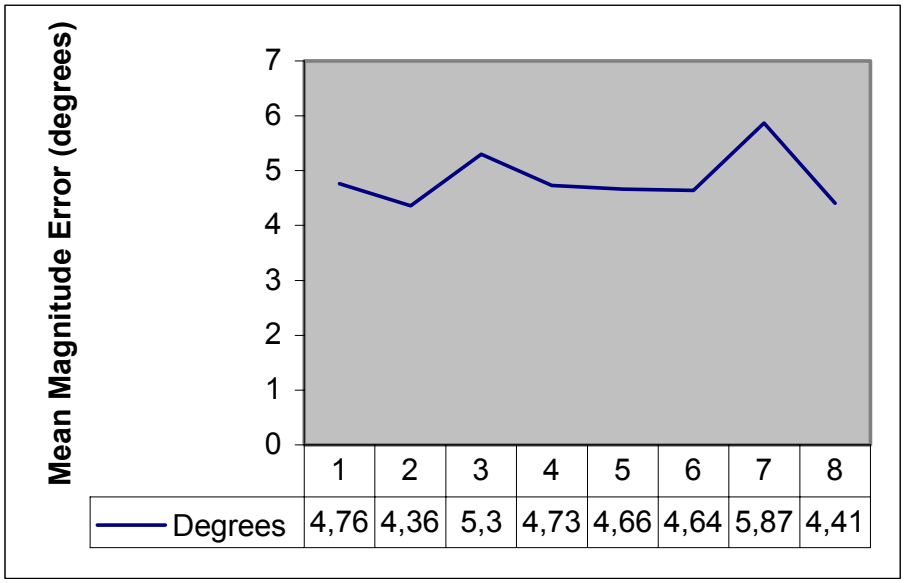
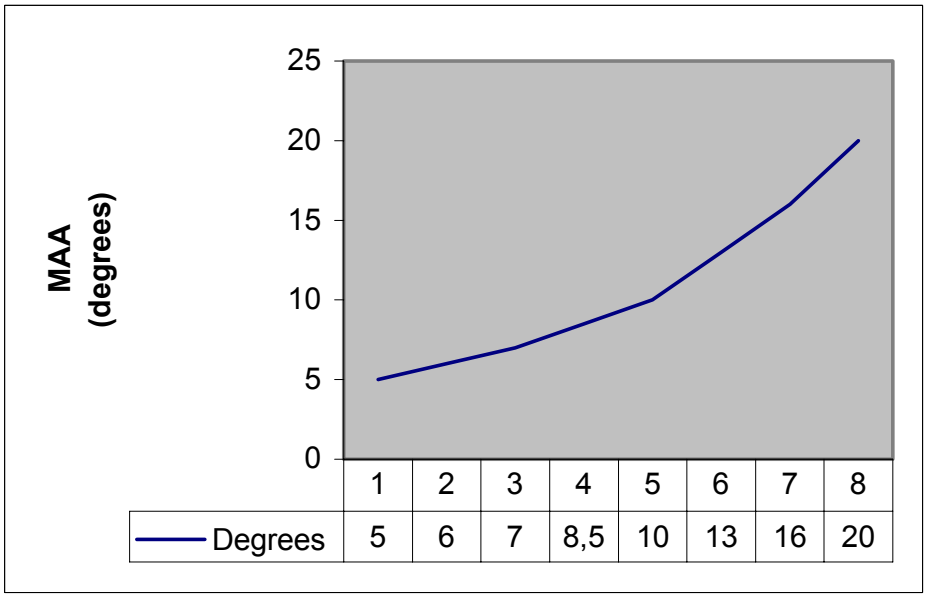


Fig.15. Results of 16 subjects measured by *Wenzel et al.* Individual and average localization angle errors are given in free-field listening and using headphones with non-individualized HRTFs [165]. Data of *Wightman and Kistler* are also shown for comparison [120].



(a)



(b)

Fig.16.

(a) Average localization error using octave band noise signal with the given center frequency: 1: Pink noise, 2: 125 Hz, 3: 250 Hz, 4: 500 Hz, 5: 1000 Hz, 6: 2000 Hz, 7: 4000 Hz, 8: 8000 Hz. [174].

(b) MAA of a 500 Hz pure tone at the given azimuth in the horizontal plane: 1: 0º, 2: 15º, 3: 30º, 4: 45º, 5: 60º, 6: 75º, 7: 90º, 8: 105º.

2.4.4 Headphone playback errors

The application of a headphone in a virtual synthesis introduces well-known errors. These are:

- in-the-head localization (the lack of "externalization")
- front-back confusion
- sources too near
- elevation shift
- ambiguity of movements symmetrical to the median plane.

The separated reproduction of interaural differences often leads to *in-the-head localization* [35-39]. In-the-head localization is when the distance of the perceived sound image is less than the head diameter. This happens generally in case of identical ear signals and monaural, median plane sound sources. Factors that are supposed to cause in-the-head localization are: lack of head movements, absence of bone conduction, distortions of the headphone, etc. In a free-field representation, ITD and ILD information cannot be handled separately, only through headphones. If signals are presented to the two ears sufficiently similar in frequency and the ITD is not too great, they are perceived as a single entity, called binaural fused auditory image. When heard from an external source, these images are typically externalized and reported to be from a source „outside of the head". On the other hand, images produced via headphones are typically reported to be „inside of the head". Simultaneously recorded direct and indirect sound or reverberations may decrease this phenomenon in a diffuse-field using an artificial head [40, 41]. Externalized sources are precisely said to be "localized", whereas those heard inside the head are "lateralized".

Front-back confusion or *reversal error* means to be confused about the direction of a sound source in the front and/or in the back. The reason is that median plane sources produce similar SPL at the eardrums and no interaural differences. It appears more frequently in case of a frontal source than in case of a source in the back. The reversal errors are natural even in real-life situations but it is strongly increased during headphone playback by losing head movements and reverberation. Results about the *rate* of front-back errors in listening test show a wide range depending on the playback situation: 1-22% [35], 6-20% [48, 120], 20-43% [116], 29% [133], 25% on average [134], 2-10% and 6% on average [117], 3,4% on average [135]. For further information of front-back judgments in case of interaural differences in the horizontal and vertical plane with noise and/or pure stimulus see [35, 93, 96, 136]. It is possible that sounds heard within the head are less precisely localized, but higher elevational errors were also found by externalised images [179, 180]. Head tracking was to be found significant only for solving reversal problems, in

contrast to previous investigations indicating that head movements enhance source externalisation. For broadband Gaussian noise a 7:1 decrease in front-back reversals and a 2:1 decrease in back-front reversals was found when head motion cues were supplied. No other treatments yielded significant results for solving reversal confusions. These results indicate that information due to the change in SPL induced by head movements (in order of 1-2°) is needed in order to localize sound images in front of a subject correctly, and this should be taken into consideration to achieve 3D reproduction using headphones, otherwise, it would seem to be difficult to avoid front-back confusion and elevation shift [65, 79].
Elevation shift is less significant. This means that a perceived sound direction tends to be "higher" than it actually was. Sound sources are also heard usually closer to the listener than their actual distance was during the recording. This is probably due to the trancducer of the headphone, which is much closer to the ears than any other sound source in real life sound fields.

2.4.5 Quality of HRTFs in the virtual audio simulation

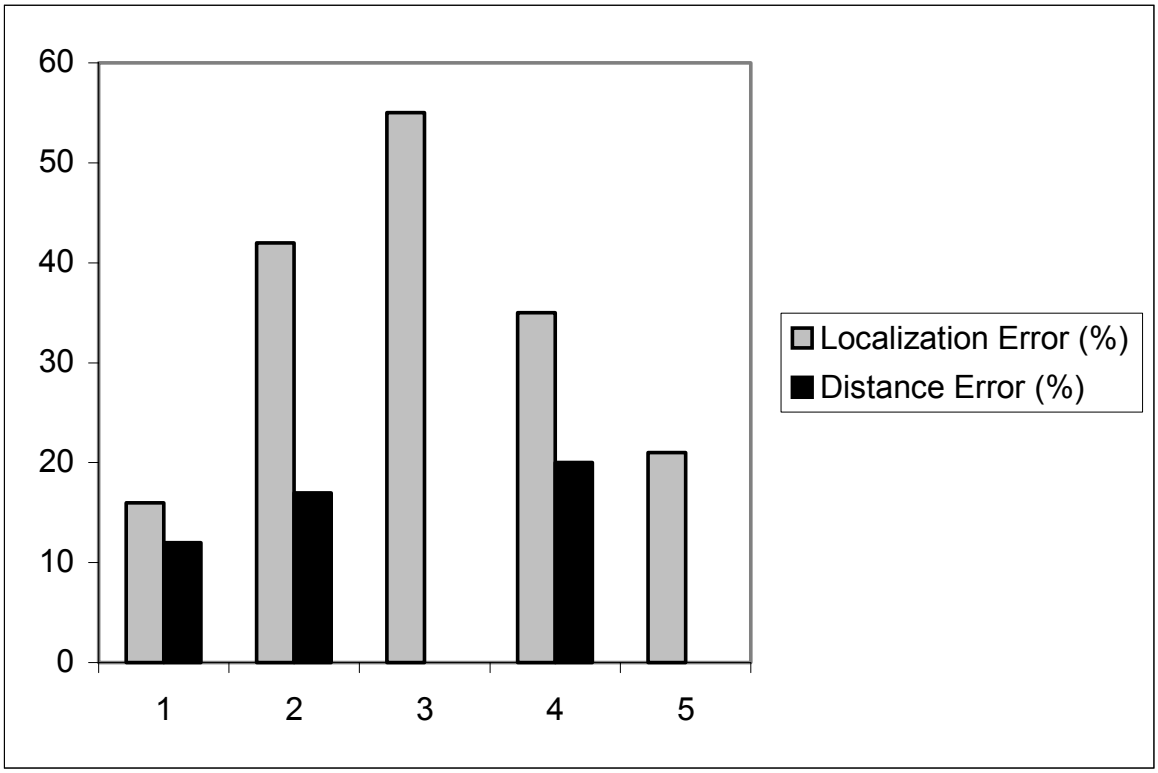
By headphone playback, the natural HRTF filtering is disabled and for a correct spatial and binaural playback the HRTFs have to be reproduced artificially. In general, this can be made with the HRIRs. The convolution of the input signal and the HRIRs is made in real-time [7, 111, 112].
Every human has his own individual HRTFs. For the reproduction we can use these individual HRTFs, HRTFs from a "good localizer" or from a dummy-head. The quality and localization performance using different sets of HRTFs has a wide range of interests [46, 70, 72].
Individual HRTFs are usually recommended for the best localization performance, but their measurement takes many efforts.
It was shown, that HRTFs from a good localizer and the use of simple methods to make them more individual results in a satisfying localization [13, 46, 116-119]. With the help of a simple frequency-scaling factor, individual properties can be included for reproduction. For this transformation (frequency shift) of the geometrical difference between the head and the pinnae of the dummy-head a human subject is needed to get better localization performance. Inter-subject spectral differences can be minimalized with a scaling factor depending on the interaural delays and the size of the head and pinnae. The increase in percentage of the localization is better than the difference in percentage between the scaled and non-scaled HRTFs [116, 118].

Non-individualized HRTFs have the advantage to measure them fast and easily, but it may lead to decreased localization performance [95, 115]. Using non-individualized (generic) HRTFs have been cited as degrading localization accuracy and externalisation and increasing reversal errors, typically for full-spectrum noise stimuli. *Møller* suggested that non-individual HRTFs result in increased number of reversals but have no effect on externalization [13]. With 19 loudspeakers in a standard listening room the largest localization errors occur in the upper median plane and subjects do not localize better with other subjects HRTFs than with their own. Using non-individual HRTFs results in more median plane and distance error but not necessarily in elevation shift or in-the-head localization.

Generic HRTF could be the HRTFs from another person or from a dummy-head. HRTFs from a HAT seem to be insufficient for a binaural playback [10, 72]. In a standard listening room, real life listening showed correct localization with some errors in the median plane (16%). Using HRTFs from a HAT introduced low median plane error increase of about 36-55% together with the increase of front-back confusion. Random humans' HRTF tend to be better (35% error) than HRTFs from a HAT (see Table 1 and Fig.17.).

Minnaar et al. reported median plane errors of 9,7% in real-life and 33-53% with HATs [4]. The distance and overall localization error also increased. 60% of real human heads are better than even the best HATs. Using HRTFs from a "good localizer" or even from a random selected human subject can also deliver better results [4, 25, 46, 119, 120].

The question is whether only the differences between a real human and a torso simulator are responsible for decreased localization performance, or whether there exist higher processes to help.



Errors [%]	Real life	BK 4128	All HATs	Random human head	Typical human head
Median plane	16	42	36-55	35	21
Distance	12	17		20	

Fig.17. Localization and distance error in listening tests measured by *Møller* [72]. Table 1 shows that free-field listening is of superior quality. Results with dummy-head HRTFs have decreased quality in contrast to the HRTFs of a typical or random selected human head. The Brüel&Kjær 4128 is shown detailed, because this dummy-head will be used in our measurement.

2.4.6 Virtual Acoustic Displays

Virtual Acoustic Displays (VAD) are widely used in several applications. VAD identifies a virtual environment, where sound sources are artificially reproduced and the listeners are able to localize and identify them. This acoustic information may appear without the visual help of the eyes or by completing the visual information. Such an application can be useful e.g. for blind PC users, flight simulators or computer games.

To realize a VAD two independent questions have to be answered. First, which sounds correspond the best to the visual and deeper meaning of the object to be reproduced [127]? In other words, what is the best mapping between sounds and events on the screen? Second, what is the localization blur like through headphone playback? How many sources can be identified simultaneously and in which resolution? The first question is more psychological to evaluate, the second one is pure psychoacoustic.
In principle, three-dimensional VADs can be realized by reproducing *depth* or *distance* information as well, e.g. by an object approaching the listener or by overlapping windows. The principal goal is to reproduce and replace the visual objects of a computer screen by keeping their spatial distribution on the two-dimensional VAD.
State-of-the-art multimedia computers and applications nowadays allow full auralization and orientation in a virtual reality (Fig.18.). Auralization means rendering audible sound fields by physical or mathematical modeling [113, 114]. Only the last decade made it possible to handle huge amount of computation data, real-time filtering of HRTFs, reverberation and head movements [25, 132, 133, 174]. In an experiment of *Begault et al.* independent variables of head tracking, individualized HRTFs and early and diffuse reflections were chosen to evaluate the influence of localization performance using headphone-delivered speech stimuli in a VAD [180]. The relative advantage of these has never been compared directly before.
An architecture presented by *Blauert* is capable of presenting complex auditory environment including head-tracking, vision and tactile/thermal modality (Fig.19.) [121]. Calculations of secondary sound sources (with time delays), direction of emission and incidence and wall reflections (auralization) is made by real-time convolution of HRIRs. Absorption of sound in the air and complex directivity characteristics of the sources can be accounted as well. FIR filters are of 80 taps, 43 sets in a resolution of 11,25° azimuth and 22,5° elevation angle. Binaural playback shows smaller error in case of consistent simulation of "static" HRTFs and ITD information [65].
These systems are still very expensive and far away for users sitting at home. Measurement of individual HRTFs, the equalization techniques of good quality headphones with the needed computational performance, are not intended for home applications. Furthermore, recently applied techniques still do not explain or fulfill the requirements for the proper simulation of the directional information, nor do they completely avoid headphone playback errors [123, 180]. The advent and the availability of the necessary computer power for real-time processing of audio signals will initiate new technology.

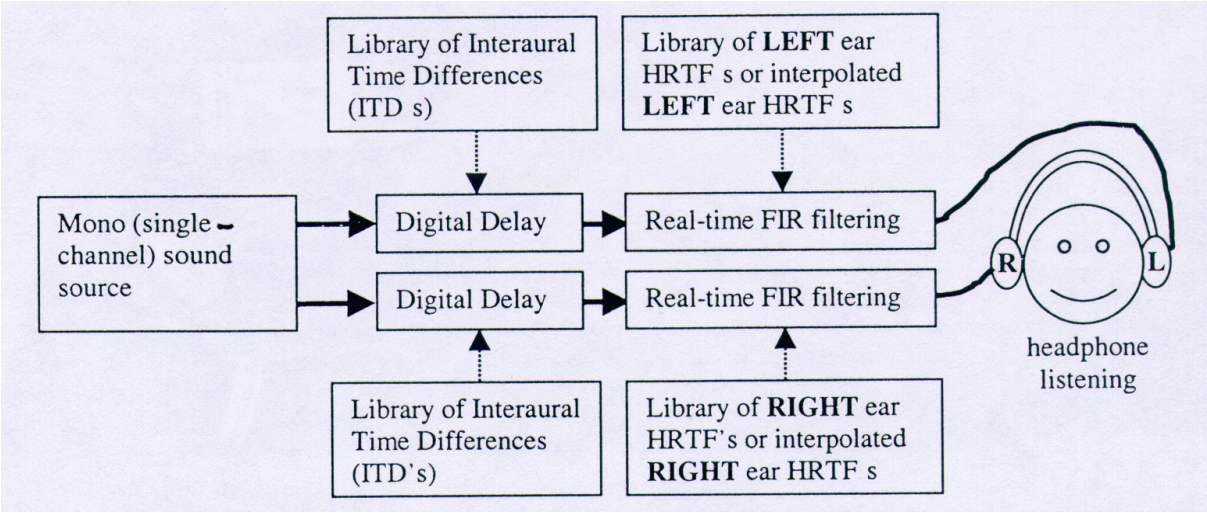


Fig.18. Virtual audio simulation using ITD and HRTF information together.

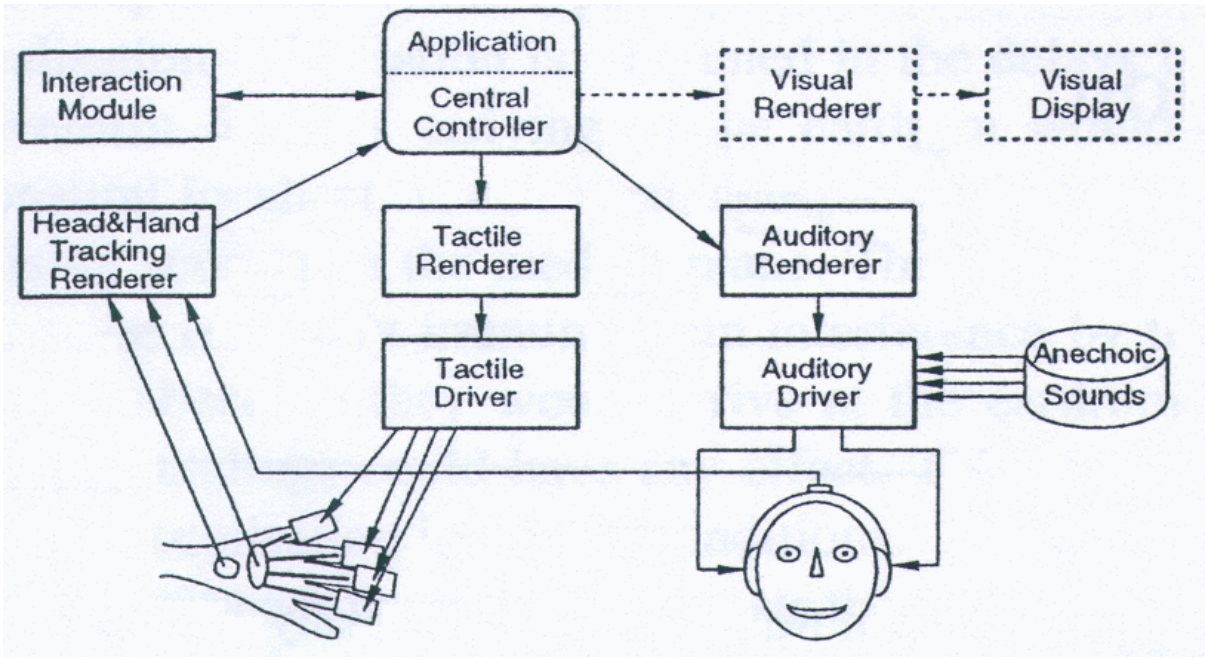


Fig.19. Architecture presented by *Blauert* for presenting complex auditory environment including head-tracking, vision and tactile/thermal modality [121].

2.5 Problems

The electrophysical analysis of the auditory system is faced with difficulties, because we do not have good ideas of the sensory features to which the auditory system is particularly responsive. The general function of the auditory cortex and higher processing algorithms are still not certain, and many hypotheses exist. A very different task, with very different implications for the function of the auditory cortex, is that for sound localization. Although, sound localization depends in part on the utilization of timing and filtering information (of the outer ears), it is not likely that the whole of the deficits obtained can be related to the temporal functions. Experiments with animals indicated that localization tasks require the presence of an intact auditory cortex and they must have a more complex basis than a simple sensory one (at the level of the outer ears) [181].

The auditory system provides an acoustical image of the world by detecting, localizing and separating external sounds through the binaural fusion and frequency analysis. Recently, instead of the linear and hierarchical model of the auditory system more complex and parallel-built models have been evaluated. The auditory system has long been assumed to function as a frequency analyser while more sophisticated functions, like acoustical recognition, have been considered the central pathway and higher processing. The traditional view requires serious revision in order to account for the auditory system's capability of encoding information under a wide range of environmental conditions. It is likely that the system uses special strategies to focus on the elements that contain meaningful components. The auditory system creates robust representations of significant information through the use of distributed, multiple representation and coding strategies.

The "place principle" of frequency coding described the neural representation of spectral content. It is becoming clear that this cannot account for all the ways in which auditory information is processed and assembled into complex representation of the external environment. Hypotheses of distributed evaluation suggest the higher processing to be active and to help already before the inner ears.

Furthermore, localization in binaural recordings made with artificial heads is inferior to real-life localization as well as to localization in binaural recordings made in the ears of selected humans. Artificial heads are still not as good for binaural recording as a well-selected human head, although some of the new artificial heads approach the performance of human heads. The question is,

whether existing binaural recording systems can be improved by adding more "accurate" HRTFs or the transmission of the directional information which fails on the playback media. This kind of recordings assumes the outer ears and the pinnae filtering, thus the SPL at the eardrums to be the one and only phenomenon to be reproduced.

2.6 Goals

The main goal is to prove that higher processing algorithms do contribute to sound localization and the perception of the acoustical environment we are in, as well as the quality of the playback situation and method we are listening to.
Results will support the need to revise and extend the physiological auditory models at the level of the outer ears as well. The statement of the binaural technique will be proved to be neither "necessary" nor "satisfactory" in its restricted form.
As mentioned above, the traditional view requires serious revision in order to account for the auditory system's capability of encoding information under a wide range of environmental conditions. We will focus especially on selected and well-defined acoustical environments near to the head based on the spectral evaluation of HRTFs and the directional information encoded within them.
The complementary results will suggest the improvement of binaural playback systems, but not through the accuracy of HRTFs. Therefore, an accurate dummy-head measurement system has been set up with lots of novel methods to increase the SNR and reproducibility in contrast to former HRTF measurements. The measured HRTFs will be evaluated in details in order to determine the effect of the environment near to the head and the effects of the HRTFs during the decoding of the acoustical information.
It will be suggested that the problems with simulated localizations or with dummy-heads are in connection with the auditory nervous system and the function of higher processing. Even with the use of individual HRTFs the headphone playback errors are present without head tracking or reverberation simulation. The extension of the auditory modeling by the peripheral effects of the higher processing will lead in the future to apply not only HRTF simulations but also other cues related to the contribution of the auditory cortex.
First, a binaural playback system will be presented with HRTF filtering and headphone playback. Then the quality and existing headphone errors will be discussed; after that the localization blur and a proposal for a virtual auditory display application will be given.

3 HRTFs in listening tests: localization blur in a 2D Virtual Audio system

3.1 Introduction

This section introduces a binaural playback system for listening tests and its capabilities. The goal was to create a virtual environment for *headphone playback* to determine the localization blur with this system, the effect of the simulated HRTFs and find whether well-known errors exist or not.
The starting point was the former GUIB (Graphical User Interface for Blind Persons) project. In this international project the researchers were searching for solutions to help elderly and disabled people to use personal computers. Blind persons do not have the advantageous properties of the commonly used graphical user interfaces (GUI) like MS-Windows, icons and the ability of orientation among multiple visual information [124]. In order to do this, visual events on the screen, like opening files, closing windows, movement of the cursor, etc. are to be replaced only by sound events. The former results of this project related to sound reproduction are:

- a collection of sounds representing visual icons and events of the screen only by acoustical information called "Earcons" based on the ideas of blind persons [125],
- the possibilities of different input media (like touch-pads, keyboard with Braille-displays etc.) – whichever is "user friendly" for a blind user [182],
- and the localization blur using a multi-channel loudspeaker playback system [112]. The surprising finding of the latest test was: blind persons cannot better localize than people with normal eyes; furthermore, loudspeaker playback is not suited for a real-life application. The so-called Sound Screen was a multi-channel array of loudspeakers behind the screen with low quality spatial resolution. It was also large and heavy, disturbed the environment as well (e.g. in an office). It was also suggested to determine the localization blur through headphone playback as well.

The test described in this section was for investigating the role of the HRTFs as well as to continue the GUIB project using the headphone playback method. For a further GUIB application with Earcons, the localization blur of different

signals have to be determined. The Earcons are short (about hundreds of milliseconds) pure tones or special noisy-like sound events. Therefore, we decided to use 300 ms long sound events of broadband and filtered version of broadband noise (signal A, B, C) to match and model in a generic but not too specific way the possible real application of Earcons in the future. Results from this test will suggest the frequency-dependent resolution of the virtual sound screen created by this system.
Furthermore, the measurement method itself suits and assists the GUIB project. It has unusual and novel methods like the 3-categorie-forced-choice in order to determine the "uncertainty" of the subjects during the localization judgments. A two-direction discrimination will be applied to determine the localization blur independent of the direction of a moving source. Instead of the commonly used method to measure the discrimination of a sound source with constant distance, a "virtual rectangle screen" is simulated (see below). On the way to specify this information we will also get the restrictions and limitations of the HRTF synthesis and headphone reproduction.
Further information about international standards for psychoacoustic measurements, definitions of accuracy and reproducibility in subjective tests are listed in [126].

3.2 Measurement method

Our virtual sound screen is a 2D square surface in the front of the listener, as mentioned above. It was selected because only less experimental results exist with non-constant source distance in the front of the listener. Usually, horizontal plane experiments are made with constant source distance around the head. Secondly, the mapping from a visual screen (PC monitor) is better to a "screen-like" 2D virtual sound screen for the orientation with the mouse (see Fig.20.). The maximal range of simulated sources is ±60° horizontal and vertical. Because the distance of the source is not constant (see Fig.21.) sources over 60° are "too far away", and the subjects make their localization judgments based only on this distance information. In addition, we assume that the listener in a real life application would be able to adjust volume, so the parameter "depth" is neglected.

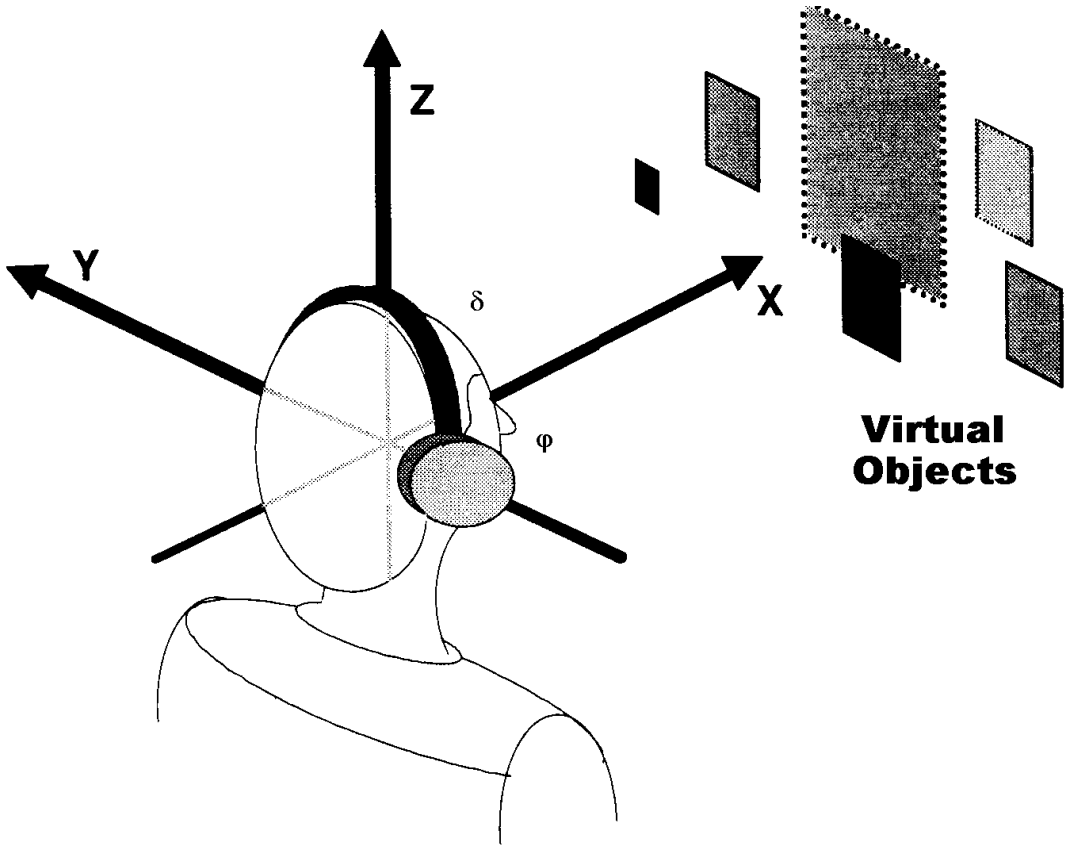


Fig.20. Illustration of virtual sources in a 2D representation [112]. The virtual acoustic surface is parallel with the Z-Y-plane. The origin is in the front of the listener: φ=δ=0°. Virtual objects move during the measurement parallel with the Y or the Z-axis in the horizontal or median plane respectively.

The measurement setup is based on a PC with the Beachtron™ DSP board. Real-time convolution of the mono input signal and the HRTFs is made in the time-domain (16 bit; 44,1 kHz). The system is precisely equalized for the circumaural, open-dynamic Sennheiser HD540 headphone. The HRTFs originate from a good localizer in a measurement of *Wightman and Kistler* [47, 111, 120]. 72 measured HRTFs are available in a form of 75-point minimum-phase-FIR-filter set in 30° spatial resolution. Duration and volume of the test signals were determined during a pre-test with 7 subjects. The main test was made with 40 untrained subjects, all with normal hearing. The individual setting of the HRTFs corresponds to measure the size of the head (distance of the ear canal entrances, see Fig.22.). Setting the ear canal distance decrease the angular error [128].

Localization depends on the signal frequency (bandwidth), duration, loudness and a-priori knowledge. To reduce the parameters we work with constant volume and duration.
Excitation signals for the MAA-measurement are 300 ms noise burst impulse-pairs: white noise (signal A), 1500 Hz low-pass (signal B) and 7000 Hz high-pass filtered version of white noise (signal C). The SPL of signal B is with 10 dB, the level of signal C is with 6 dB greater than the level of signal A (Fig.23.) for an almost constant sensation of loudness. This value is an average based on the subjects' opinion. They had to determine the SPL for signal B and C to be as loud as signal A. Broadband noise bursts must exceed 100 ms to be the subjective loudness independent from the length [129]. Stimulus frequency and duration are widely investigated in this context [130].
Novelties and general conditions in our measurement:

1. Use of a 2D virtual sound screen in the front of the listener. Sources can move only in the horizontal (left and right) and in the median plane (up and down) from the origin in 1° resolution. The source distance is not constant and the source is *not moving around the head* as usual.
2. Subjects have to report in a 3-categorie-forced-choice (MAA): "no difference between the sources", "different sound sources" and "I'm not sure". This is, because the subjects have spatial domains where they are uncertain. The size of this domain can be determined.
3. Source-pairs have to be discriminated first as the second source is moving away from the static reference source, then as it moves toward the reference point. We are looking for the nearest point to the reference, where (from both directions) the subject is able to discriminate the sources with certainty. The auditory system has deteriorative accuracy and localization performance in case of an "incoming" sound in contrast to an "outgoing" sound event. If we determine the localization blur from both direction of moving, we will get the direction-independent localization performance of the subjects.

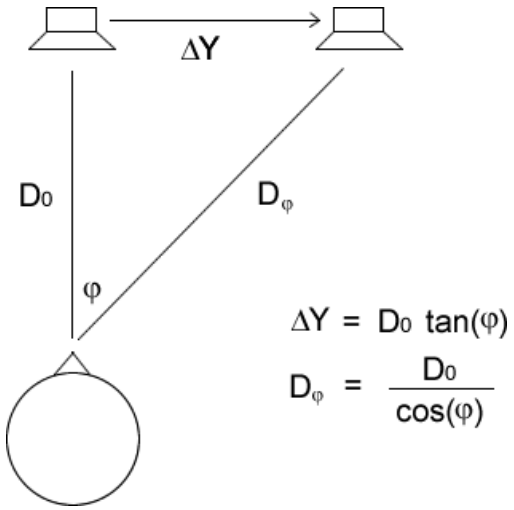


Fig.21. Source distance is 1/COS function of the azimuth angle φ by a moving sound source in the horizontal plane.

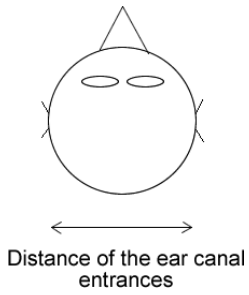


Fig.22. Setting the HRTFs: measuring the size of the head.

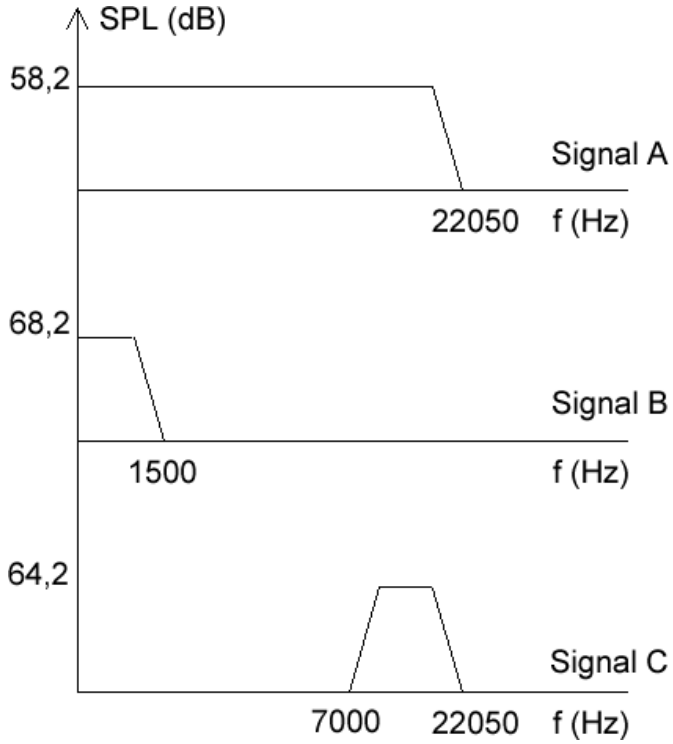


Fig.23. Spectra of the input signals. Signal A is white noise, signal B and C are derived from signal A with 1500Hz low-pass filtering and 7000 Hz high-pass filtering respectively.

The first impulse of the burst-pair is always a fixed reference point, and the second is moving first away, then toward the reference point. During the MAA measurement subjects were asked to report in a 3-categorie-forced-choice. Possible answers are: “no difference” if the subject is not able to discriminate the sources and they seem to come from the same direction. „Different sound sources“ means that he is able to distinguish between the signals. He may have the possibility to choose „uncertain“ as the answer if he is not sure which is the case.
At the beginning, the reference point is always in the origin. The second source is moving away from the reference point to the left (Fig.24.). After the subjects have reported “different sound sources” the moving source moves backward. The nearest point where the subject in both direction of moving was able to distinguish the sources will be selected as the new reference point. Maximal total number of reference sources is 13 horizontal (6 left and 6 right) and only five vertical (2 up and 2 down). The origin is always included. The duration was chosen to be 300 ms, because the Earcons are that long and signals over 250 ms are to localize the best (see Fig.10.). The pause should exceed this value for a correct separation of the burst pairs.
In [108] a similar method was used, but only in a 2-alternative forced choice as the subject’s response was used to initiate the next trial. In [101] the subjects had also to report in a forced-choice using pulse-pairs and they had the possibility to be uncertain. But this was not investigated deeply.

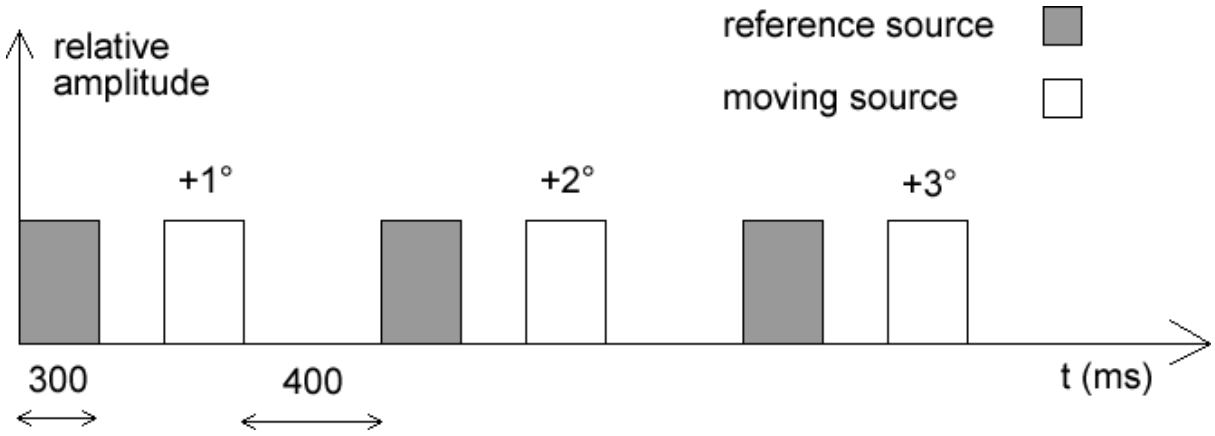


Fig.24. Presentation of the signals in a MAA measurement. The first impulse of the 300 ms signal burst pairs is the reference point and the second is moving in 1° steps away or toward the reference point.

3.3 Results

The test is divided in two parts. A **preliminary test** was made with 25 subjects in order to find well-known headphone playback errors, like in-the-head localization, elevation shift and front-back confusion using female speech signals [131].
The results of the **main test** were obtained from 40 untrained subjects. 20 male and 20 female between 23 and 50 years of age took part in this listening test under the conditions mentioned in Sec.3.2.

3.3.1 Capability and errors in headphone playback

During the preliminary test, special sound source locations were generated and specified questions had to be answered. These questions and answers may be confusing in order to find whether the subjects are easily influenced or not. After the subjects adjusted the volume of the speech signal for his "most comfortable volume", a common sound level was chosen for all subjects by averaging.
Answers from the subjects were evaluated as follows: first the static location or the movement of the source is given, which was not known by the subjects. Then the questions they had to answer and the percentage of the answers.

I. The virtual source is *in the front* in the horizontal plane.

"Where is the sound source?"	FRONT: 31%	BACK: 69%
"Reverse direction is possible?"	YES: 58%	NO: 42%

II. The virtual source is *in the back* in the horizontal plane.

"Where is the sound source?"	FRONT: 4%	BACK: 96%
"Reverse direction is possible?"	YES: 30%	NO: 70%

In both cases 80% of the subjects reported "in-the-head localization".

III. The source is *moving around the head to the left* in the horizontal plane.
"How does the sound source move?"

BEHIND THE HEAD: 43%
AROUND THE HEAD TO THE LEFT: 50%
OTHER: 7%

"Is the source moving around the head in the horizontal plane?"
YES: 67% NO: 33%

"Is the source moving around the head above/below the horizontal plane?"
YES: 58% NO: 42%

"Is the source moving this way?"  YES: 11% NO: 89%

"Is the source moving this way?"  YES: 90% NO: 10%

"Is the source moving this way?"  YES: 58% NO: 42%

IV. The source is moving *up*.

"Direction of the moving?"	UP: 92%	DOWN: 8%
"Reverse direction is possible?"	YES: 8%	NO: 92%

V. The source is moving *down*.

"Direction of the moving?"	UP: 12%	DOWN: 88%
"Reverse direction is possible?"	YES: 41%	NO: 59%

The role of the HRTFs is essential in this test, because in the median plane no interaural differences appear; only the filtering of the HRTFs delivers the directional information in case of static sound source locations (I.-II.). According to former results poor localization performance was observed:

- only 20% of the subjects were able to "externalize" the sound source and avoid in-the-head-localization. Out-of-head localization has been found to be possible, when the acoustic energy ratio of reflected sound to direct sound in a room is controlled properly or by controlling special networks connected to the earphones [40, 132].
- Front and back directions were mostly confused as the source was in the front (69%); only one third was able to localize the source at its correct position. If the source was in the back, only 4% answered false. The uncertainty was also greater in the front: 58% believed the reverse direction as well - independent from their answers. This indicates a

well-known error: sources in the median plane often are localized in the back hemisphere during headphone playback.

We have to point out that we did not measure the *rate* of the front-back reversals, only the relative number of the subjects is indicated (69%) who made a reversal error during a single measurement in this test.

During the third section, the source was always moving around the head in the horizontal plane, only the questions were different in order to see if the subjects could be influenced (III.).

- Many of the subjects reported the source moving only behind the head (in the back hemisphere). Only 50% was able to detect the correct movement.
- 58% believed that the turning-movement is *above* the horizontal plane. This so called elevation shift is also a well-known error during headphone playback.
- Only 11% was able to detect the source moving in the frontal hemisphere.
- It is interesting that for 58% the movement was acceptable as a moving source in the frontal plane.

Changing of the elevation was surprisingly easy to recognize (IV.-V.). Some authors reported decreased performance from the lower hemisphere. Our results do not support this finding relevant: 92% and 88% were able to detect the correct direction. Only the uncertainty was greater “down”.

This test proved that even a carefully made headphone-equalization, the use of HRTFs of a good localizer and individual setting of the size of the head are maybe insufficient. The subjects were confused and undecided in case of static median plane sources and by moving sources symmetrical to the median plane.

All of the subjects were easily influenced and they reported all kinds of answers by the same signal reproduction, which suggests low quality localization in the median plane. The directional judgments of the subjects show that well-known errors of the headphone playback are present. In-the-head localization and front-back confusion are more significant than elevation shifts. Their eyes and will can influence the subjects very easily. This phenomenon is independent from the signal processing and suggests alternative headphone design and basic problems with headphone playback systems [137]. Possible solutions in headphones design for decreasing the in-the-head localization, elevation shift and front-back errors could be a physical displacement of the loudspeaker in the headphone as described in [40, 122, 138-141] or direct concha excitation, where the trancducers are angled from the front to face the concha area resulting in four-times better localization than conventional headphones [115].

3.3.2 Localisation blur and discrimination skills

The main test includes listening tests using noise impulse pairs in the horizontal and in the median plane in order to determine the localization blur. 40 untrained subjects all with normal hearing participated in the test, and the results are presented below showing average (AVG), maximal (MAX) and minimal (MIN) values of measured data.
The test was carried out in the anechoic chamber to avoid external environmental disturbances. Subjects were sitting on a comfortable chair with a signal button in the hands. During the 10 minutes of accommodation time the distance of the ear canal entrances (size of the head) was measured, a detailed explanation of the procedure was given and a trial run was made in one direction. Maximal, minimal and averaged values of the measured head diameter and age of the subjects are shown in Table 2 and 3 respectively.

Ear canal distance [cm]	AVG	MAX	MIN
Male	13,6	15,2	12,0
Female	12,4	13,3	10,5

Table 2. AVG, MIN and MAX values of the measured distance between the ear canal entrances. Total average over every subject is 13 cm.

Age [years]	AVG	MAX	MIN
Male	28,3	39	21
Female	27,7	39	22

Table 3. AVG, MIN and MAX values of the ages of the subjects. Total average over every subjects is 28 years.

First, signal A was presented in the directions „down”, „up”, „left” and „right”. After a few minutes break we continued with signal B and signal C. Overall time for the test was about 60 minutes (15 minutes for each test signal on average).
At the end, subjects had to fill out a questionnaire about personal data (sex, age), computer skills (59% „professional or engineer“; 41% „everyday user“) and headphone user routine (7% „everyday”; 24% „often”; 59% „seldom”; 10% „never”).

Results were found to be independent of age and computer skills, but little improvement in the localization performance was found by subjects using headphones often. The spatial resolution is as good as independent of gender.
Tables 4 to 7 (see Appendix F) show the data in degrees broken down into sub-tables of signal, direction and gender. Figure 25 shows the average values from the median plane for all signals up and down as well. In *vertical* directions no significant differences appear between female and male subjects. The average resolution for signal A is about 15-17°, 19-24° for signal B and 18-23° for signal C. The maximum values can reach the double of the average value; the minimum values could be 10-50% of the mean value.
In the *horizontal* plane signal A is localized the best with an average resolution of 7-9°, signal B with 9-11° and signal C with 8-10° (Fig.26.). In general we can support the finding that broadband sources are localized the best as well as signals with lots of high-frequency information, but the differences in our measurements are relatively low: the results of signal A are only 1-2° better than results of signal C.
It is interesting that the resolution (the difference between nearby source locations) is almost constant. Small differences between the averaged values of new reference points over 50° are due to the large min-max-domain: some were able to locate the last source at 35° and some only at 70°. Minimum values of 3-4° and maximum values of 20-24° were measured in the horizontal plane (Fig. 27 to 30).
The possible source locations are shown on Fig.29. in the median plane and horizontal plane respectively (on average).
Figure 30 shows all the *individual* results for signal A on the left side. The colors refer to data for each new source location from 40 subjects. Note that only the first four new reference points could be determined by all subjects: some were not able to detect 6 within the 60° domain (see „missing locations" below). That is the reason for less data for the fifth and sixth reference point (filled orange and brown).
Our data are comparable with other results from the literature. Table 8 and Table 9 contain comparative results from the median and the horizontal plane achieved by headphone playback under the given conditions and signals.

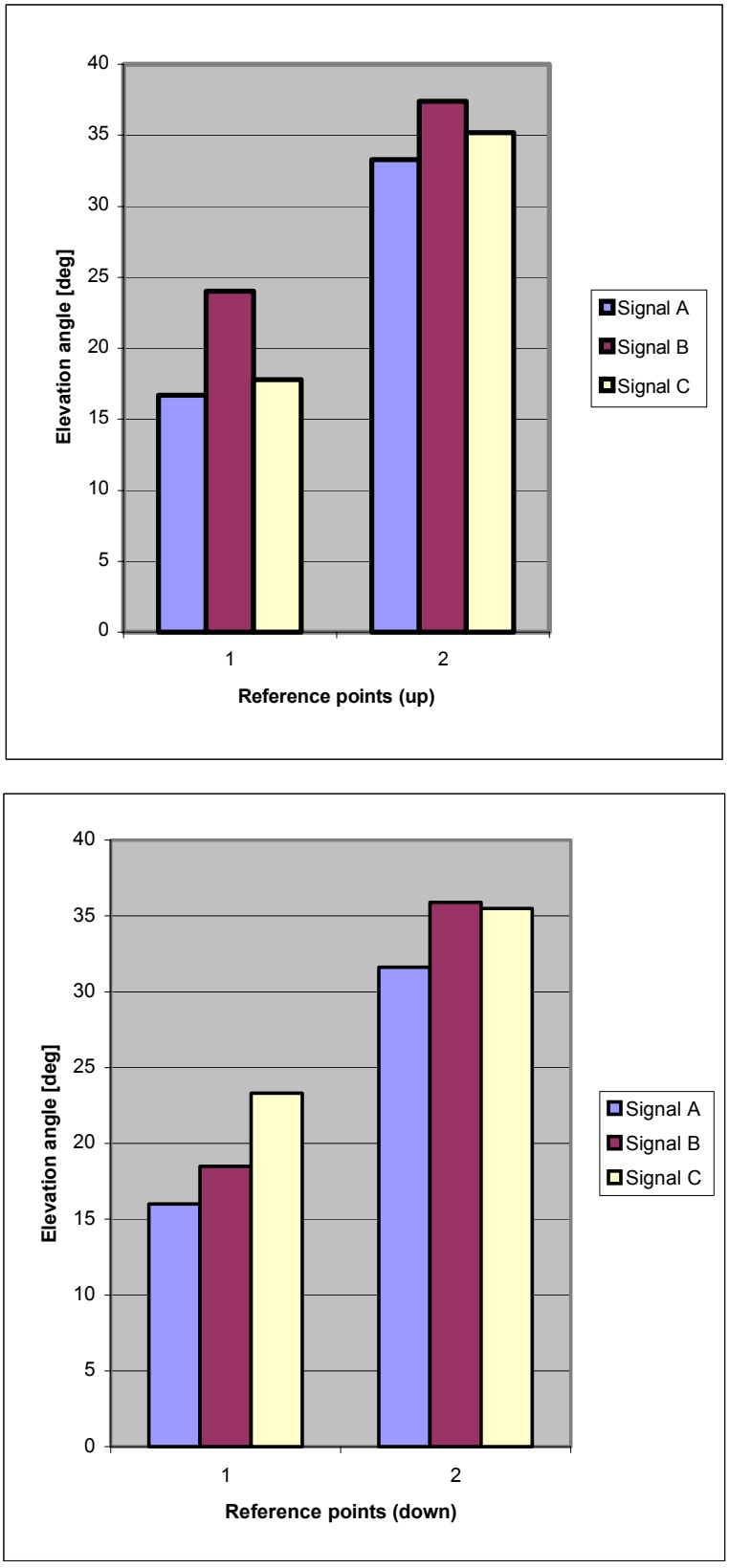


Fig.25. AVG values from the median plane for all signals.

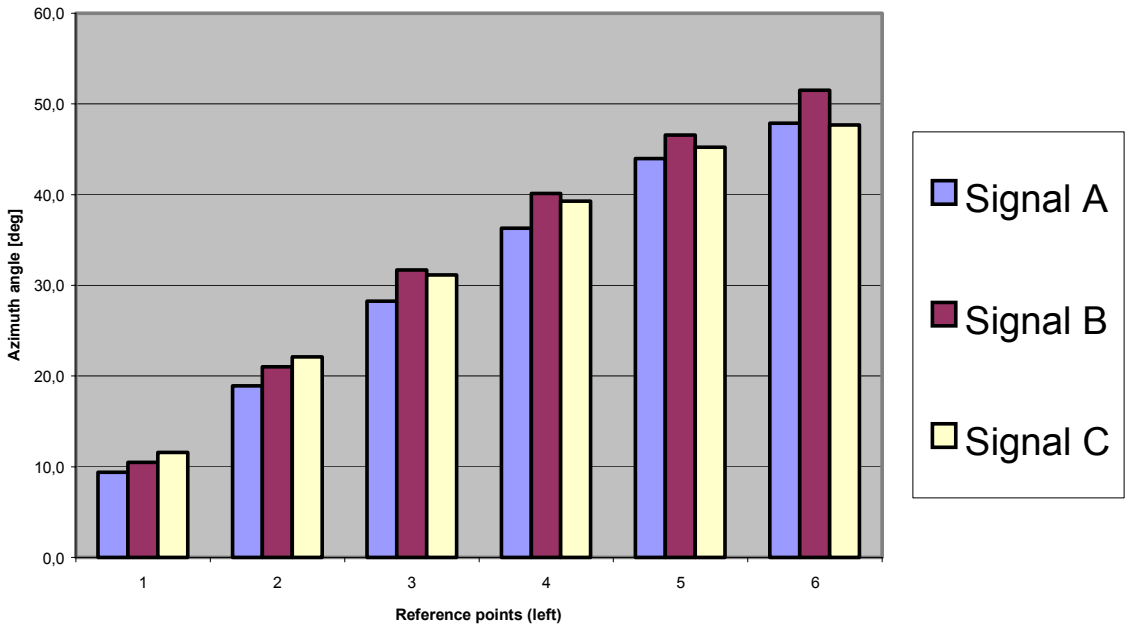


Fig.26. Localization of signals with different spectra (AVG values, left side).

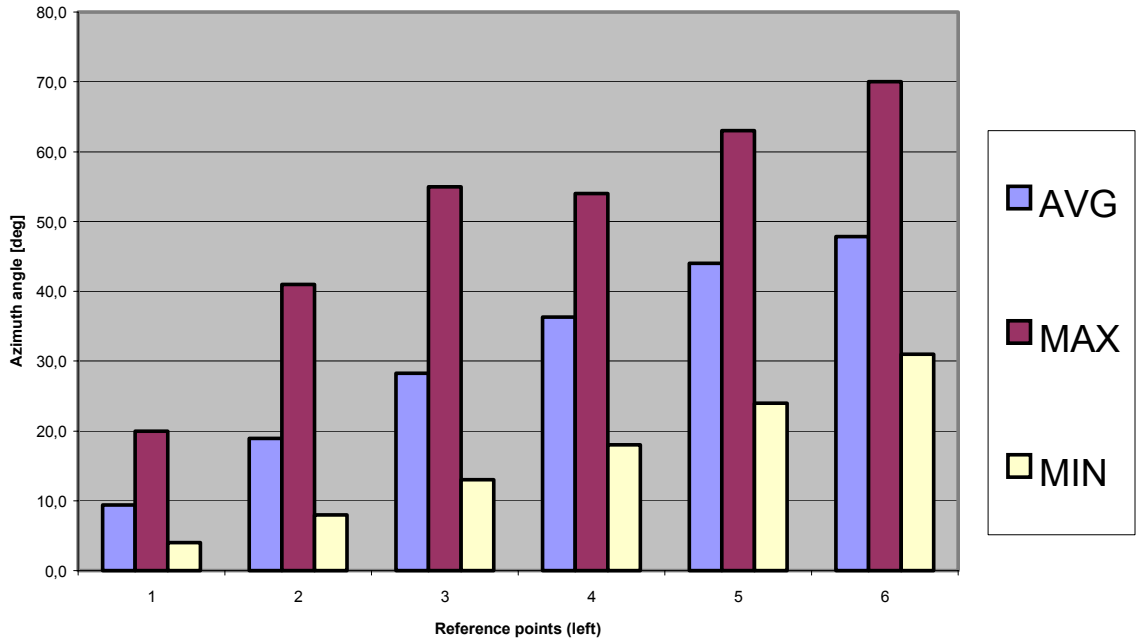


Fig.27. MAX, MIN and AVG values for new reference points (signal A, left).

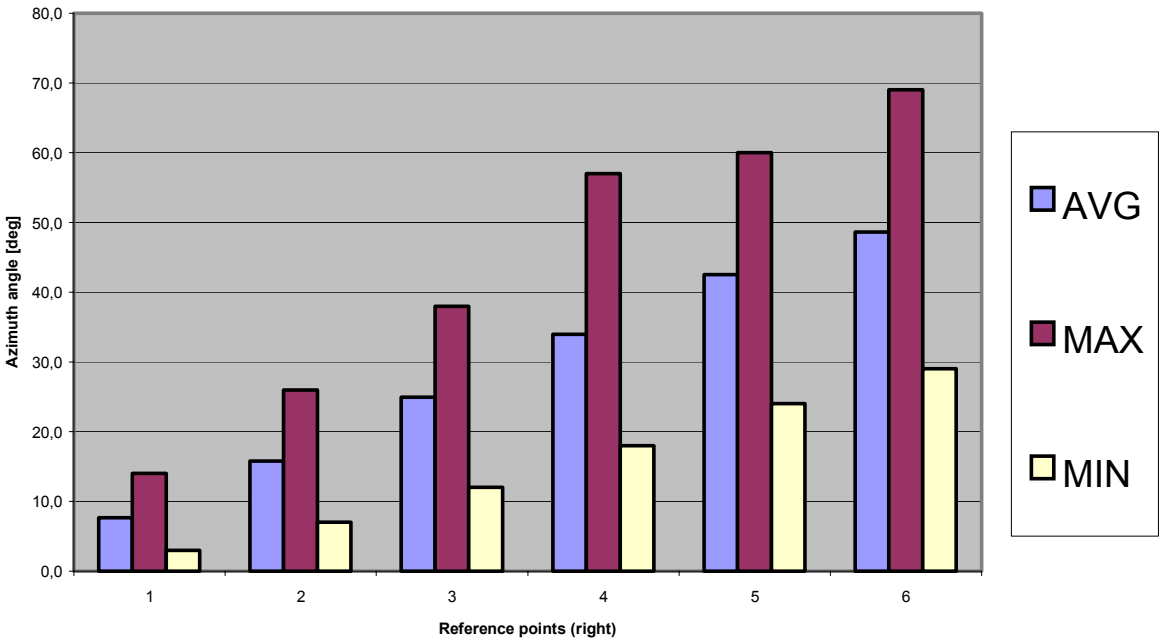


Fig.28. MAX, MIN and AVG values for new reference points (signal A, right).

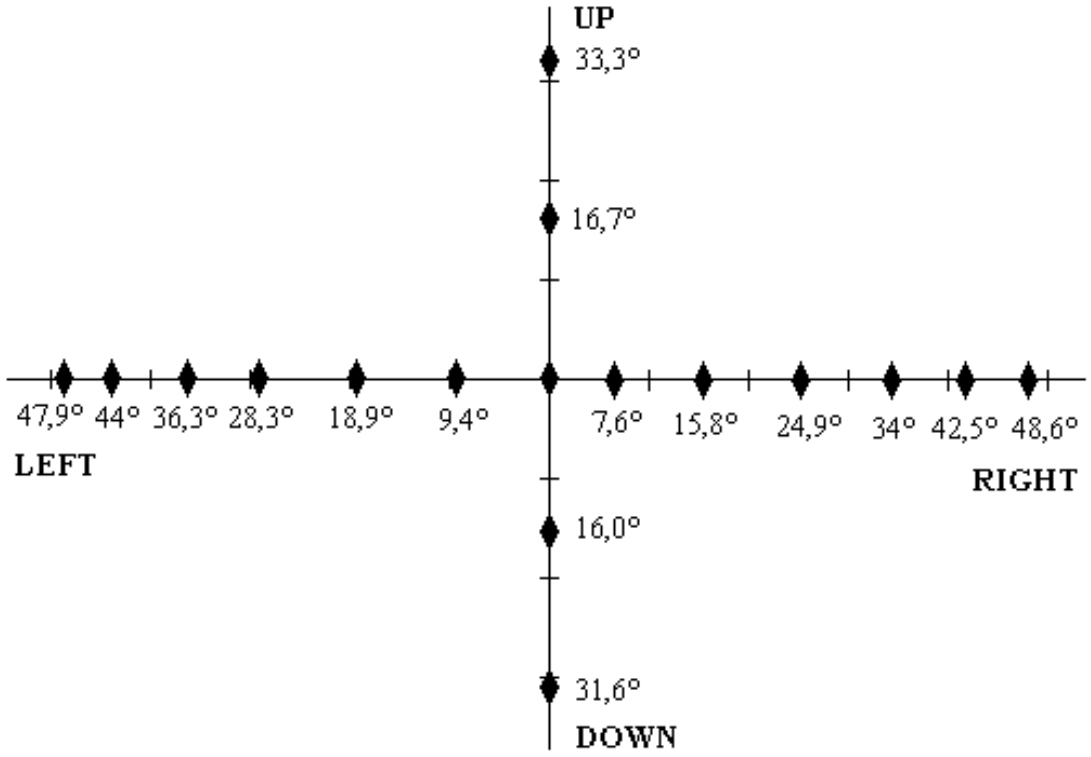


Fig.29. AVG values as possible source locations for signal A based on Table 4 and Table 5.

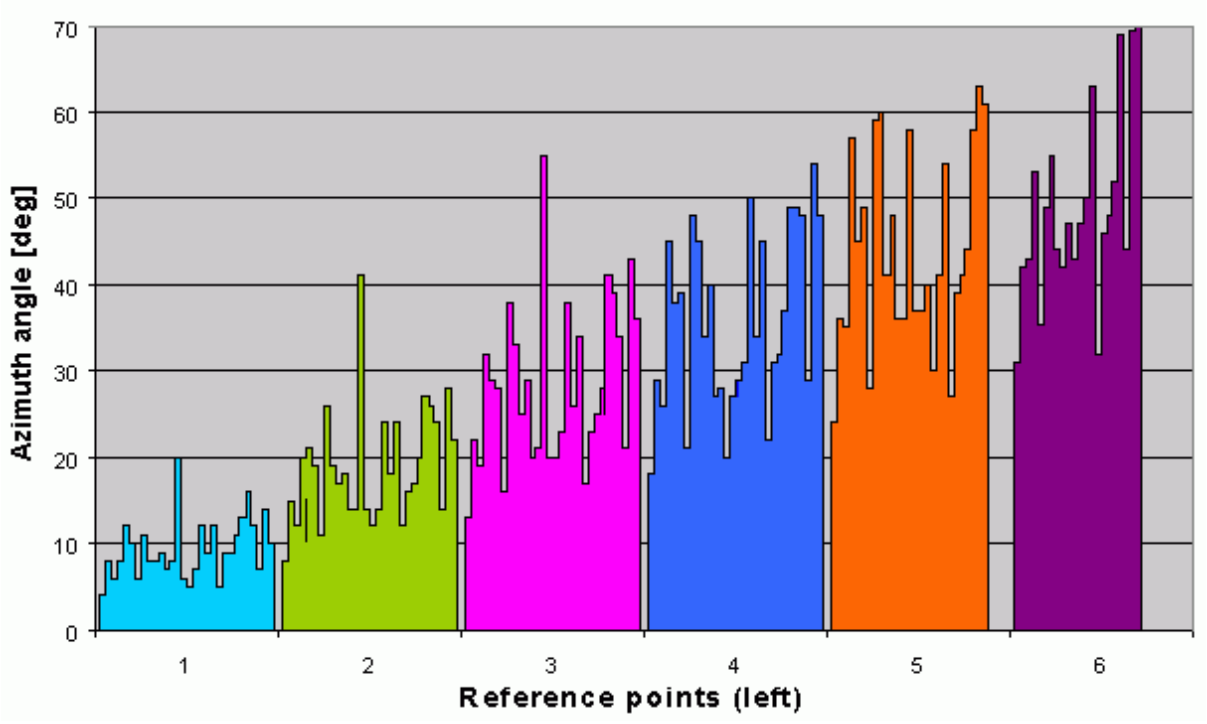


Fig.30. Individual results of all subjects for new reference points (signal A, left). Note that only four reference points could be determined by all subjects.

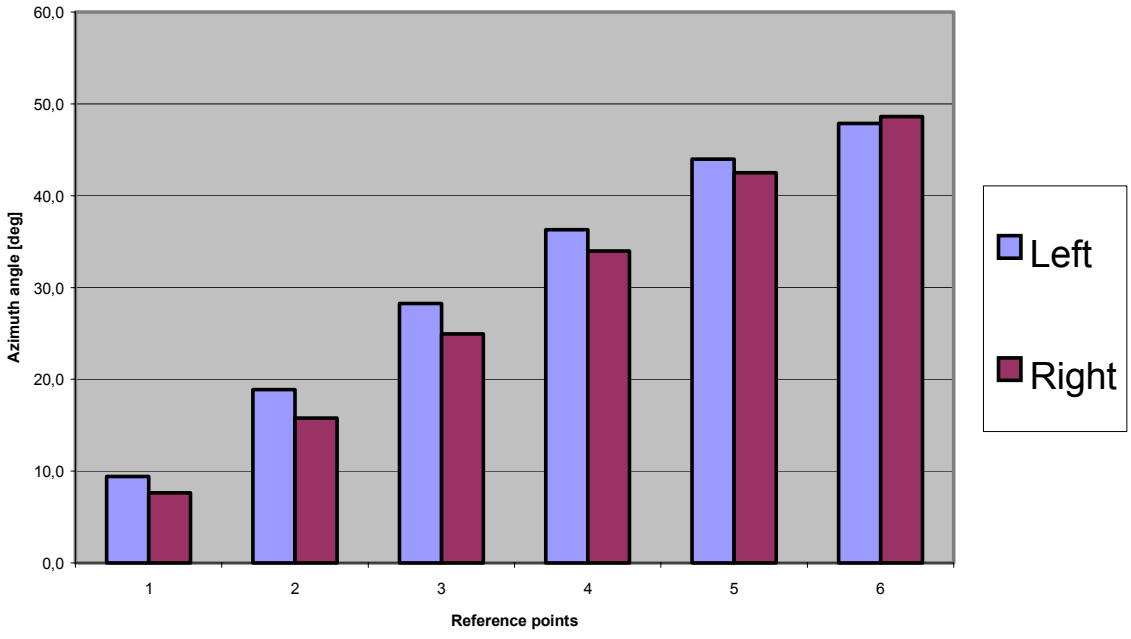


Fig.31. AVG values from the left and right side for new reference points (signal A). Right-handed subjects show systematically lower resolution on the left side.

AUTHOR	SIGNAL, REMARKS	RESULTS
Oldfield, Parker [135, 172]	azimuthal mean value	9°
	azimuth errors with HRTF filtering	4°-6°
	azimuth errors without HRTF filtering	11,9°
Wersényi	MAA values, 300 ms broadband noise, non-individual HRTFs of a good localizer	7°-10°
McKinley, Ericson [174]	average error, MAA value	5°
Middlebrooks [118]	average error, non-individual HRTFs (other-ear-condition) average error, individual HRTFs (own-ear-condition)	17,1° 14,7°
Duda [175]	average error with human HRTFs average error for broadband signals (12kHz)	4,5° 3,4°
Gardner [179]	average angle error, pink noise bursts of 250 ms	14,3°
Begault, Wenzel [180]	average error (generic HRTF) average error (individual HRTF)	21,7°-23° 20°
Martin [84]	average error, 328 ms noise signal maximal error	9,6°-9,7° 13,1°

Table 8. Comparative localization results in the horizontal plane using headphones.

AUTHOR	SIGNAL, REMARKS	RESULTS
Oldfield, Parker [135, 172]	elevational mean value	12°
	elevational error with HRTF filtering	6°-8°
	elevational error without HRTF filtering	21,9°
Wenzel, Foster [165]	non-individual HRTFs, 16 subjects lower elevations, front lower elevations, side	 ca. 24° ca. 23° (side)
Wightman, Kistler [120]	average error lower elevations, front lower elevations, side	 ca. 21° ca. 20°
McKinley, Ericson [174]	MAA value, dummy-head HRTF	30°-35°
Wersényi	MAA values, 300 ms broadband noise, non-individual HRTFs of a good localizer	15°-24°
Duda [175]	average error with human HRTFs average error for broadband signals (12kHz)	19,2° 17,2°
Gardner [179]	average angle error, pink noise bursts of 250 ms	34,2°
Begault, Wenzel [180]	average error (individual HRTF)	17-19°

Table 9. Comparative localization results in the median plane using headphones.

3.3.3 Localization judgements

Left-right and up-down symmetry

Other studies reported asymmetries on the left and right sides of the hearing system in connection with right or left-handed persons [35, 142]. *Burke* measured less front-back errors on the left side using broadband stimulus from an array of loudspeakers with 20 left -and 20 right-handed subjects [143]. The right hemisphere of the brain is perhaps responsible for the interaural evaluation and the left for the monaural. *Abel et al.* investigated left/right asymmetries in sixteen (14 right-handed) normal-hearing young adults in a semi-reverberant room using 300 ms broadband noise stimulus from an array of loudspeakers. Averaged percentages of image reversals tended to increase on the right side of the space, but in contrast, error patterns on the left and right side were fairly symmetric for azimuths close to the interaural axis. The peak ILD was found 10-15 dB greater for sounds on the right horizon than on the left [95].
Our results also showed systematic asymmetry but we had only right-handed subjects. Further measurements are suggested to find regularities on this field.
Sources from the left side are typically harder to localize. The results show 2-4° average differences that correspond to a difference of 20-40% (Fig.31.)!
Table 10 shows how many subjects had decreased localization performance. By signal A 67% of all subjects had decreased resolution on the left side and only 6% on the right side. The difference is much greater for signal B.
Every second subject showed decreased resolution on the left side for *all signals*. Only one subject delivered better results on the right side for all signals.

Decreased performance [%]	signal A	signal B	signal C
To the left	67	75	57
No difference	27	21	30
To the right	6	4	13

Table 10. Typically decreased localization performance on the left side. All subjects were right-handed.

Vertical localization

In the median plane the localization is only made based on the HRTFs because no interaural differences are present. This results in a decreased localization performance in contrast to horizontal plane localization.
This fact is supported by our results as well: first, the spatial resolution is poorer, second, some were not able at all to localize the sources. 67% of the subjects reported correct localization. But 33% made the MAA-judgments only why the impulses „sound different“ (based on the spectral distortions of the applied HRTFs) but they could not really localize the sources. The same observed *Mills*: subjects reported that the difference between the stimuli seemed to be in the loudness or quality of the sound rather than its location [51].
Tan et al. reported asymmetry in the vertical localization: higher elevations can be perceived more frequently than lower elevations (below the horizontal plane), all the subjects had up-down reversals and only very high and low positions could be separated [115]. Table 11 shows the symmetry in our measurement between the directions up and down. For all signals the resolution by lower elevations is poorer. By signal A 29% of the subjects had decreased performance „down” and only 7% „up”. The most significant difference appears for signal C.

Decreased performance [%]	signal A	signal B	signal C
Down	29	32	57
No difference	64	42	38
Up	7	26	5

Table 11. Results of up-down comparison show more symmetry than Table 10. Signal C seems to be localized significantly better up than down.

Vertical localization [%]	
YES	67
NO	33

Table 12. One third of all subjects were not able to localize the sound source in the median plane, they made the MAA-judgments based on spectral distortions only.

Missing locations

Subjects had to discriminate new source locations (reference points) within a domain of ±60° left/right and up/down from the origin. The number of possible source locations is limited: maximal 6 in the horizontal plane and maximal 2 in the median plane. Subjects, who can not determine so many different source locations, have poor localization performance ("missing locations").
In the *horizontal* plane 50% of the females and only 45% of males could discriminate 6 sources for all signals. In *vertical* directions 70% of females and 62% of males were able to detect 2 new sources. This shows a bit poorer performance of males. The best performance was observed by signal A, followed by signal C and signal B the last.

Uncertainty in discrimination skills

The subjects reported in a 3-categorie-forced-choice, so they determined a domain in which they were uncertain. By some of the subjects this domain is quite large: by 57% it reached 3-5° or more independent from the signal. 43% of all subjects reported only "different sources" and "no difference". The uncertainty is by them 1°, maximal 2°.
It was expected that subjects have better resolution if the second source is moving toward the reference point, because the distance at the start is large and then decreased. But this was not common at all, some subjects could better localize as the moving source was moving away from the reference point.
Another interesting observation was that some subjects still believed to perceive a difference by a moving source backward on the left side even if it was already on the right side. They put the mark „no difference" as the moving source was over the reference point about 1-2°, but there were values of 5-8° observed as well.

3.4 Summary

Minimum Audible Angle measurements were made in order to determine the localization blur for signals with different spectral contain. 40 untrained subjects reported in a 3-categorie forced choice using headphone playback and synthesized HRTFs. The goal was to determine how many virtual sound sources

can be placed in the horizontal and in the median plane respectively and in which spatial resolution:

- localization is poorer in the median plane than in the horizontal plane,
- the lack of individual HRTFs and head movements cause in-the-head localization, front-back reversals and elevation shift. (The last is not very significant in our measurement.)
- in the median plane, one third of the subjects could not localize the sources at all, and there is an increased number of "missing locations".
- Source movements symmetrical to the median plane are confusing and hard to perceive, sources are often localized only in the back hemisphere
- Age, sex and computer skills do not influence the localization, but subjects wearing often headphones delivered better results
- Broadband signals are to localize the best, followed by high frequency stimulus and low frequency tones at last.
- The hearing system is not symmetrical: different resolution can be measured on the left and the right side as well as up and down.

The 2D virtual acoustic display is suited for replacing the screen and visual information for blind and elderly people in case of proper mapping between acoustic and visual information, so these results can be the basis for further GUIB applications and investigations.

Average resolution of 7-11° and 15-24° were measured in the horizontal plane and median plane respectively dependent on the spectral content of the signals. White noise is to localize the best, low frequency filtered noise the least. It is also suggested for a GUIB application to use broadband noisy like sound events and/or tones with more high frequency content. Earcons are already available based on the decisions of blind people. The localization depends neither on the age nor size of the head (ear canal distance) nor the computer user's routine in this set of subjects. Females seem to be as good at localizing as male subjects.

Subjects who wear headphones often deliver better results. People who were not able to detect 2 sources vertical and/or 6 horizontal only seldom or never wear headphones.

Based on these results, for a GUIB-based simulation it is recommended

- not to use vertical displacement of simulated objects, because one third of the users are not able at all to localize virtual sound sources in the median plane. One possible solution could be timbre or pitch modulation based on psychoacoustic observations: signals having higher frequency components are „above"; signals with lower frequency elements are „below",

- to partitioning in the horizontal plane for maximal 9 source position in a resolution of 10 degrees.

On the other hand, the preliminary test showed that low-cost real-time system with many efforts to a correct binaural reproduction have all kinds of headphone playback errors. This assumes that the problem of insufficient localization is not due the „quality", fine structure or overall accuracy of the HRTFs. This will be investigated next.

To find out more about the role of the HRTFs and their fine structure, we need an accurate, precisely controllable measurement system for measuring the HRTFs and the differences. The human auditory system does not seem to utilize all the information included in the HRTFs, only the easily recognizable, significant information [96]. What is the role of the HRTFs in the decoding cue of the directional information of a sound source?

4 Measurement of dummy-head HRTFs

4.1 Introduction

There are many methods to measure HRTFs (see Appendix B). In the time-domain: impulse response followed by an FFT, which can be made fast (with real human subjects) but only with limited SNR and spatial resolution. Frequency-domain methods need broadband stimuli, e.g. noise excitations. They have increased SNR due to averaging and a long measurement time. For reproducible measurements we need an objective and accurate measurement system. Head and Torso Simulators (HATs, dummy-heads) are suited for long-time measurements and they try to model the "average human head".
Former measurements were made at the Békésy György Acoustic Research Laboratory at the Technical University of Budapest in order to identify the minimal-phase property of the HRTFs of a dummy-head in connection with the visual capability of the eyes [45, 144, 145]. This investigation also showed parts of the system, which were inappropriate (horizontal resolution of 5°, reproducibility, SNR). We updated and re-installed a full automatic, computer-controlled measurement system for measuring huge amounts of HRTF data using novel methods to increase the precision and SNR.
This section introduces new methods, like generating a pseudo-random noise excitation (algorithm), methods to decrease the disturbance of the 220 V mains periodicity, effects of averaging, test measurements and settings of azimuth and elevation with a computer controlled turntable and a laser targeting system.
It is necessary to reach this kind of accuracy, because we will later work only with the *differences* between measured HRTFs. So most of the individual properties and differences as well as the "undesired transfer functions" in the measurement chain will be eliminated [46, 74].

4.2 The measurement setup

4.2.1 General parameters

To investigate the small differences and changes in a HRTF database, we definitely need an objective system, which allows precise and reproducible measurements. Our goal was to install a system with increased SNR compared to the general used systems by keeping a good spatial resolution [8, 9, 45, 61, 76, 146].
The HRTFs were measured using a Brüel&Kjaer Head and Torso Simulator Type 4128 placed on a turntable in the 125 m^3 anechoic room. The elevation of the loudspeaker is adjustable by strings from -45° up to +90°. From former psychoacoustic investigations we decided to use a spatial resolution of 1° in the horizontal plane, and 5° in the elevation according to the best resolution of the auditory system. Pseudo-random broadband noise signal is used as stimulus, and two channel responses are collected and averaged in a reference measurement.
The measuring software controls the turntable, delivers the stimulus from the DSP card, and stores the responses of both ears simultaneously with 50 kHz sampling frequency, 16 bit resolution and 4096-point FFT. Eq.7. shows the linear spectral resolution of the measured transfer functions based on the sample frequency and points of FFT. The DSP card is an AT&T Ariel mainboard with a DSP32C processor. Brüel&Kjaer 2706 power amplifier, and 2636 measuring amplifiers are used (Fig.32.).

$$RES_{FFT} = \frac{50000Hz}{4096} = 12{,}2Hz \qquad (7)$$

For the proper setting of source elevation a laser targeting system is used and for increasing the SNR a robust averaging procedure was built based on reducing the high voltage mains effect (see below).
The most important property of the loudspeaker is the deviation of its transfer characteristics according to its reference axis. This was measured using a Brüel&Kjaer microphone type 4166. By lateral movements of 6° of the microphone no significant errors occurred in the measurements (Fig.33.). The same kind of fluctuating transfer function of the loudspeaker was measured and found to be proper for reference measurements in [120]. The effects of the undesired transfer characteristics in the measurement chain were eliminated by the reference signal and by calculating the HRTFs as usual:

$$HRTF(j\omega) = \frac{H_{outerears}(j\omega)}{H_{reference}(j\omega)} \qquad (8)$$

The reference signal was measured with the same Brüel&Kjaer 4166 microphone (Fig.34.). The validity of the HRTFs is above 200 Hz.

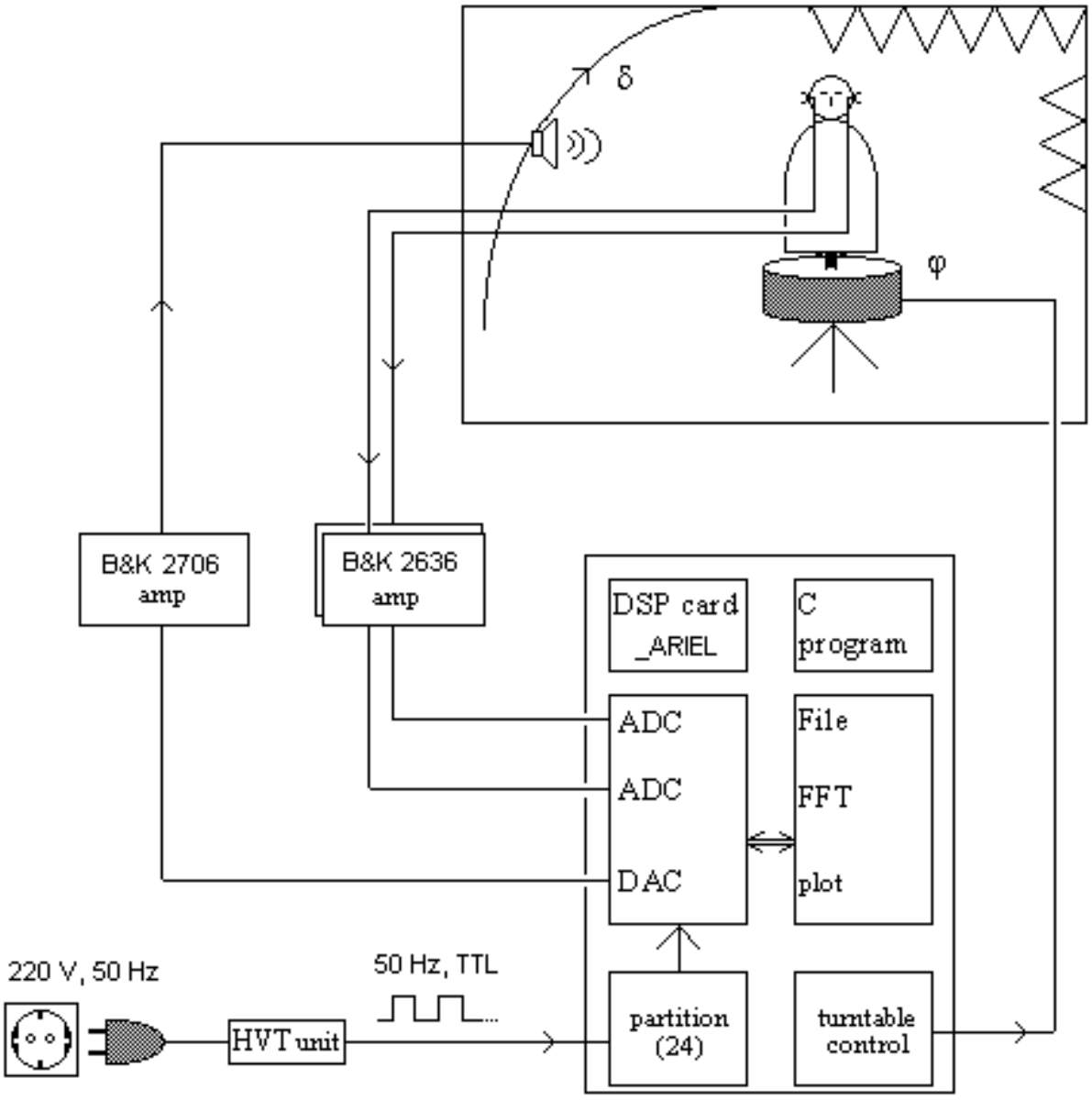


Fig.32. The measurement setup

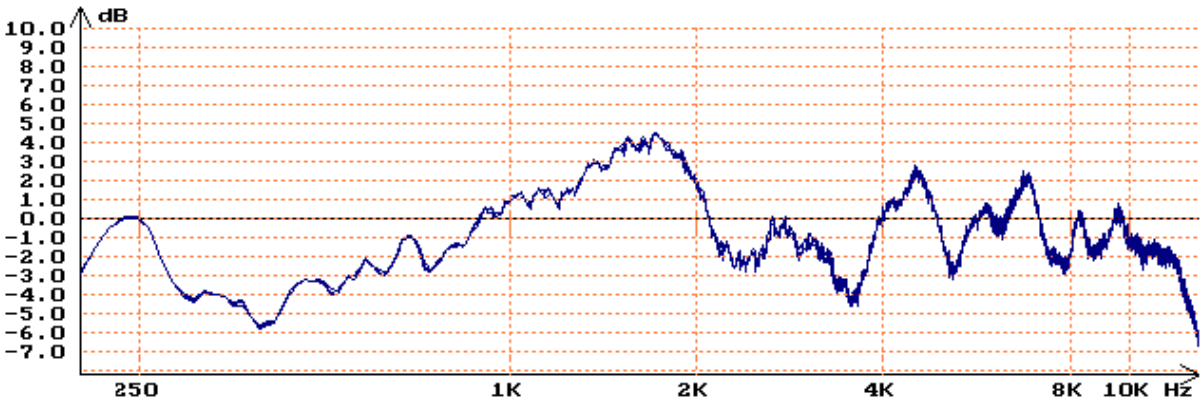


Fig.33. Transfer characteristics of the loudspeaker within ±6 degrees from the reference axis measured with a linear, unidirectional microphone type BK 4166.

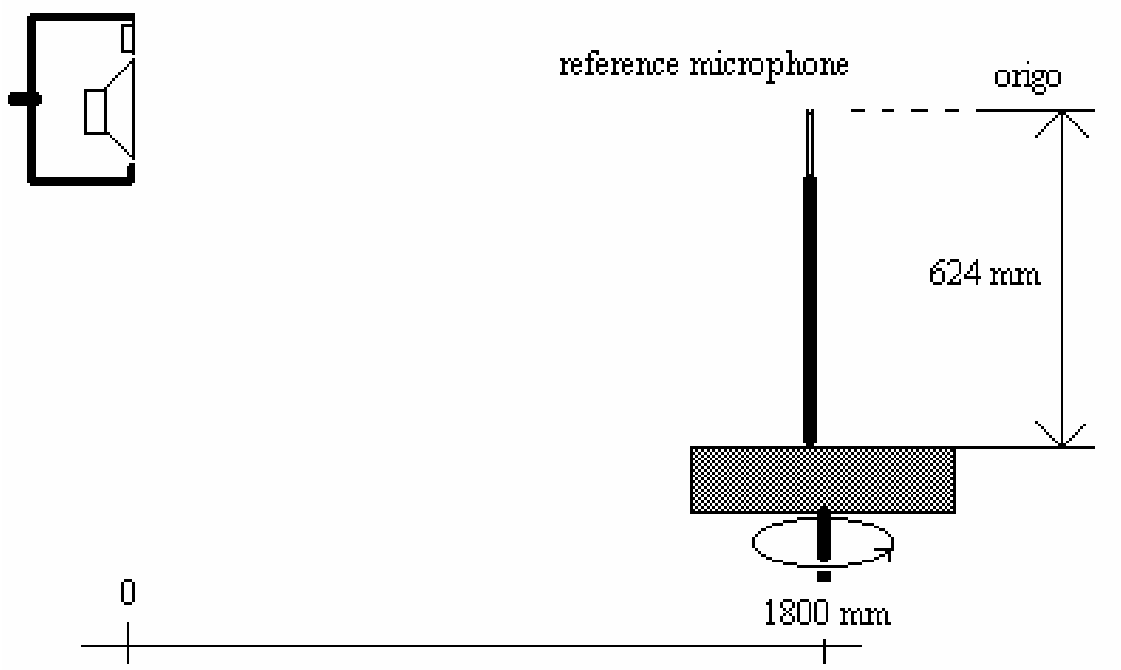


Fig.34. Setup for measuring the reference signal, the transfer characteristics of the loudspeaker and for the impulse response.

4.2.2 Setting of the elevation

The elevation of the source is adjustable by strings in $\Delta\delta=5^{o}$ steps between –45 to +90 degrees. The first task is the precise setting of the sound source elevation. A small mirror is placed on the box of the loudspeaker and a laser targeting system (LTS) onto the head of the torso. This is a red laser beam focused on the mirror. The elevation of the beam (δ') can be set with a precision of 5" which results in a 0,77% precision of setting the sound source elevation assuming $\Delta\delta=5^{o}$ resolution.
The setting procedure is based on geometrical calculations of the reflected beam (Fig.35.). As the elevation of the loudspeaker changes, both the elevation of the beam and the distance of the reflected spot vary. The measuring software calculates the data needed for the setting. The inputs are the desired source elevation and the source distance, and the outputs are the elevation of the LTS-beam and the distance of the reflected spot measured from the origin. The beam elevation is linear; the distance is non-linear function of the source elevation. The LTS has to be calibrated carefully and it is removed from the head after setting the elevation (possible error of the uncalibrated LTS without calibration would be 3,1 % elevational and 10% azimuthal).
Similar method was used by *Gardner*. KEMAR HRTFs were measured using the MLS impulse response method, motorized turntable and ray projection from the center of the KEMAR face to set elevational positions in five degree steps, but only with a SNR of 65 dB [179].
Fig.36. shows HRTFs measured in 45° azimuthal steps (left ear) at source elevation $\delta=90°$ (above). No significant difference appears, because turning of the torso should not influence the results in this case. Fig.37. shows comparative result between our measurement (dotted line) and the original HRTF data of the torso as given in the instruction manual (solid line).

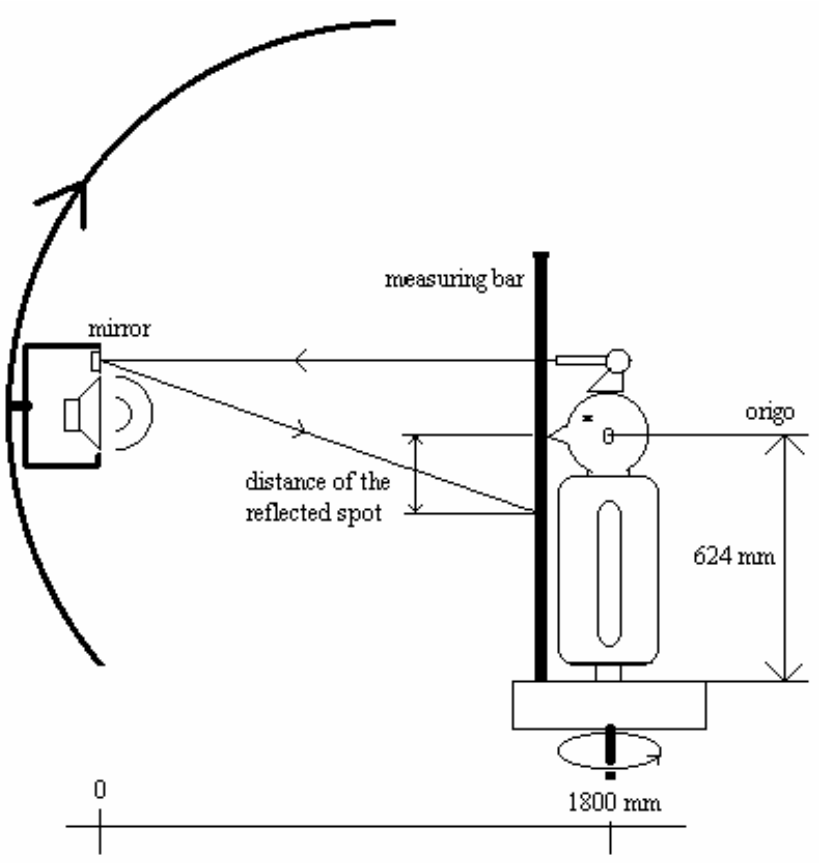


Fig.35. Setting the sound source elevation.

Fig.36. Testing the elevational settings. Set of HRTFs measured in 45° azimuthal steps (9 HRTFs, left ear).
Source elevation is δ=90°. There is no significant difference if the sound source is above the head, supporting the theory.

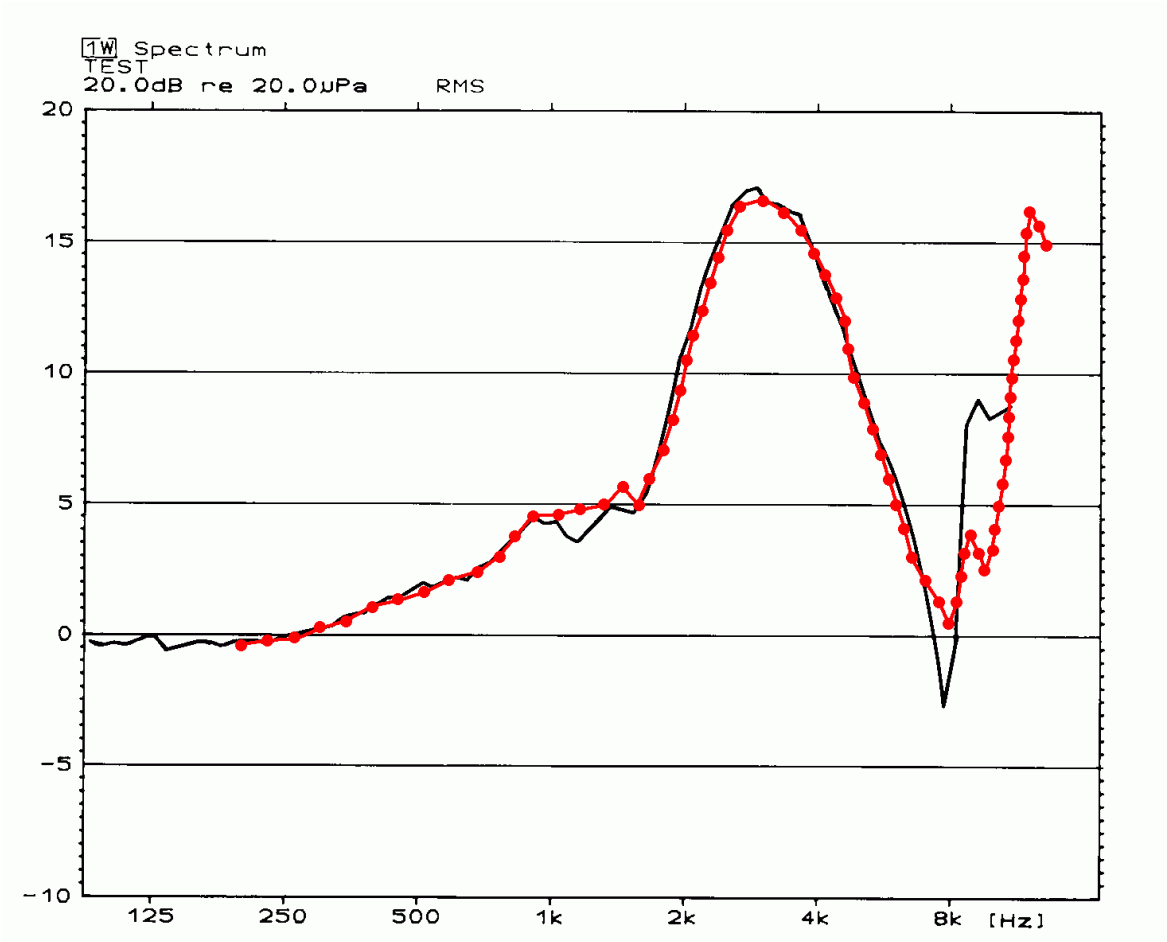


Fig.37. Comparative figures of the dummy-head HRTFs at frontal incidence. Solid black line shows the HRTF given in the Instructions Manual, dotted red line is the HRTF measured in the present investigation.

4.2.3 Setting of the azimuth

The goal is to control the turntable in 1° steps. The original built-in motor was replaced in a Brüel&Kjaer turntable with a stepping motor. This can be controlled more precisely. The motor makes 32000 steps until the turntable makes 360°. This is 88,88 steps/degree on average. The turntable is synchronized to the φ=0° azimuth.

Corresponding to the number of the steps of the motor, any arbitrary position can be set with a precision of 1/32000. The 1° azimuthal positions can be set with an average precision of 1/88,88. This corresponds to a satisfactory relative value of 1,14%. We found that this precision is necessary to find "hidden effects" in the HRTFs, e.g. the pinnae reflections at 11 kHz between 60-90º in the horizontal plane (see below). The measurement software controls the

stepping motor in the turntable. A similar method for comparison of settings using a laser pointer is in [95]. Fig.38 shows three HRTFs from φ=359°, 0° and 1° in the horizontal plane. At this resolution no significant difference appears between the measured transfer functions.

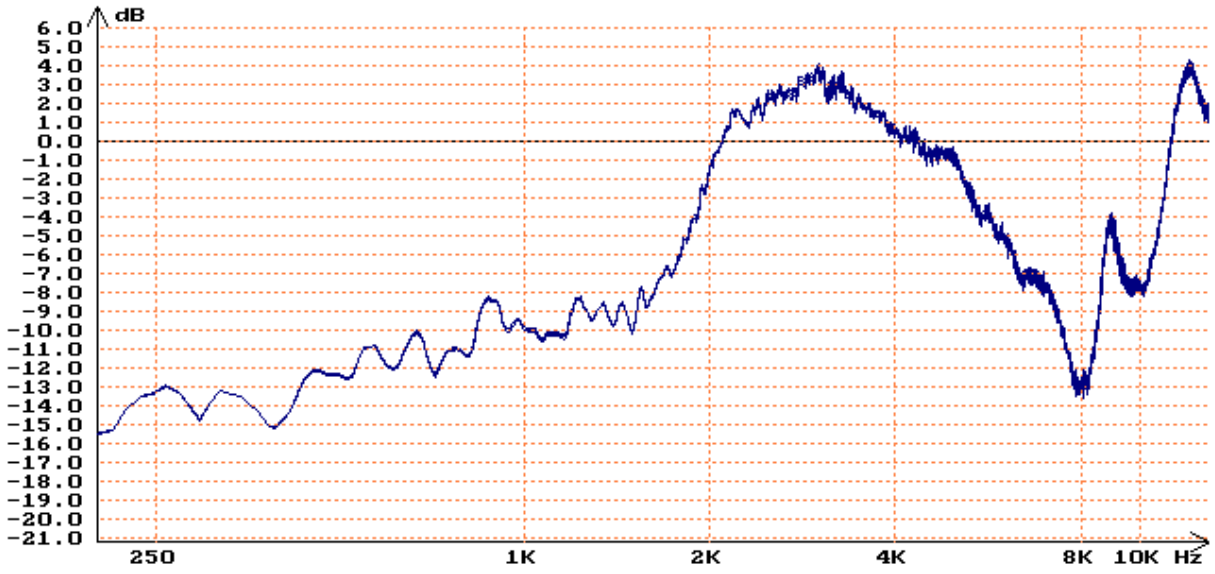


Fig.38. Set of HRTFs measured in φ=359°, 0° and 1° for the left ear in the horizontal plane.

4.2.4 The pseudo-random noise excitation

For the measurement a pseudo-random noise stimulus is used, presented periodically. A broadband noise excitation is advantageous, because more signal power can be produced in contrast to impulse excitation. That means larger dynamic range and higher SNR. The pseudo-random noise signals have the advantageous properties of the white noise signal (it is generated by a random algorithm), and on the other hand as a deterministic signal (stored as numbers on the hard disk) it can be repeated exactly [147, 148]. This makes possible the use of repeated measurement and increase of the SNR by averaging.

It is clear, if we use white noise excitation, the SNR of the measurement is frequency dependent due to the non-linear transmission of the system. This can be easily represented by the noise spectra of the system. Because this noise spectra vary with the frequency, the use of white noise excitation results in

frequency dependent SNR. Frequency independent SNR can be achieved if the spectrum of the stimulus "looks like" the spectrum of the systems' noise: the higher the noise is in a given frequency domain, the greater should be the energy contained in the stimulus.
To generate a noise signal the MLS technique is a widely used procedure through the fast Hadamard transformation [149, 150] or using Golay codes [95, 151, 152]. An alternative way to generate a pseudo-random noise signal is presented here (Fig.39.). Description of a similar signal is in [35]: the generation of a pseudo-random phase spectra with uniform distribution between $-\pi$ and $+\pi$ and flat magnitude response, IFFT of 4096-points and 44100 Hz sampling frequency is used in an averaging of 100 frames against uncorrelated noise elements and for increasing the SNR. Nowadays, sweep measurements are also preferred instead of MLS methods, because they have a quite large SNR (of about 90 dB) without the effect of loudspeaker non-linearity and harmonic distortion [153].

The algorithm was created in order to approximate the average power spectra of the entire measurement system. To get this input information we made repeated measurements with the system by zero-excitation, and the averaging was made based on signal power:

$$\mathrm{Re}_{avg}[i] = \sqrt{\frac{1}{N}\sum_{j=1}^{N}\left(\mathrm{Re}_j^2[i] + \mathrm{Im}_j^2[i]\right)} \qquad (9)$$

$$\mathrm{Im}_{avg}[i] = 0 \qquad (10)$$

$$\mathrm{i} = 0\ldots2047,\ \mathrm{N}=18.$$

The measurement was made with a unidirectional microphone in the horizontal plane turning in 20° steps (j=1…18). The algorithm will disregard the measured phase information.
The stimulus has to suit the following 3 requirements:

1., it has to be presented periodically, and the length of the stimulus should exceed *T,* where *T* is the effective length of the impulse response computed from the transfer function of the actual system.

2., the spectrum of the stimuli has to be a good approximation of the average noise spectra of the system calculated above. This results in a frequency independent SNR.

3., its crest factor has to be small, near to unity, because the power of the stimuli can be the largest this way without any distortion or overload. Furthermore, the quantisation noise will be the smallest. The crest factor is

defined as the ratio of the peak to RMS voltage (Eq.11.). The crest factor indicates how much energy is lost using a signal compared to the ideal case of a stimulus whose RMS equals to the peak value; and through the normalization the maximal energy in a measurement can be extracted [153].

$$\text{Crest Factor} = P_n = \frac{n_{peak}(t)}{n_{RMS}(t)}. \tag{11}$$

$$n_{RMS}(t) = \lim_{T\to\infty}\left\{\frac{1}{T}\int_0^T x^2(t)dt\right\}^{1/2} \tag{12}$$

where T is the linear averaging time.
The exact mathematical solution of this problem is not known, but based on numerical analysis the following algorithm is suitable to generate a sufficient signal within a reasonable running-time. A different algorithm for noise with gaussian magnitude distribution is described in [154].

Algorithm

The time difference between the samples of the stimuli is:

$$t = \frac{1}{f} \tag{13}$$

where *f* is the sampling frequency of 50 kHz. For the block-length it is:

$$N > \frac{T}{t}. \tag{14}$$

It is comfortable to choose N as the nearest power of 2 for a rapid FFT. In our case N is 4096, so the length of the period of the stimuli is 81,92 ms.

1. For starting the algorithm take the Re_{avg} average power spectra calculated above. Let the Im_{avg} phase be a random variable with uniform distribution over the 0...2π interval. The IFFT of this spectrum satisfies the first two requirements but does not satisfy the third.

$$n(t) = IFFT\left\{\mathrm{Re}_{avg} + j\,\mathrm{Im}_{avg}\right\} \tag{15}$$

2. Compute the crest factor and its square root in the time-domain:

$$P_n = \frac{n_{peak}(t)}{n_{RMS}(t)}, \tag{16}$$

$$Q_n = \sqrt{P_n}\,. \tag{17}$$

3. If there are samples which absolute values exceed

$$Q_n^{'} = n_{RMS}(t)Q_n, \tag{18}$$

reduce these samples to this Q_n value without affecting their sign.

4. Compute the FFT. The spectrum usually does not satisfy the second requirement. Normalize the spectra so that its total power will be equal to the power of the spectra at the start.

5. Let it be

$$q_n = 20\log Q_n \tag{19}$$

If there are spectral components which absolute value exceed more than q_n dB the corresponding component of the spectra at the start, cut these amplitudes so that they will be exactly q_n dB greater than the corresponding component without affecting the phase information. Similarly, modify the

component if the absolute value is more than q_n dB less, than the according component of the starting spectra.

6. Compute the IFFT again, and repeat the steps from point 2 until the crest factor seems to be small enough and the target spectrum is „close enough" to the starting spectra.

Properties and Signal-to-Noise ratio

Under a short running-time a 1,1 crest factor and a 0,2 dB deviation is obtainable. This stimulus is used for the measurements, recording the reference signal and the transfer characteristics of the loudspeaker.
Periodic signals of length 2^N are better for the FFT than binary MLS sequences that usually are generated with an N-staged shift register and an XOR-gate connected to each other, so they only have 2^N -1 states running through [153].

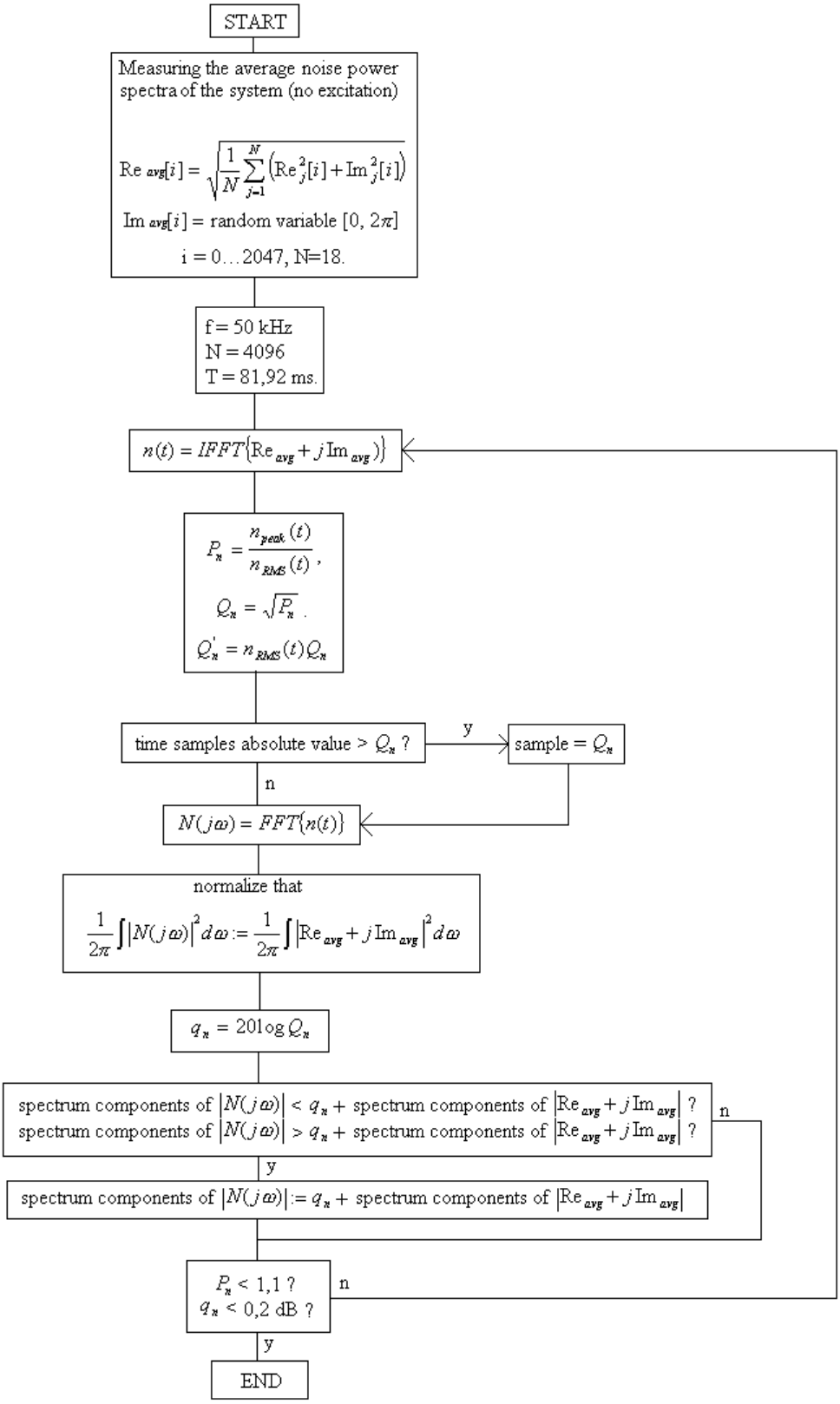


Fig.39. Block diagram and build-up overview of the algorithm for generating the pseudo-random noise excitation.

4.2.5 Effect of the high voltage periodicity

It is a well-known disturbance of acoustic signals if the periodicity of the high-voltage current (220V) can be heard. The European alternate current in the socket has a periodicity of 50 Hz, and disturbances could appear at multiple frequencies (100, 150 Hz etc.) as well.
This section shows shortly a method to decrease this disturbance by 18 dB. This method measures the real-time state of the 50 Hz current and de-synchronise the measurement. This means, every single measurement of the system is an average of 24 phase-delayed measurements inside a period of 1/20 sec. This "filtering" was tested, and the optimal number of sub-measurements was declared (24). The 24-time increase of the measurement frames and measurement time increases the overall SNR as well.
A small High Voltage Transformer (HVT) unit is plugged into the socket generating a TTL-level logical output signal according to the instantaneous frequency fluctuation of the mains as seen in Fig.40. This output signal is plugged into the DSP card.
The principle of the noise reduction is that the measurement - which alone includes 32 frames (periods) of the stimuli - will be repeated in different restart positions during 20 ms. By partitioning this interval and averaging the responses, we will get a result independent of the phase and the fluctuations of the mains periodicity. Using this procedure we can decrease the amplitudes depending of the number of restart positions.
In test measurements we had to find the optimal number of the restart positions. Fig.41. shows the results by inactive module, and restarting in 24 and 120 positions. Test signal is a 50 Hz rectangle signal.
It is clearly seen that the partition for 24 restart position is optimal with an 18 dB reduction of the amplitudes which can not be increased significantly. Increasing this number does not result in improvement of the SNR, but the measurement time increases gratuitously. All in all, we make the averaging over 768 frames of the stimuli to increase the overall SNR in one measurement.

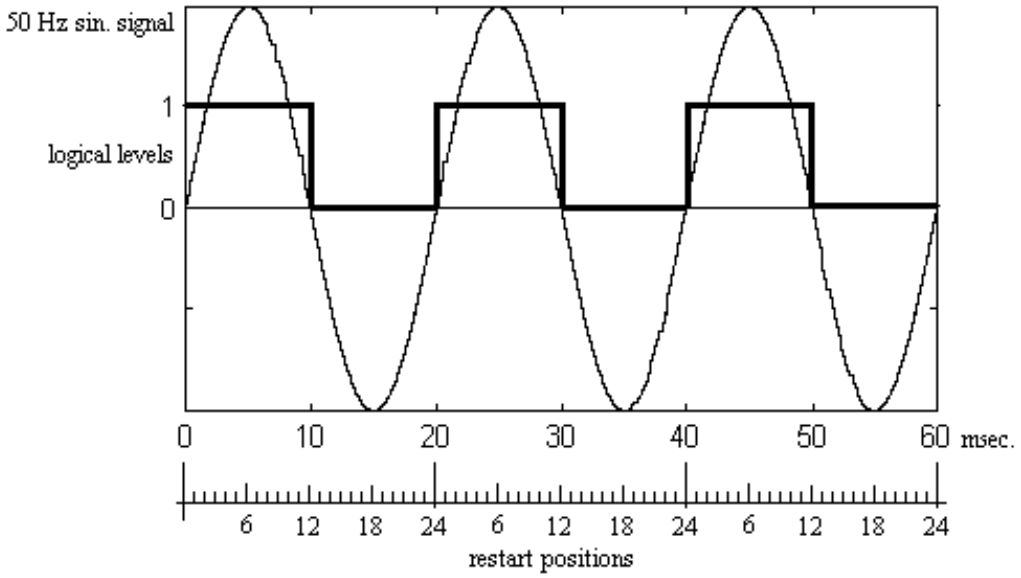


Fig.40. The function of the High Voltage Transformer (HVT) unit. Timeslots are also shown according to 24 restart positions.

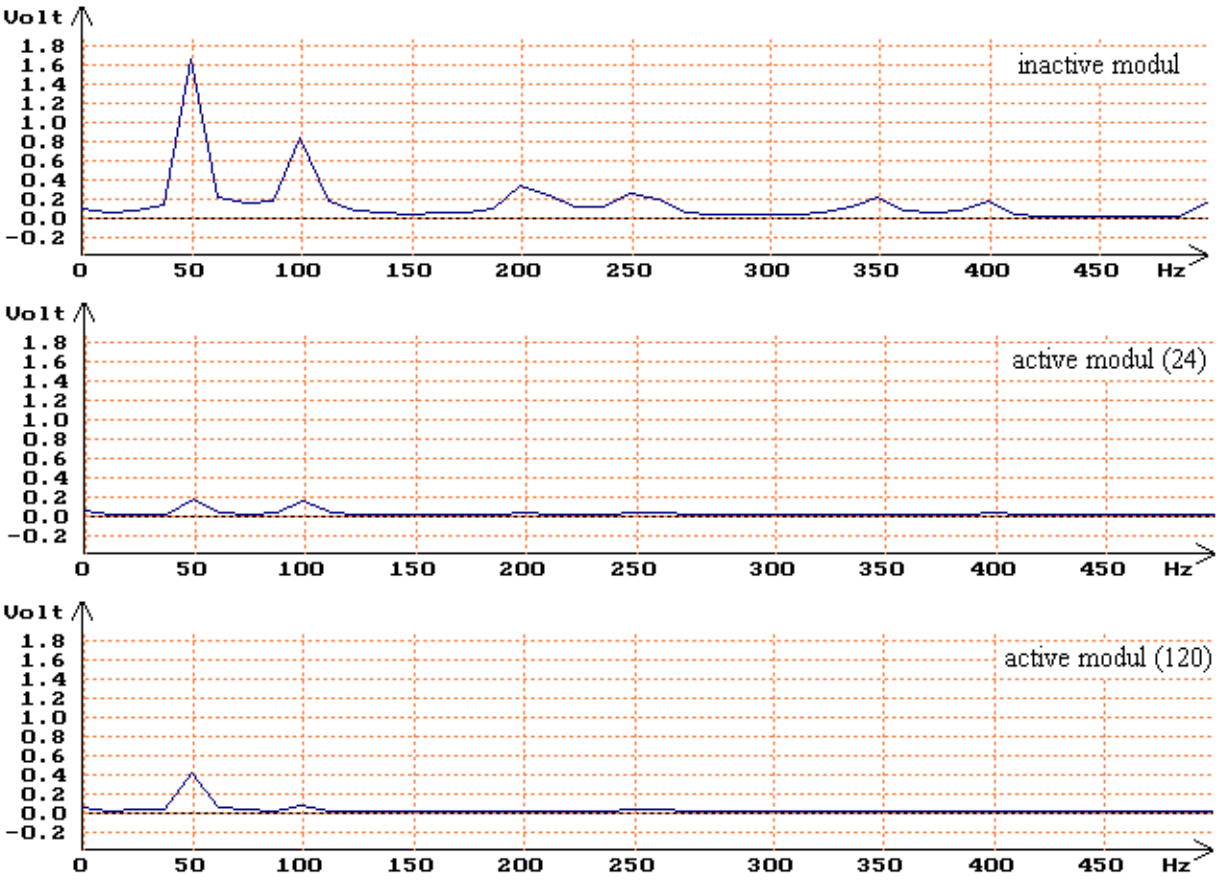


Fig.41. Testing the reduction of the components at 50 Hz and multiple frequencies. The figures show amplitude values by inactive module and by restarting in 24 and 120 positions respectively.

4.2.6 Environmental reflections and room impulse response

Eliminating the reflections is elementary in every free-field measurement situation. There are more objects producing reflections in the anechoic room, so their covering with sound-absorbing materials is necessary. Most of the reflections came from surfaces parallel with the horizontal plane near to the ears. Fig.42 shows the smoothing effects at low frequencies of the absorbing materials placed on the turntable. Over 3 kHz this effect is not significant. Our former measurements also suggested using absorbing cover on the frame holding the loudspeaker.

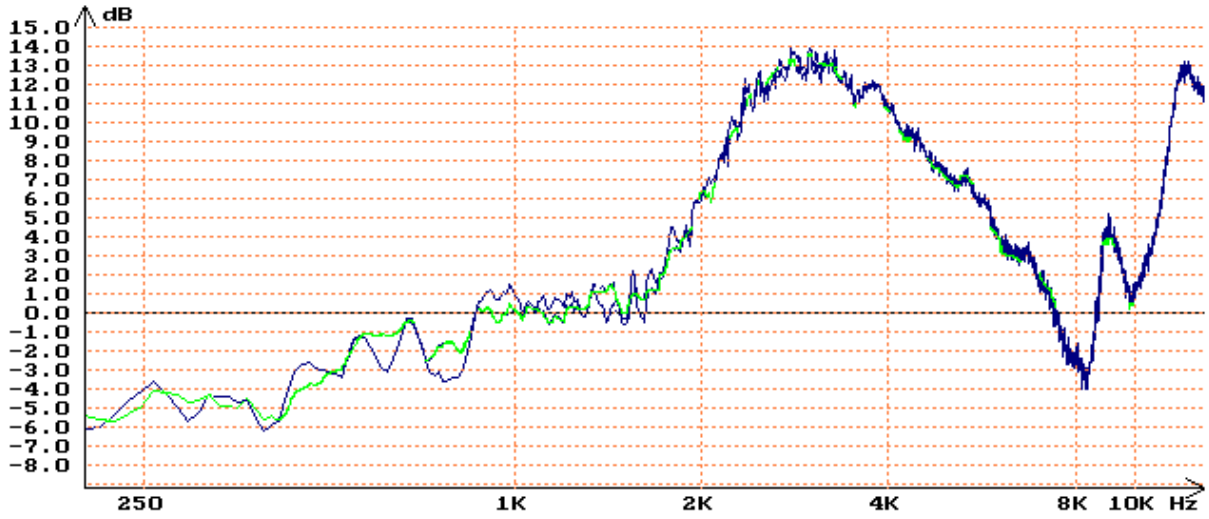


Fig.42. Low frequency smoothing effect due to the covering of the turntable surface and the rails. Blue line is for the original HRTF and green is for HRTF measured with absorbing materials.

Measuring of the impulse response will give the answer to the remaining reflections (their existence, direction and time delays) and by spectral evaluating we can find marks of intraperiodic or interperiodic time variances [150]. Both single-frame and averaged impulse responses were measured. To generate an impulse, a 4096-element array was filled with zeros, except the first element (value 1000).

$$x_{IR}(n) = \begin{cases} 1000; n = 0 \\ 0; 1 \le n \le 4095 \end{cases} \quad (20)$$

By evaluating the response array we calculated with the value of 344 m/s as the speed of sound. The resolution of the response is 20 ms, the total length of the array is 81,92 ms. The primary incidence arrives after 6,22 ms. The traveling time in the air is 4,8 ms. The first 500 elements are „useful", the rest is due to remaining reflections. These could be eliminated by truncating the impulse response. Fig.43. shows the impulse response in linear scale.

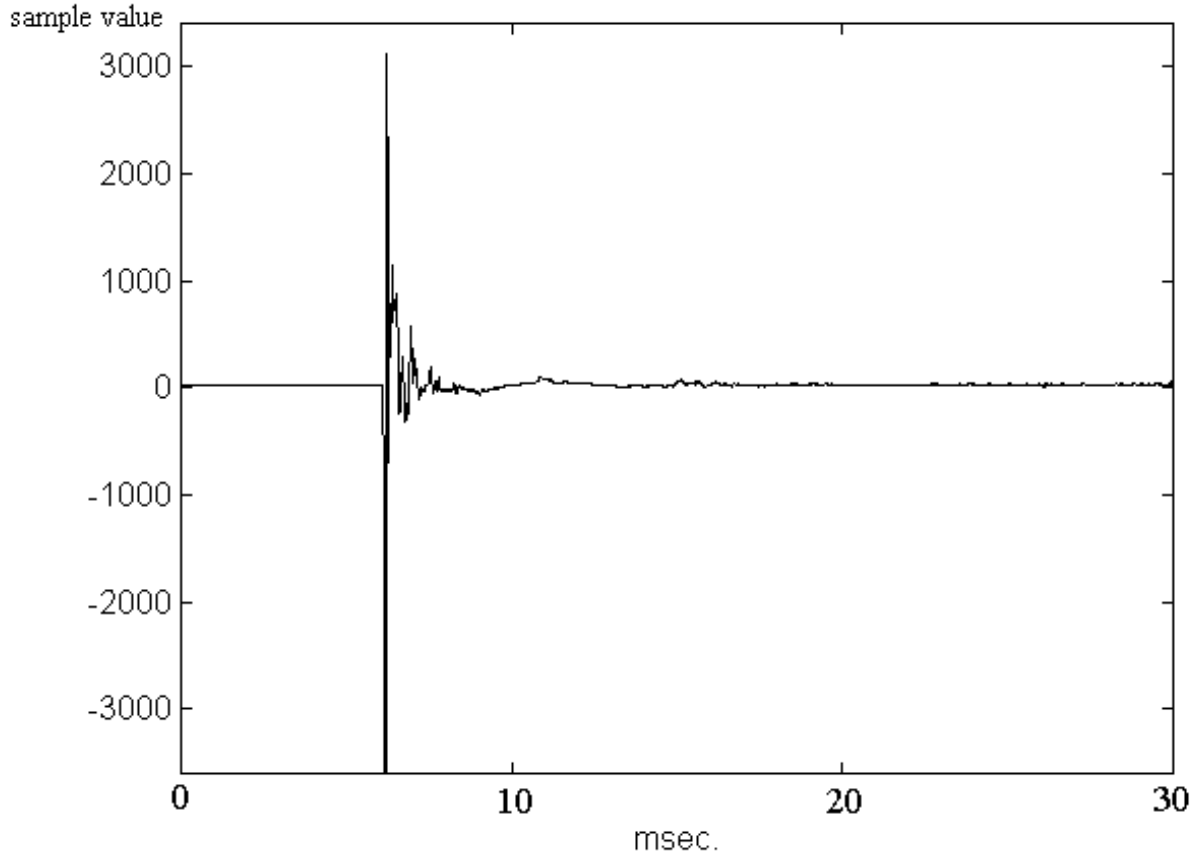


Fig.43. The impulse response on linear scale.

4.3 Testing

As a general rule we use the fact that every time we double the number of the same measurement, the SNR will increase +3dB as the result of the averaging. This is only valid if there are no time variance effects within a single measurement or between measurement frames and the noise is uncorrelated with the signal. The presence of intra and interperiodic time variance can be detected from the spectra of the impulse response [150]. In contrast to a system measuring with impulse excitation getting the same SNR needs less time using broadband signals [118]. The effect of the 768-times averaging is clearly visible on the spectra of the noise in the left-channel. No excitation is present. Fig.44. shows the same spectra measured in single-frame and using the 768-times averaging for comparison. The same smoothing effect is achievable by truncating the impulse response function in time-domain.
Our former measurement with this system had an averaged SNR only about 60 dB [45]. In the literature of spatial hearing and acoustical measurements signal-to-noise ratios were reported from 20-30 dB up to 60-70 dB [8, 9, 10, 12, 61, 180]. The DSP card normally satisfies within the half-bit error limit the following SNR-equation:

$$SNR \approx 1{,}74 + 6{,}02n \approx 98{,}1dB \qquad (21)$$

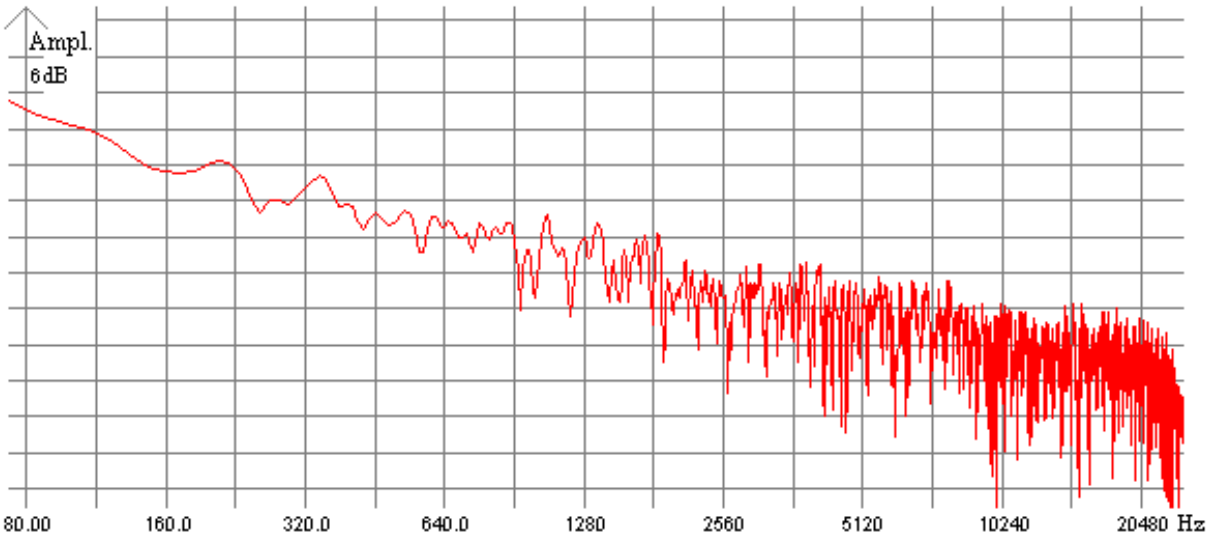
in case that the entire dynamic range is maximally used (n=16). As Eq.21. shows, if the signal to be converted from analog to digital does not use on average the 16 bit resolution, the SNR will decrease (see also Eq.6.) [155, 156].
During test measurements we determined the highest amplitude value to be converted. It is from φ=90° incidence in the horizontal plane. In the origin this corresponds to a 74,6 dB SPL. The maximal number of bits needed to convert the peak value of the analog signal is 15,5 bit, but samples in a two-channel measurement use only 9,78 bits on average. Using Eq.21. indicates that the average SNR based only on the used signal processing system is ca. 60 dB.
Random noise can be seen as a stationary random variable with uniform distribution. If a stimuli is deterministic and is presented periodically, by doubling the measurement frames a +3dB improvement in the SNR can be achieved by averaging. The 768-times averaging used delivers a +28,85 dB increase, and thus, the average SNR is about 89 dB. In addition, this SNR is increased by reducing the high voltage mains disturbing effect at special frequencies.
The number of the measurement frames is limited by the memory of the DSP card and by the measurement time. The highest number allowed by the on-board

memory is 15840 frames, which correspond to a +42 dB improvement. On the other hand, the time of a „single measurement" will be 20 times higher, than with 768 frames. We have measured HRTFs with both frame numbers. Fig.45. shows the left ear HRTFs in the frontal direction averaged over 768 and over 15840 frames. There is no visible difference which would be worth an increase in the measurement time that much.
The reproducibility of a system tells us if we are able to re-measure the same transfer function (TF) under the same conditions with the same precision. There is an easy way to find out the precision and reproducibility if we divide the transfer functions from the same direction in repeated measurements:

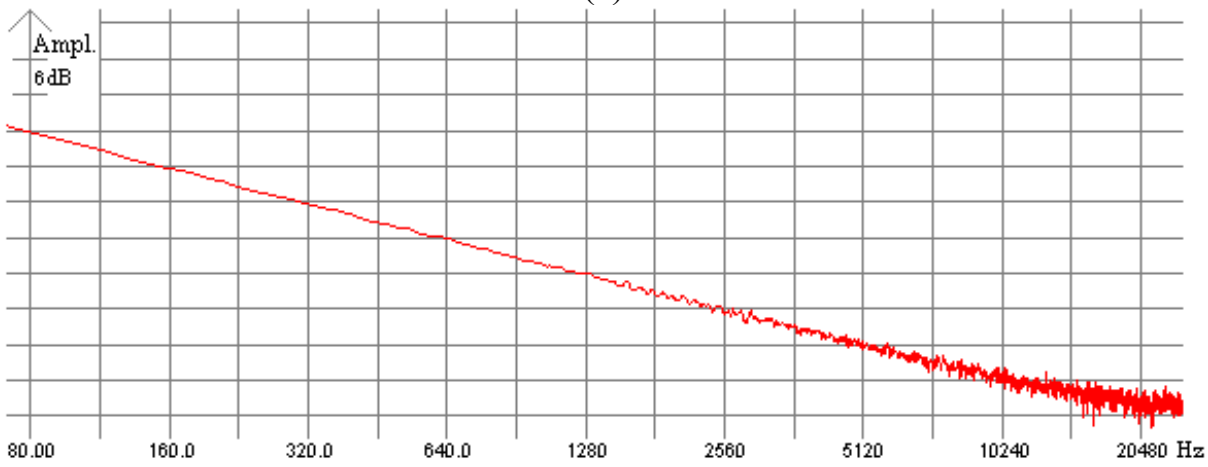
$$TFD_i = 20\log\left(\frac{TF_1}{TF_i}\right) \qquad (22)$$

where TFD_i are a set of transfer function differences showing the deviations in dB as the function of the frequency (i is the number of measurements).
Basically, the system accuracy depends on reflections, signal processing, and on the accuracy of the elevational and azimuthal settings. Deviations between the transfer functions in the entire frequency range are less than 0,5 dB in repeated measurements with unidirectional microphone independent of azimuth and elevation.
Using a periodic chirp source signal (sine-wave sweep) in 1 degree resolution, where the number of periods was equal to the number of the FFT points, a reproducibility of ±1,5 dB was reached and declared to by satisfying [52]. Existing methods of obtaining TFs are reviewed in [153].
Later, Fig.46. will show that head-shadow effect occurs only over about 4 kHz, which supports former results [10]. *Shaw* observed a constant decrease of the magnitude of the HRTF in the shadow zone (–15°–75°) up to –75° between 3-6 kHz due to waves passing over the head, and interference is between 0-60° [6]. Due to our concurrent measurements it depends on the azimuth, elevation and on the acoustical environment near to the head as well [45, 137].
Measurements with the head and torso simulator have deteriorative accuracy depending on azimuth and elevation caused by the head shadow, the shoulders, the asymmetry of the head and the pinnae. KEMAR HRTFs have been shown to produce a good „typical" head shadowing model [179]. The polar histogram in Fig.47. shows the deviation-ranges in repeated HRTF measurements in the horizontal plane. The distended circles represent linear frequency scaling with

the appropriate center frequency. As expected, in the most sensitive range the fluctuation is small, less than 1-2 dB. As the source moves into the head-shadow area the HRTFs will be hard to evaluate. The shadowing-effect of the head produces random effects and thus, even from the same direction, a large deviation is natural.



(a)



(b)

Fig.44. Effect of averaging. The noise spectrum of the left channel is shown by zero excitation.
(a) No averaging was used (single-frame). (b) After averaging over 768 frames of the stimulus.

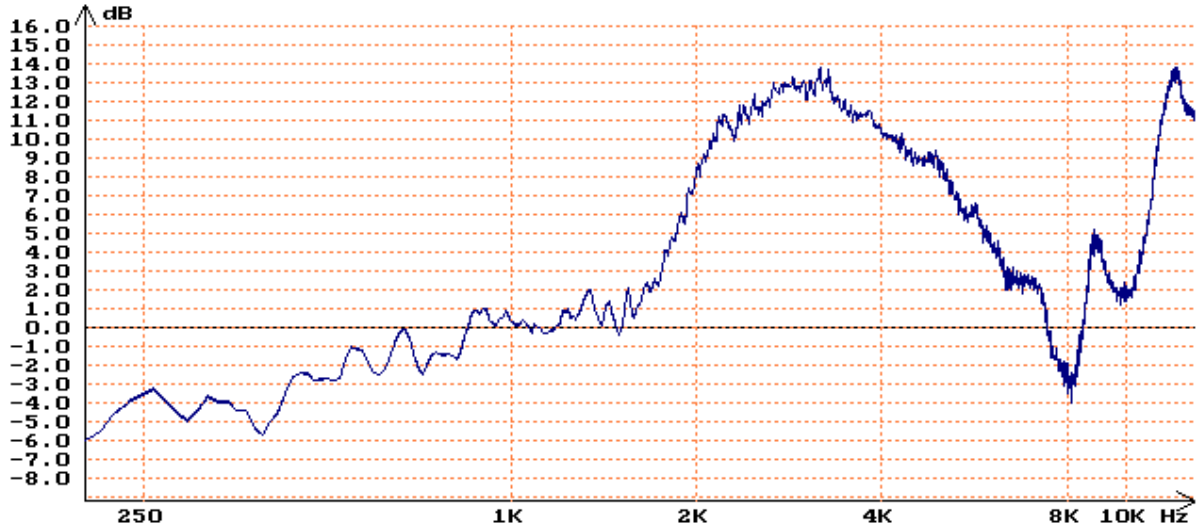


Fig.45. HRTF measurements by averaging of 768 and 15840 frames in the direction δ=φ=0° (left ear). These correspond to an increase of the SNR of +28,85 dB and +42 dB respectively. There is no visible difference.

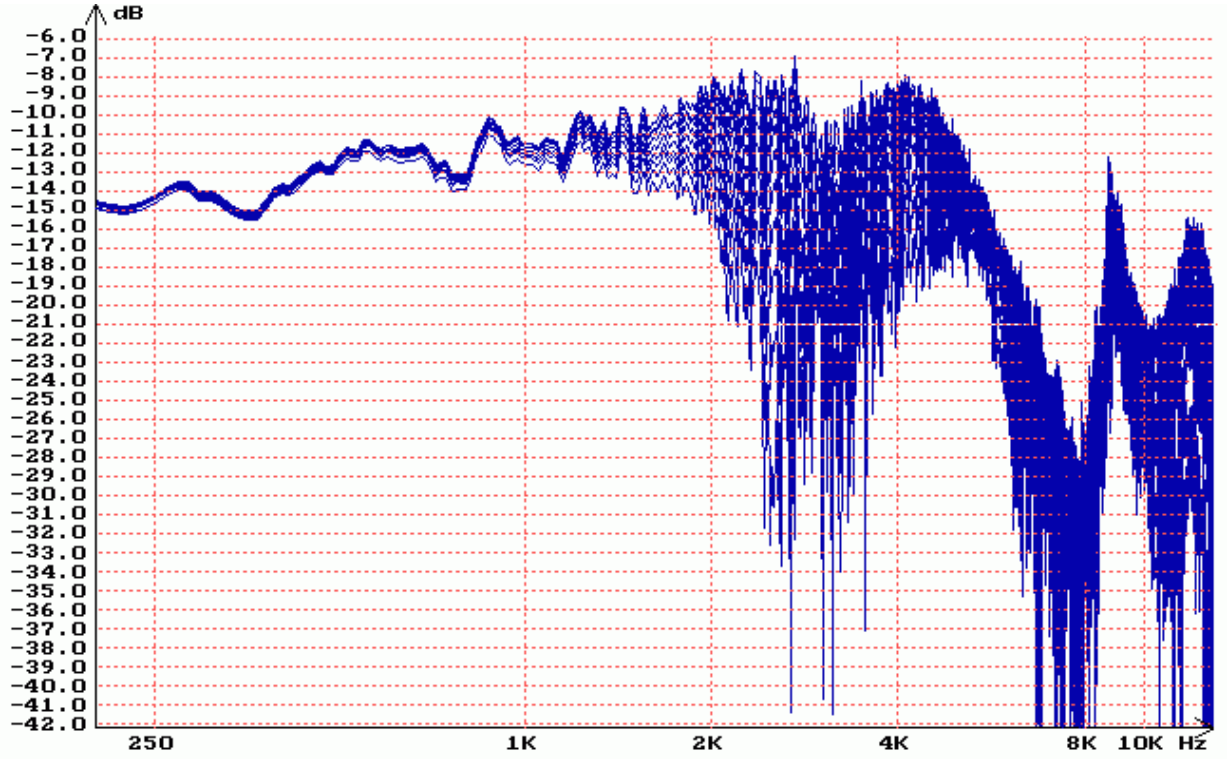


Fig.46. Example of the variable domain caused by the head-shadow. Set of 10 HRTFs measured between φ=250°-260° azimuth (left ear) in the horizontal plane. Above 1600 Hz the HRTFs vary to rapidly.

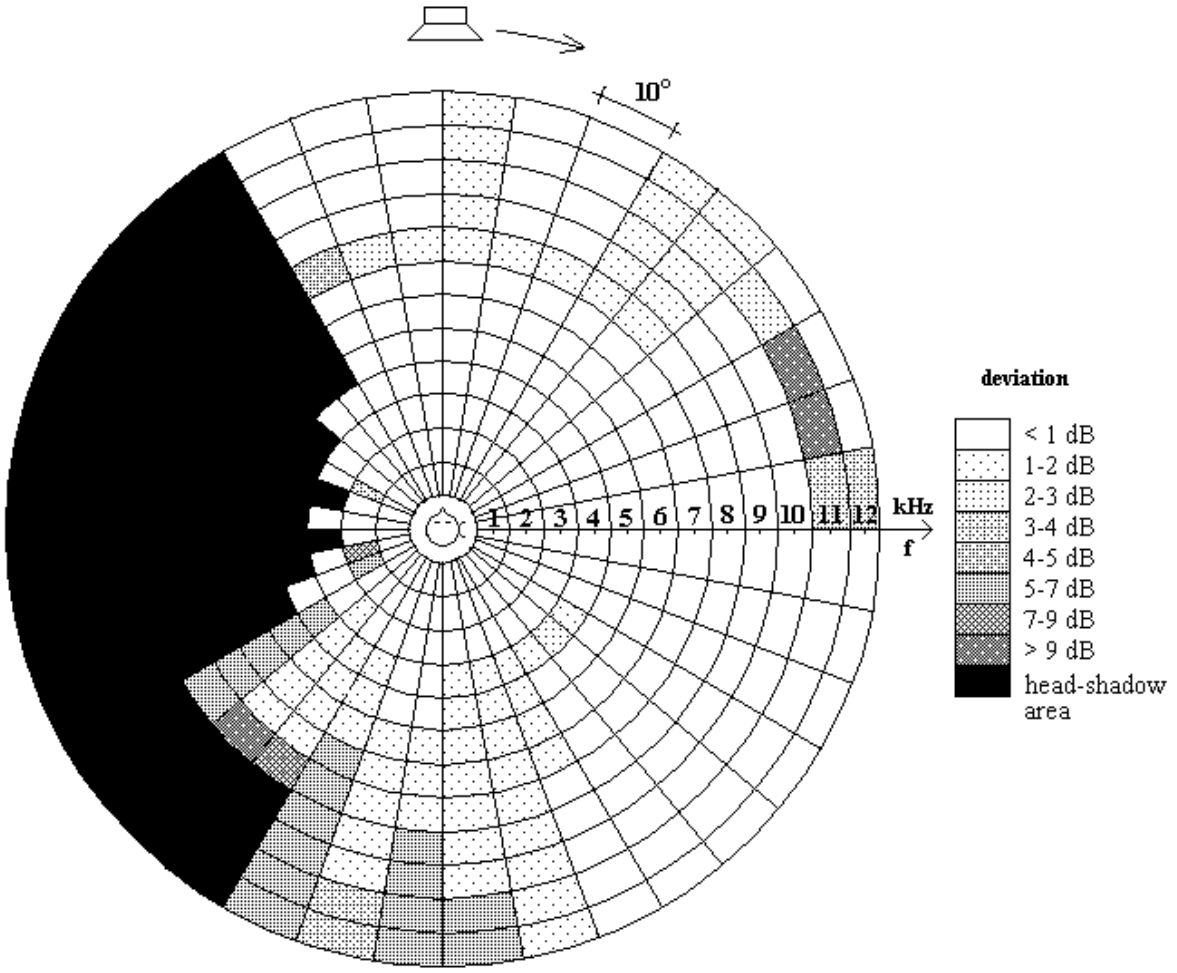


Fig.47. 2D spatial representation of the magnitude of HRTF data for a fixed elevation as a function of azimuth and frequency. The polar histrogram shows the deviations between HRTFs from repeated measurements in the horizontal plane for the right ear based on Equation 16. The natural deviations of the HRTFs caused by the filtering and shadowing effects are shown as unsigned absolute values in dB. The circles correspond to frequency domains with 1 kHz bandwidth marked with the center frequency (linear scale). Note the head-shadow area (filled black) and the domain caused by the pinnae between 60 and 90 degree at 11 kHz and the contralateral side at 2 kHz ("bright spots" [37]). The "noisy domain" is in the head-shadow area, where HRTFs vary more than 9 dB after independent measurements from the same direction. This leads to disturbed and rapidly variabled high frequency components.

Note the large deviation in the φ=60-80° domain around 11 kHz and on the contralateral side at 1600-2500 Hz. The HRTFs from this direction in repeated measurements are almost identical, but between 11 and 12 kHz they are shifted about 25-30 Hz and this difference is enough to affect the TFD and produce large deviation. Small shifts in the frequency of sharp notches in the HRTFs from recording to recording result in relatively large variations over very narrow frequency bands [95]. This suggest the importance of the microstructure of the HRTFs. Removing the shoulders or covering them with absorbing materials also decreases this effect but does not eliminate it completely like removing of the pinnae. This high frequency random effect of the pinnae caused by its reflections can be handled by the „multipath-theory" calculating secondary sound paths in the time-domain [54].

4.4 Summary

HRTF database was recorded in the anechoic room using a dummy-head placed on a turntable with the following properties (Table 13):
Spatial resolution is 5° from -45° to +90° in the median plane and 1° in the horizontal plane. Setting the proper sound source direction can be made using the calibrated LTS with the precision of 0,77% elevational, and with 1,14% azimuthal calculated with 1,8 m source distance.
The signal processing includes 50 kHz sampling frequency, 16 bit resolution in two channels simultaneously and 4096-points FFT. The resolution of the response is 20 μs and 12,2 Hz linear in the time-domain and frequency-domain respectively. The validity of the measured transfer function is above 200 Hz.
Electronic disturbances like the high-voltage mains influence and random noise effects are decreased during long-time averaging. The high-voltage components are reduced by 18 dB and, moreover, the SNR allowed by the signal processing is increased by 28,85 dB. The overall SNR is frequency independent and reaches 89 dB or more. The optimal repeat of the signal is 768 frames.
The applied non-MLS pseudo-random noise stimuli can be used for every transfer function (TF) measurement with this system. The algorithm is able to generate easy and fast the proper stimuli for every system other than ours.
The deviation (precision and reproducibility) between the measured TFs is less than 0,5 dB considering that HRTF measurements have increased uncertainties caused by the head and the torso. The next section shows how we can evaluate a huge amount of HRTF data and spectral differences of about 1 dB.

source	
elevation	[-45°,+90°], 5° steps, 0,77% precision
azimuth	[0°, 360°], 1° steps, 1,14% precision
distance	1,8 m
signal processing	
sampling frequency	50 kHz
conversion	16 bit maximal; 9,78 bit on average
FFT	4096 points
resolution	20 μs., 12,2 Hz (linear)
channels	2 channel, simultaneously
averaging	>768 frames
reduction of high-voltage disturbance	18 dB at 50Hz, 100 Hz etc.
Loudspeaker	
diameter	0,12 m
Transfer function (TF)	constant (within ± 6°)
stimuli:	non-MLS 81,92 ms. pseudo-random noise
SPL in the origin	74,6 dB (max.)
validity of the TF	above 200 Hz
overall SNR	> 89 dB (frequency independent)
deviation in TFs	< 0,5 dB (reproducibility)

Table 13. Datasheet of the measurement system

5 Evaluation of differences in dummy-head HRTFs caused by the acoustical environment near to the head

5.1 Introduction

In section 4 we introduced an accurate dummy-head measurement system for automatic recording of the HRTFs. The precision allows us to evaluate differences in a range of about 1 dB in the fine structure of the HRTFs and in the so called HRTFDs (HRTF Differences). We do not need individual recordings or human interaction.
We have also seen that the measurement of the HRTFs shows significant "random effects" due to torso reflections and shadowing. The term random in case means if we measure the response of the ears in the "head shadow area" we do not measure the same response over time in repeated measurements. This is due to the natural shadowing and reflecting effects (existing secondary sound paths and no primary incidence) of the head and/or the modifications of the acoustical environment.
The connection between the variations of the HRTFs and the acoustical environment near to the head is discussed here. Our goal is to search typical properties of the magnitude response of the HRTFs as the source is moving in the 3D space. We will show how the head-shadow and pinnae reflections influence the sensation on the lateral side (closer ear) and at the contralateral ear. The primary (direct) wave reaches only the closer ear but the contralateral ear only gets secondary reflections (diffuse-like sound field) – therefore no high frequency information.
"Small changes" in the environment near to the head influence the HRTFs significantly, up to 15-20 dB. We can find typical properties and effect of glasses, hair, caps or clothing. In the real life we do not hear differently, we do not have decreased localisation performance after a hair-cut or without wearing the glasses. On the other hand, smaller changes in the HRTFs during a binaural playback may lead to insufficient or distorted localization. Based on Fig.47. we show polar histograms showing an "overlapped" effect: the normal deviations and their extensions and changes are due to these everyday life objects. The effect is showed on some figures only as the function of frequency as well. The

evaluation is mostly based on spectral *differences* between the HRTFs. Comparison of different methods can be found in [157].

5.2 Terms of use

The mathematical analysis uses the following definitions and abbreviations. Assuming that the complex HRTFs are divisible mathematically, the free-field HRTF Difference (HRTFD) is defined as a quotient of HRTFs from the same direction but under modified conditions:

$$HRTFD = \frac{HRTF_{C1}}{HRTF_{C2}} = 20\log\left|HRTF_{C1}\right| - 20\log\left|HRTF_{C2}\right| \quad (23)$$

where C_1 identifies the reference and C_2 the modified condition. We plot the 20log/HRTFD/ magnitude response as the function of frequency (Fig.63-65) or as 2D polar histogram as function of frequency and azimuth (Fig. 57-62).
The complex quotient refers to subtraction of two logarithmic magnitude responses. This difference gives us the deviation in dB between two HRTFs measured in the same direction but under modified conditions at all frequencies. For analyzing the HRTFDs we do not need individual recordings on real human heads because the dividing will eliminate the individual differences. The dummy-head HRTFs can be regarded as a particular individual set of HRTFs. We are only interested in changes and deviations caused by modifications of the acoustical environment near to the head.
All the interaural differences can be calculated as well. E.g. the interaural HRTFD from φ=30° is defined as the quotient of the monaural HRTFDs:

$$HRTFD(f)_{\text{int}\,eraural,\Phi=30^\circ} = \frac{HRTFD(f)_{monaural,\Phi=330^\circ}}{HRTFD(f)_{monaural,\Phi=30^\circ}} \quad (24)$$

showing the difference between the HRTFDs of the two ears. For further applications we can calculate the difference of second order: differences between two HRTFDs may represent the effects better.
Properties of the HRTFDs are:

- They can be easily calculated (complex division)
- No individualism needed (it will be eliminated by the dividing)

- The system above is able to measure them accurate, automatic and in a huge amount
- Differences of about 1 dB can be evaluated
- They can determine the measurement accuracy (using a unidirectional microphone)
- In simple cases they are able to detect primary reflections, their distance and show the affected spectral regions by the reflections without any time-domain measurement [158].

First we have to measure all the HRTFs in a so-called "normal" situation: the torso is placed on the turntable without any additional materials ("bare torso"). The $HRTF_{C1}$ database is recorded according to the resolution and accuracy given in Table 13. The next step is to repeat the whole measurement the same way - except with certain modifications in the acoustical environment (database $HRTF_{C2}$). These are: wearing a cap, glasses, clothing and having hair. Calculating now all the HRTFDs the results will show the acoustic role of the measured objects together (overlapped) with the natural deviations of the torso. With an extensive analysis we are searching for significant and representative effects in the frequency domain to find how, where and how much do they influence the HRTFs. It will be shown, that these small changes in the acoustical environment influence the HRTFs significantly.
Originally the HRTFD has up to seven dimensions:

$$HRTFD=F\{\delta, \varphi, r, j\omega, channel, condition\}. \qquad (25)$$

Because it is hard to handle with so many parameters, some simplifying is required. In the first approximation the HRTFDs are independent from the source distance *r*, from the complex variable *j* (only magnitude responses are evaluated) and from the number of channels. The latest one is based on the fact that the head and torso simulator is completely symmetrical and our observations support that all results from the left ear are identical to those from the right ear (360°-LEFT). If the conditional variable can be set only as the three different objects mentioned above, we get the following functions to evaluate:

$$\begin{aligned} &HRTFD_{hair}(\delta,\varphi,f) \\ &HRTFD_{cap}(\delta,\varphi,f) \\ &HRTFD_{glasses}(\delta,\varphi,f) \end{aligned} \qquad (26)$$

where $f = \frac{\omega}{2\pi}$ in Hz; δ identifies elevation; φ identifies azimuth in degrees.
The magnitude responses of the HRTFDs are plotted for a *fixed elevation and condition* as a function of azimuth and frequency on the polar diagrams of figures 57 to 62. Other 2D spatial-domain representation of HRTF data can be found in [157].
The rise of the spectral curves both for HRTFs and for HRTFDs is defined in dB/Hz. With this variable we can quantify the edges of peaks and valleys in the HRTFs. After the derivation the DHRTF is defined as

$$\mathrm{DHRTF} = 20\log\left|\frac{\partial HRTF(j\omega)}{\partial f}\right| \qquad (27).$$

This curve needs smoothing, e.g. by moving windowing with variable bandwidth. Complex FFT and power spectral smoothing using one-third-octave band or non-uniform filter-bank performing a moving average over frequency (reduced resolution) is described in [159].

5.3 Evaluation of dummy-head HRTFs in the horizontal plane based on deviations in the peak-valley structure of the bare torso

In this section we analyze the horizontal plane HRTFs from the HRTF_{C1} set recorded in a resolution of 1°. Due to the median plane symmetry the analysis is made only for one ear. We search for typical changes in the peak-valley structure both in frequency and magnitude by azimuthal movements of the sound source. The procedure corresponds to *plotting of ten nearby HRTFs together* as the source moves from φ=0° in the horizontal plane around the head. The "thickness" of the plotted lines delivers relevant information about the effect of azimuthal turning: if the figure containing ten HRTFs is "thin", the similarity between the HRTFs is large; if the figure is "thick", the HRTFs vary significantly as the sound source is moving.
In the region 0°-30° there is a constant increase of the overall HRTF level up to 3-5 dB independently from the frequency (Fig.48). Furthermore, the peak at 9 kHz increases by 7-9 dB. Other deviations of the nearby HRTFs are limited under 1 dB except between 2-10 kHz where this limit is 2 dB.

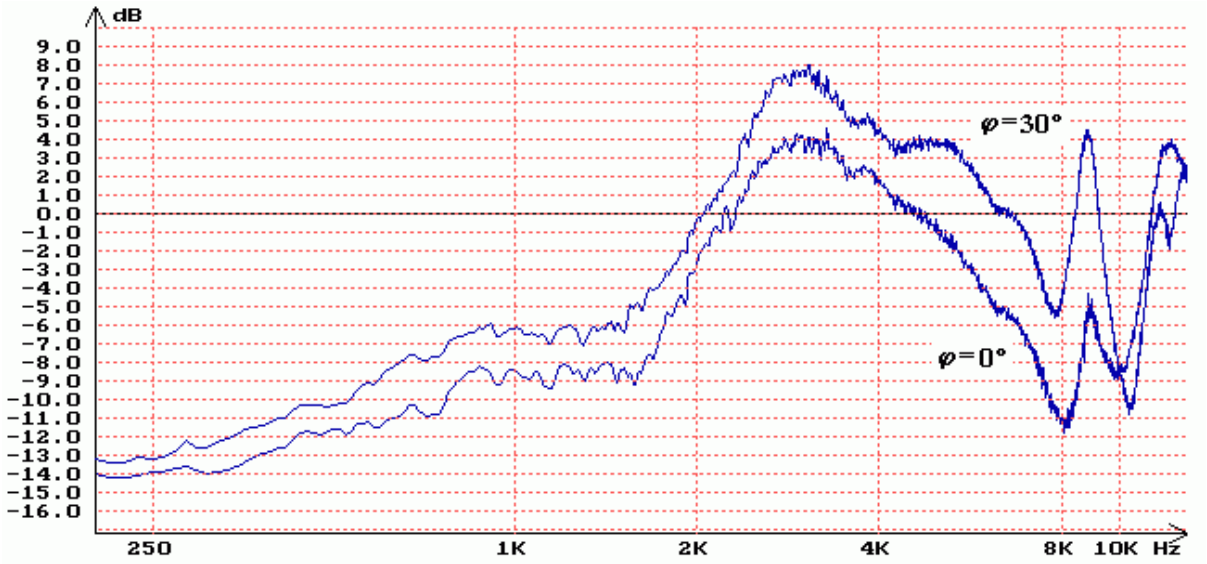


Fig.48. Horizontal plane HRTFs from the directions φ=0° and φ=30°. The overall signal level increased without significant changes of the peak-valley structure. The peak at 9 kHz increased about 9 dB.

The signal level reached at 30° remains constant until 80°. This region seems to be the most sensitive monaural domain of the hearing system in the horizontal plane: the plotted HRTFs are similar. The positive-going edges are very thin; the changing of the azimuth is only noticeable on the height or deepness of a peak or valley (Fig.49.). Only the changes in the domain between 7-8 kHz are not limited under 1 dB. Some increase of the peaks and valleys at 8, 10 and 12 kHz is also noticeable. This region can be identified as the "monaural sensitivity domain" with the axe φ=45° (Fig.50). It was also found, that measured ITDs are minimal between 1,2-1,6 kHz between 15° and 60° [51]. The same minimal changes (1-1.5 dB) within repeated measurements and asymmetrical spectral variations of the HRTFs about the interaural axis were also found by *Carlile and Pralong* supporting our observations [95]. They show the so called minimum audible field (MAF) sensitivity function, which describes the minimum detectable pressure level, determined at the position of the subject's head for a free-field stimulus in the median plane. This is also defined as a binaural measure of sensitivity for a free-field sound but it can be applied to the monaural HRTFs. It seems there is a marginal increase in sensitivity under binaural listening conditions.

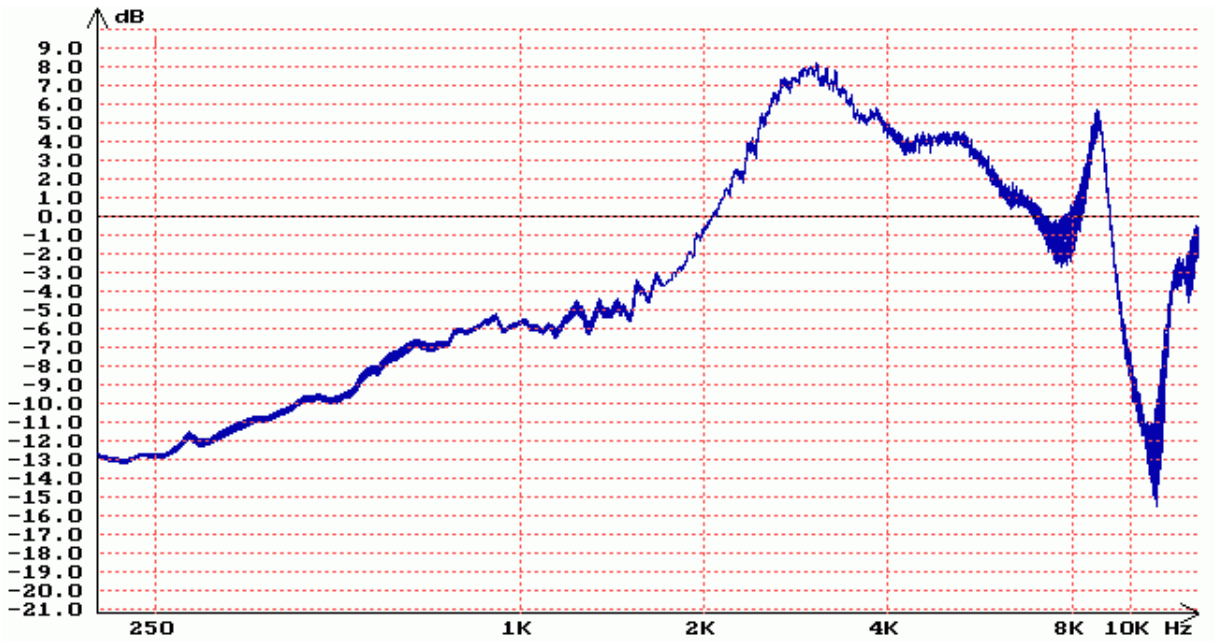


Fig.49. Typical changes in the peak-valley structure of the HRTFs in the horizontal plane. Ten figures are plotted between 40 and 50 degree in 1° resolution. The positive-going edges are very "thin" and changing of the azimuth is only noticeable on the height or deepness of a peak or valley.

As the source moves from the most sensitive domain the "thickness" of the plotted HRTF curves increases. Between 70°-110° the most important peak at 3 kHz and the valley at 4 kHz is falling down by 4 and 9 dB on aggregate respectively. The positive- and negative-going edges are still very thin, but the height of the peaks and valleys is changing significantly, up to 5-7 dB (Fig.51).
The effect of the pinnae at 11 kHz between 70°-90° is discussed above and in [144]. The HRTFs from this direction have a random frequency-shift effect. This means that the HRTFs are almost identical during repeated measurements, except between 11 and 12 kHz, where a small frequency shift of about 25-30 Hz appears causing large differences (up to 15 dB) in the quotient of the magnitude responses.
Decrease of the overall signal level at the middle frequency components is conspicuous between 90°-140°. At 4 kHz this can reach 20 dB (Fig.52). The signal level increases between 140°-180°. This area can be influenced very much by affecting the acoustical environment near to the head.

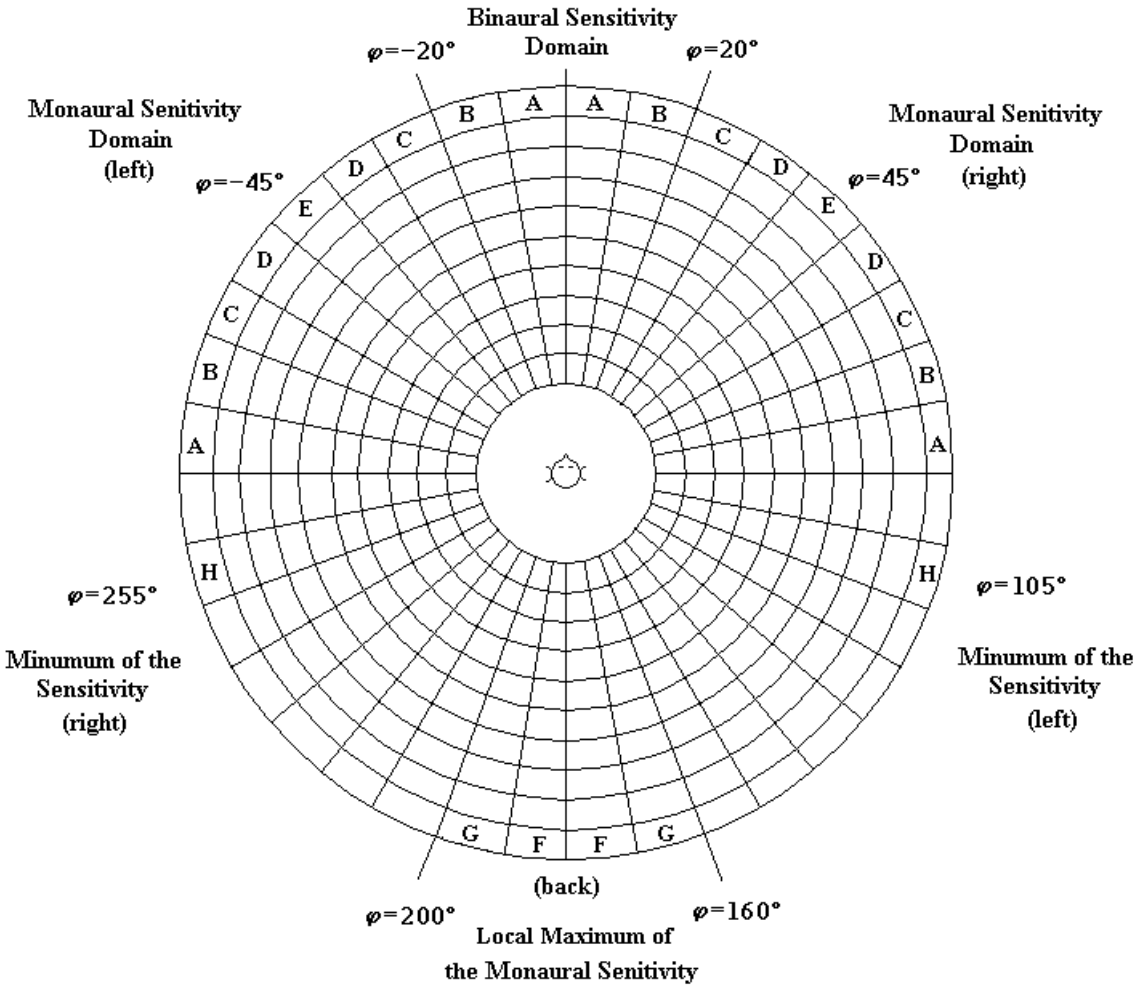


Fig.50. Monaural sensitivity domains for the left and the right ear. The overlapping area between ±20° defines the binaural sensitivity domain (see Fig.54.). Letters show the symmetry to the 45°-axe. HRTFs from the items with the same letter are similar for the according ear. The local maximum area is symmetric to the 180°-axe (see Fig.53.). The minimum of the monaural sensitivity is on the contralateral side: at 255 degrees and at 105 degrees for the right and left ear respectively (see Fig.55.).

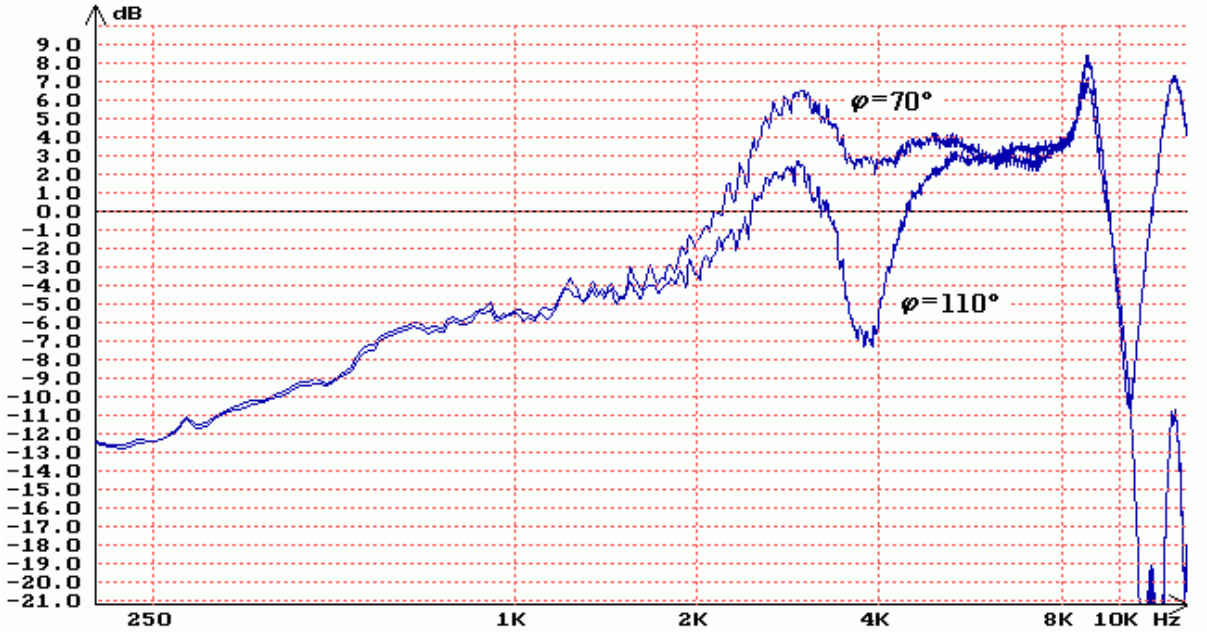


Fig.51. Horizontal plane HRTFs from the directions φ=70° and φ=110°. The valley at 4 kHz decreased about 9 dB. The positive and negative going edges do not vary in this domain.

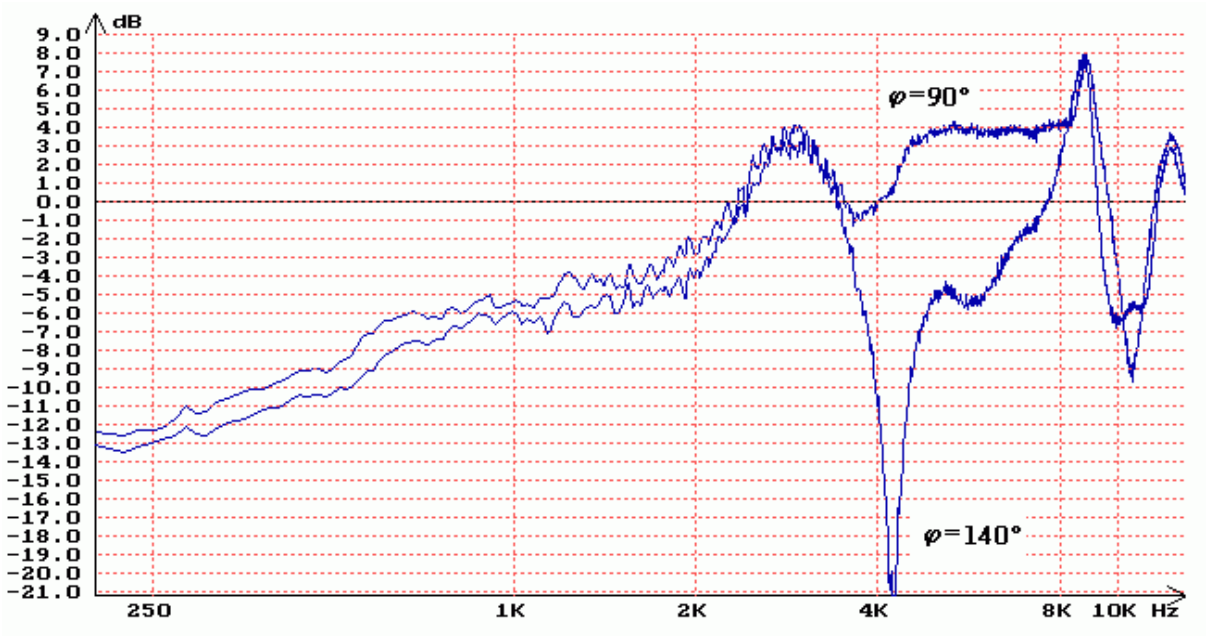
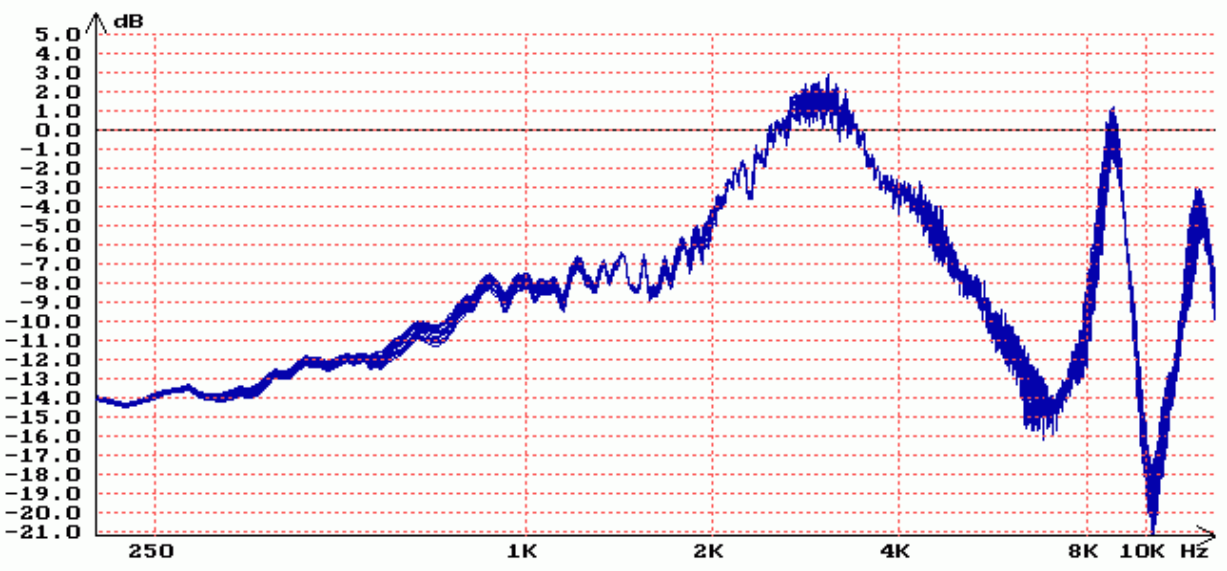
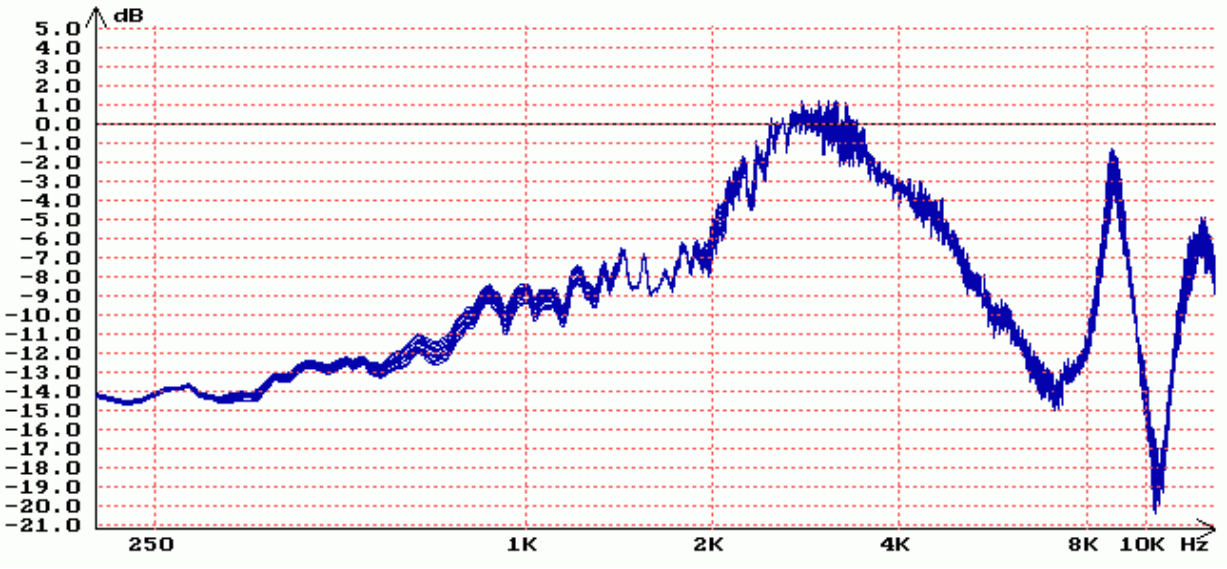


Fig.52. Horizontal plane HRTFs from the directions φ=90° and φ=140°. Only the domain between 4-8 kHz changes significantly.

In the direction “back” we have thin lines again referring to a median plane source, where no interaural level differences appear and the auditory system needs all the HRTF information for the localization. The sensitivity of the hearing has a local maximum at 180° with a ±20° symmetry. An interesting result is that the same ±20° symmetry is visible at the “frontal” direction (Fig.53, Fig.54).
The head-shadow causes level decrease and random effects in the HRTFs [5, 6, 7, 10]. Over 200° the overall signal level decreases ca. 2 dB/10° and the overall line thickness is getting thicker also ca. by 2 dB/10° above 1 kHz. The local and absolute minimum of the sensitivity of the hearing is between 250°-260° (Fig.55). Symmetrical to this region a little improvement begins, but only after 300° are the usual peaks and valleys recognizable (3, 9, 12, 15 kHz) with a thickness of 4-5 dB. The domain 340-360 degrees are comparable with 0-20° (Fig.54 a-b).

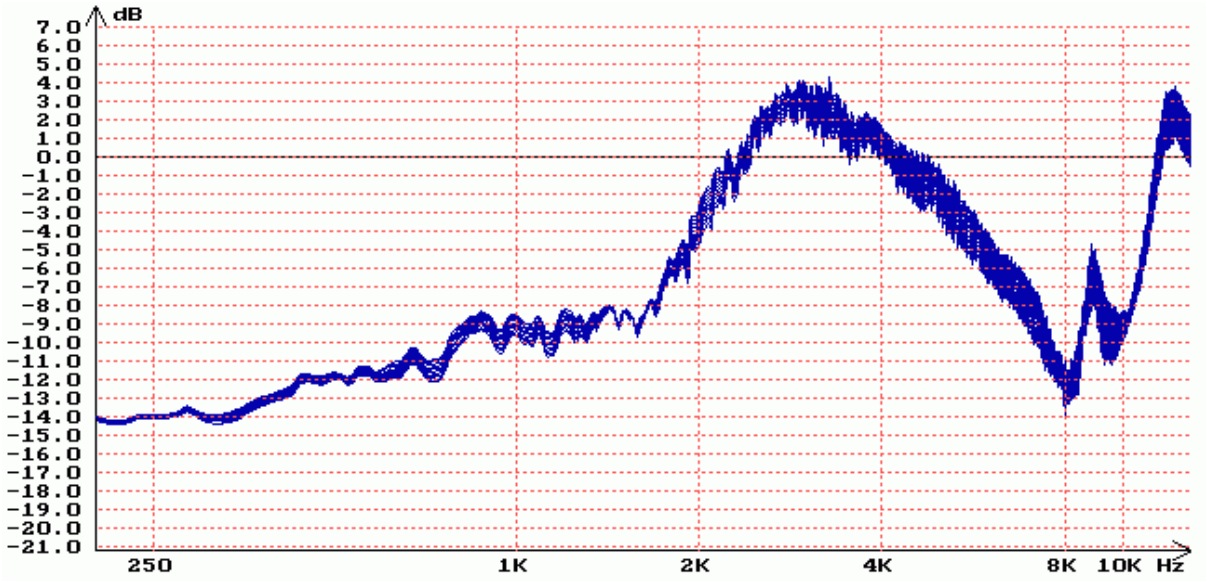


(a)

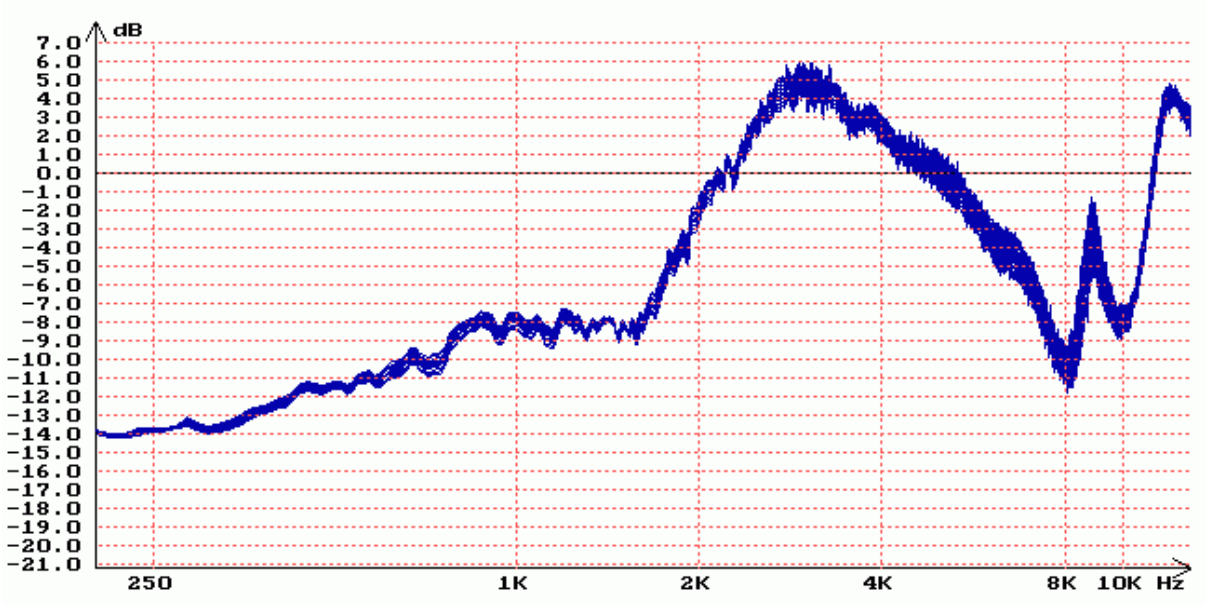


(b)

Fig.53. Two figures show ten plotted HRTFs in 1° resolution in the horizontal plane for comparison.
(a) φ=170°-179° (b) φ=180°-189°.
Note the median plane symmetry to the φ=180°-axe in the local maximum area of the monaural sensitivity. The HRTFs in figure (a) „look like" those from figure (b). Compare with Fig.54.

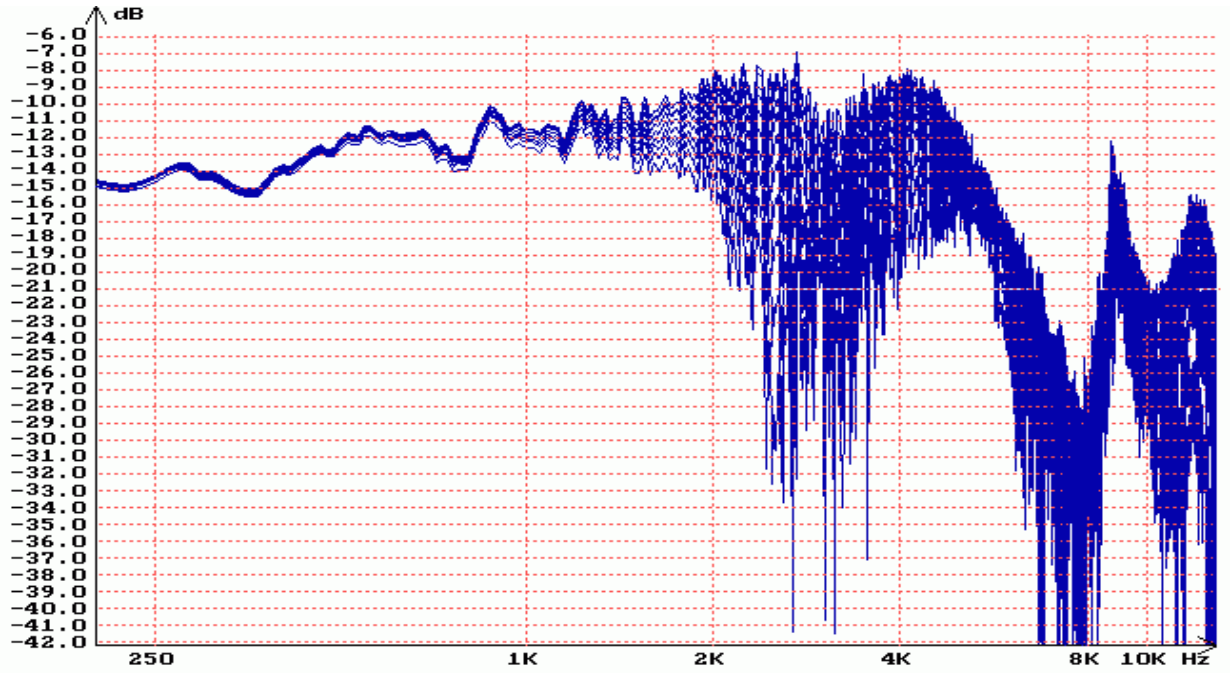


(a)

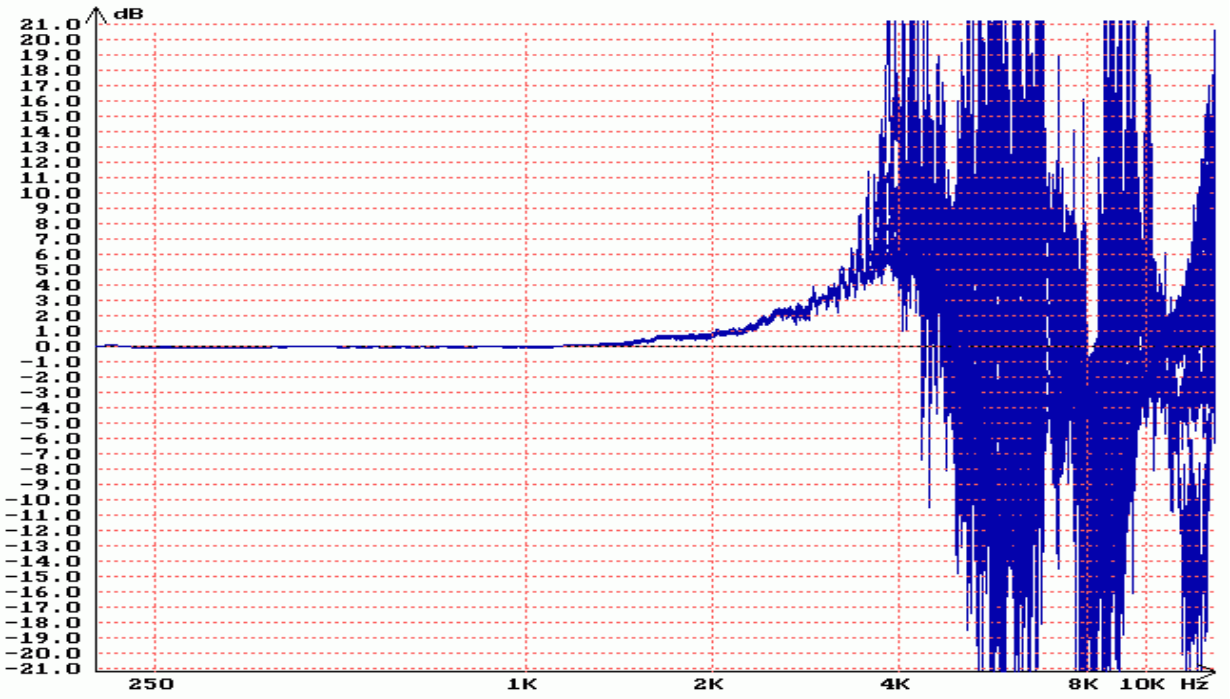


(b)

Fig.54. Two figures show ten plotted HRTFs in 1° resolution in the horizontal plane for comparison.
(a) φ=350°-359° (b) φ=0°-9°.
Note the median plane symmetry to the φ=0°-axe in the binaural sensitivity domain. The HRTFs in figure (a) „look like" those from figure (b). Compare with Fig.53.



(a)



(b)

Fig.55. Minimum of the monaural sensitivity in the head-shadow area. The components above 2 kHz are too variable to allow evaluation of high frequency directional information, but there is no difference below 1600 Hz. Ten HRTFs (a) and calculated HRTFDs (b) from the horizontal plane are plotted between φ=250° and φ=260° in 1 degree resolution. Compare with Fig.49.

In general we can say that the “thickness” of the edges in the HRTFs do not vary significantly which indicates that changing of the azimuth does not really influence the peak-valley structure in the frequency (no shifting) only the height of the peaks and valleys. This is important because the effect of the cap and hair produces relevant shifting in the frequency and create new peaks and valleys. The only frequency shift was observed at the 10 kHz valley, which moved up to 11 kHz and back again.
Fig.56 shows the role of the pinnae filtering effect at frontal incidence. Two HRTFs were measured with and without the artificial pinnae of the head and torso simulator. The sound collecting effect at 3 kHz and above 8 kHz is significant. Average differences between the spectra of the torso below 3 kHz with and without pinnae of 0,86 dB was reported in [24]. Our measurement could not show differences even less than 0,5 dB.

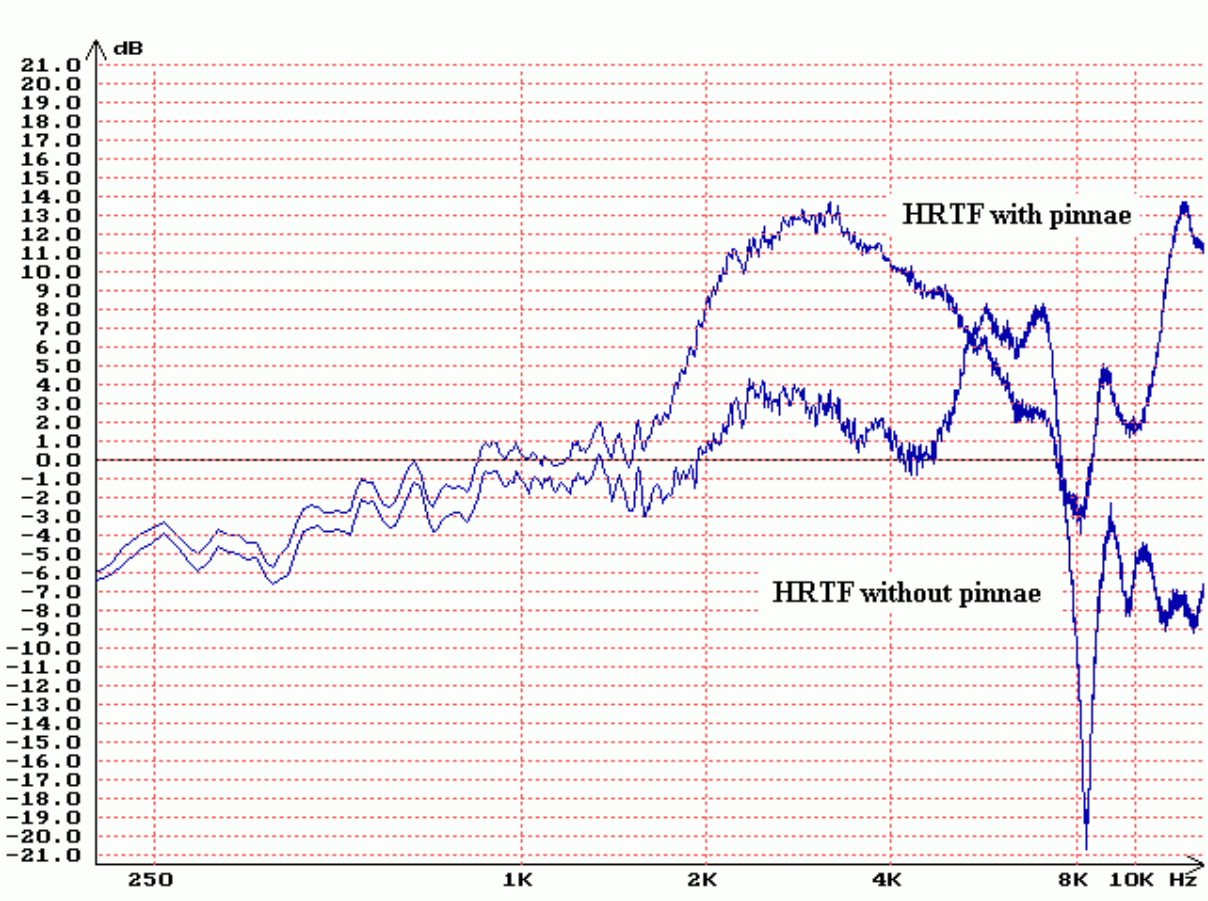


Fig.56. Effect of the pinnae at frontal incidence (φ=δ=0°). Both HRTFs contain the effects of the torso and the head. The reflecting and amplifying effect of the pinnae is clearly visible at the main resonance frequencies of 3, 9 and 11 kHz.

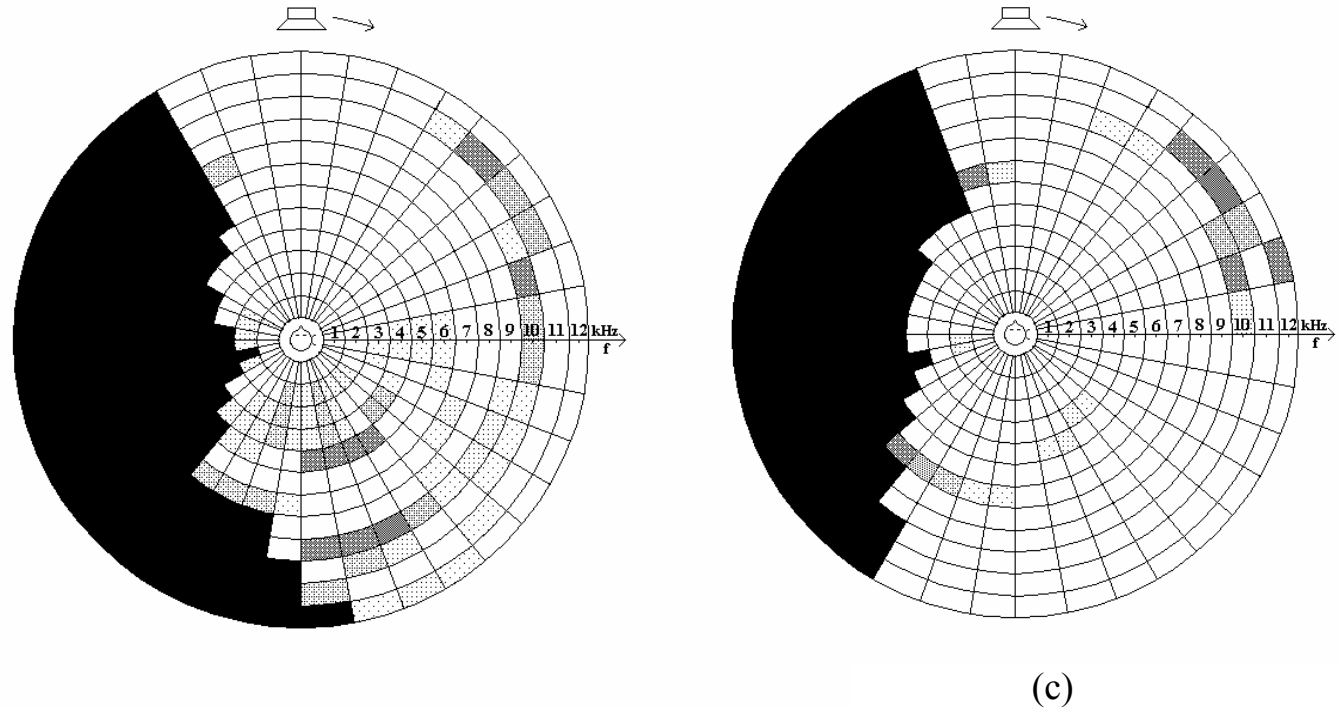
5.4 Effect of the acoustical environment near to the head

The previous section showed how the HRTFs vary by a moving sound source in the horizontal plane. This section analyses the typical effects and deviations in the HRTFs by changing the acoustical environment near to the head. These modifications on the torso are completely usual in our everyday life. If we put on a hat, a cap, get a hair-cut or take off the glasses - we do not recognize any differences in our localization performance or in the sound quality [163].
The resolution in the measurement is 1° horizontal and 5° elevational from –10° up to +30° and we have results from +45° and +60° elevations as well. The objects we have been focused on are: four different kinds of glasses, four different but similar baseball caps and three toupees with different length and haircut. Moreover, some results we obtained from measurements with clothing. All of these objects are quite symmetrical and we made the effort to put them symmetrically to the median plane. The short-cut hair toupee was placed always without covering the pinnae and the long-cut hair always completely covering the pinnae. This fact did not influence the results at all.
During the evaluation of the results we did not observe any significant differences among the different kinds of caps, toupees or glasses. These objects have common properties and thus common effects on the HRTFs, which are represented by the HRTFDs. We present figures containing the averaged effects only for the right ear.
Figures 57 to 67 show the variations in dB of free-field HRTFDs in function of azimuth and frequency as introduced on Fig.47 from -10, -5, 0, 5, 10, 15, 20, 25, 30, 45 and 60 degree elevation for the right ear. The values are unsigned absolute values. The conclusions of the next section were made based on these figures: hair (a), baseball cap (b) and glasses (c).
Differences of zero dB in a HRTFD (white domains) may indicate “real directional information” in the frequency, because these components are constant and independent of the environmental modifications. The monaural sensitivity domain is affected at least.
In the series of Fig.68 to 70 we plotted some representative horizontal plane HRTFDs only as function of the frequency to show regular variations. We zoomed in to the interesting part of the frequency axis and so the plotted figures may have different dB/div value and x-axe scaling. Table 14 contains the appropriate frequency domain and azimuthal steps. Note the different scaling of the x and y-axis before comparing.

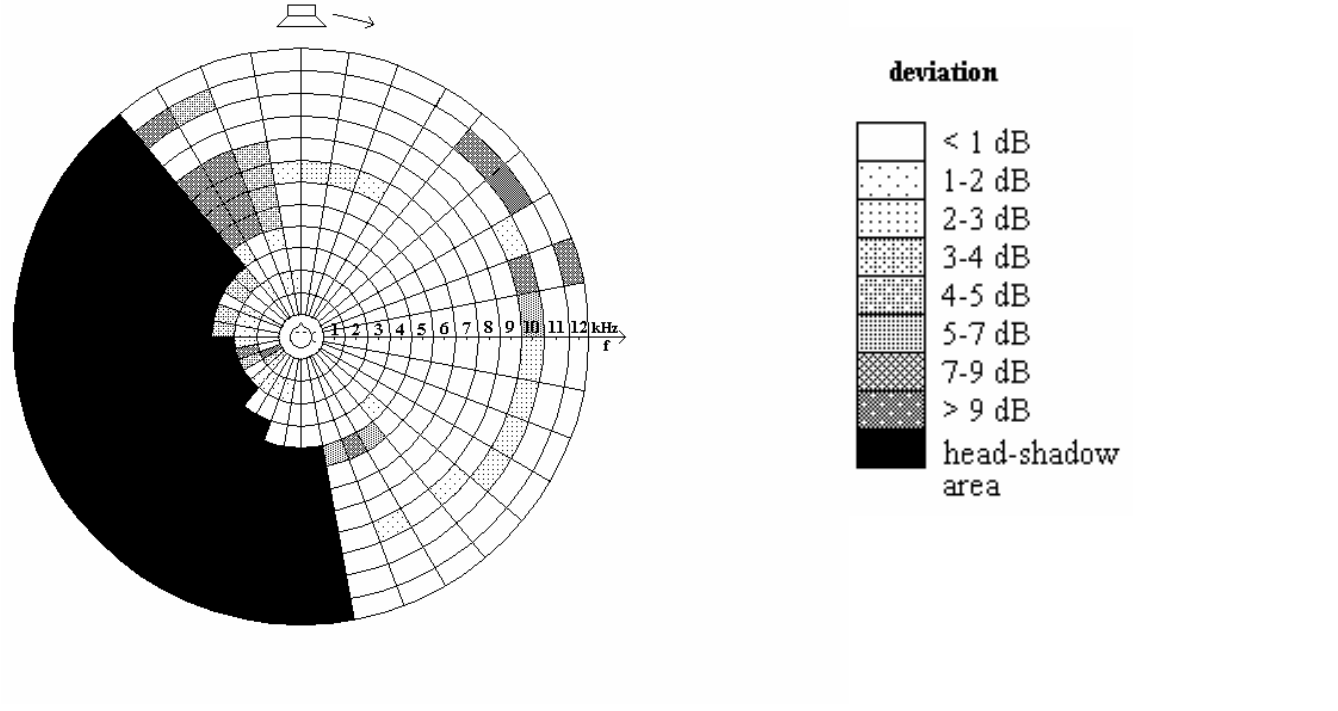
	Fig: (domain; step)	Fig: (domain; step)	Fig: (domain; step)
HAIR	68a: 150°-195°; 5°	68b: 80-170°; 10°	68c: 297°-321°; 3°
GLASSES	69a: 105°-150°; 5°	69b: 265°-300°; 5°	69c: 250°-270°; 2°
CAP	70a: 90°-170°; 10°	70b: 140°-185°; 4°	70c: 230°-260°; 2°

Table 14. Resolution and azimuthal domains. Note the different scaling of the x and y axis.

Effect of the objects can be both amplification and damping, and they influence not only the height of existing peaks and valleys. They produce new frequency components and shifting as well. The evaluation outside the head-shadow area (from –20° to ca. 180°) is made in the entire frequency range and inside only for the low-frequency components. As a general rule, we never found changes under 1600 Hz as expected [5, 12]. The rigid spherical model of the head predicts amplifying effects near to the head due to diffraction even if the head directly blocks the contralateral ear [128]. This suggestion is supported by our measurements below 3 kHz (see later).

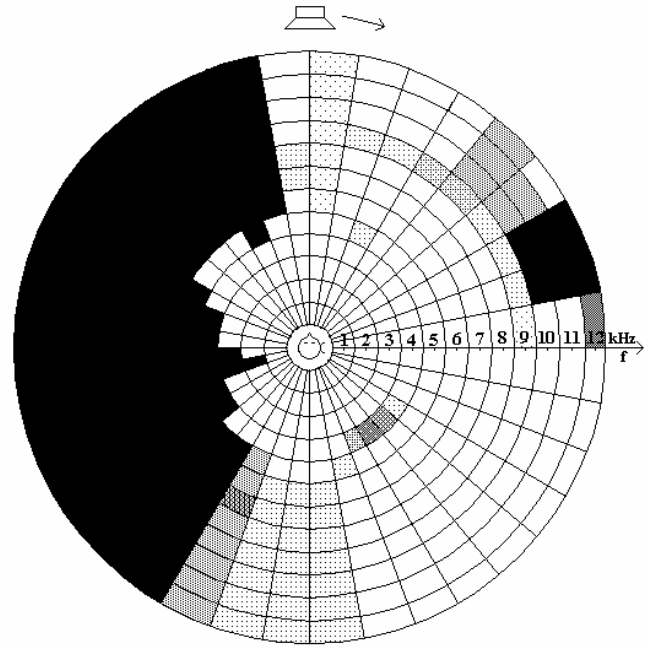


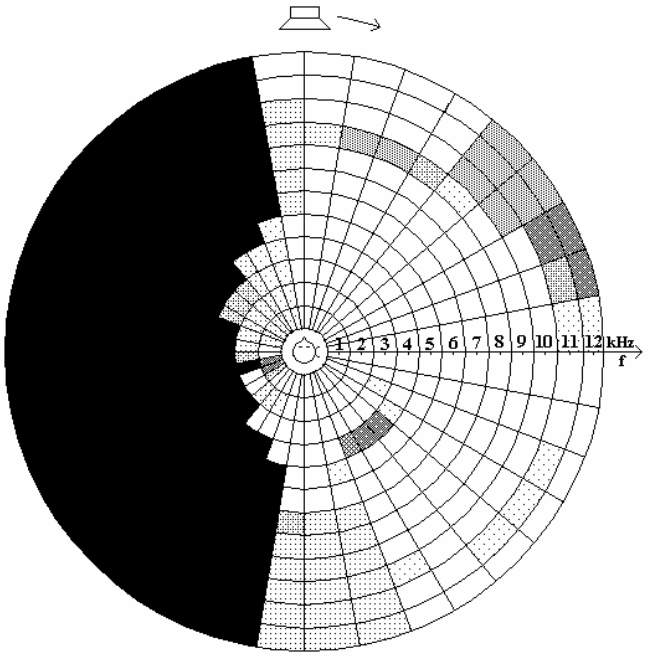
(c)



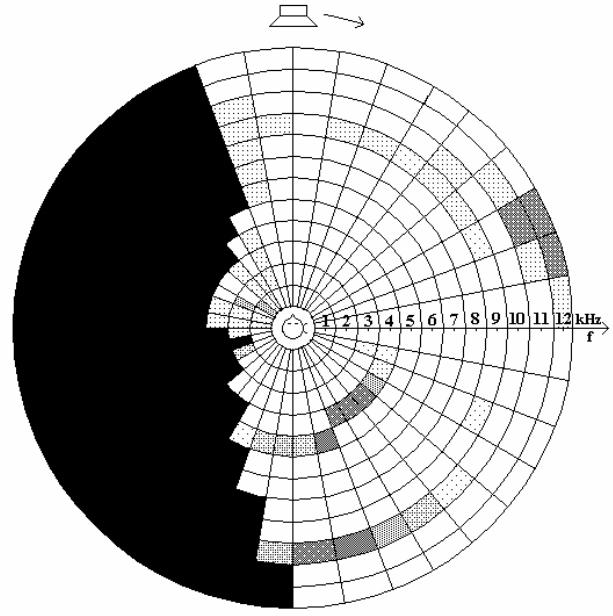
(b)

Fig.57. HRTFDs showing the effect of hair (a), baseball cap (b) and glasses (c) from the elevation –10°.





(c)



(b)

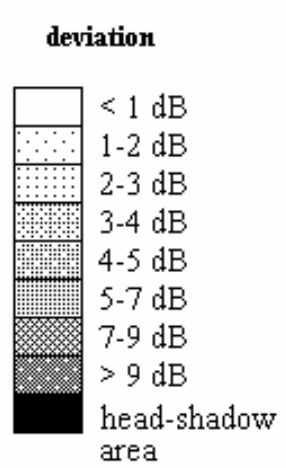


Fig.58. HRTFDs showing the effect of hair (a), baseball cap (b) and glasses (c) from the elevation –5°.

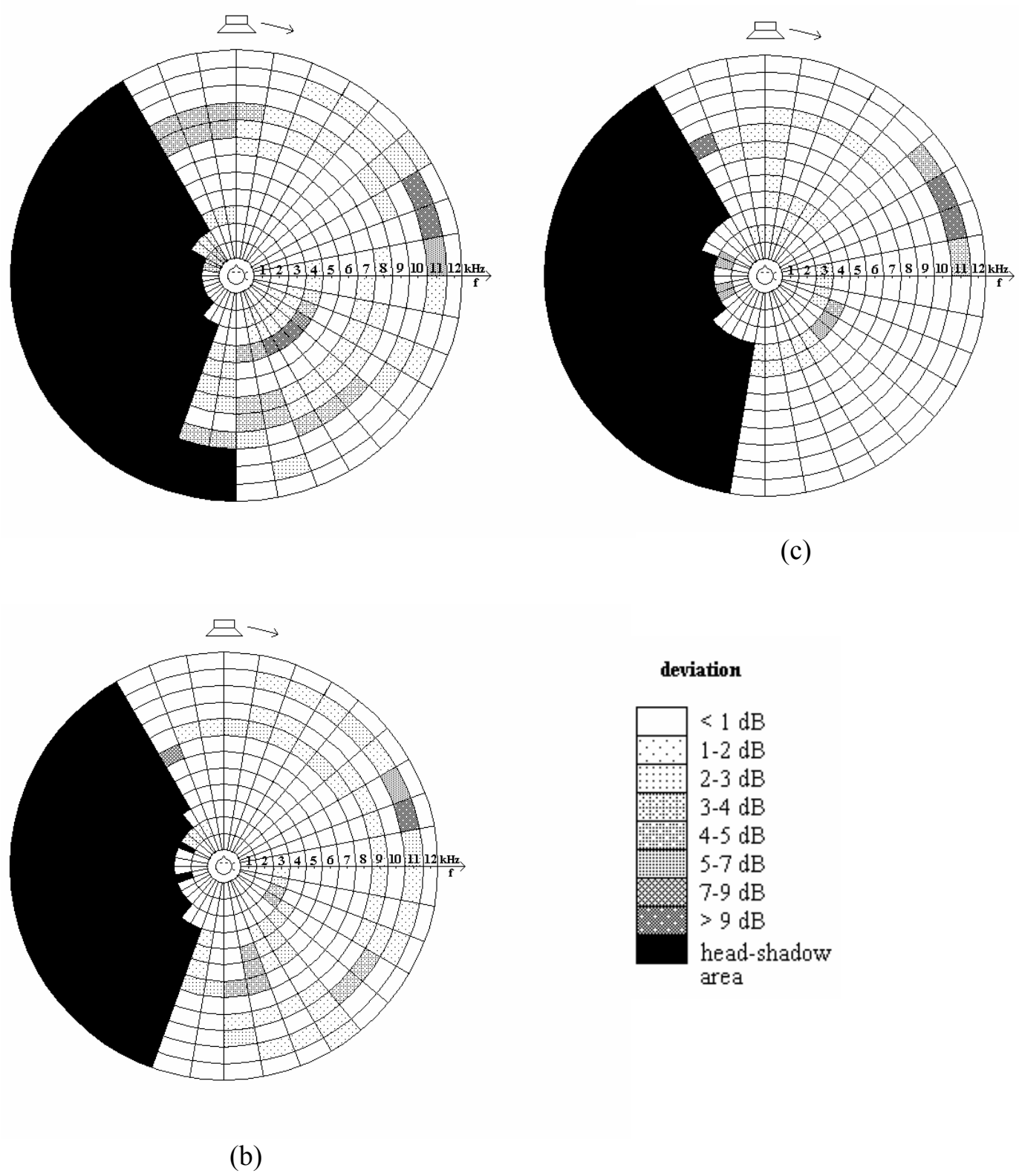
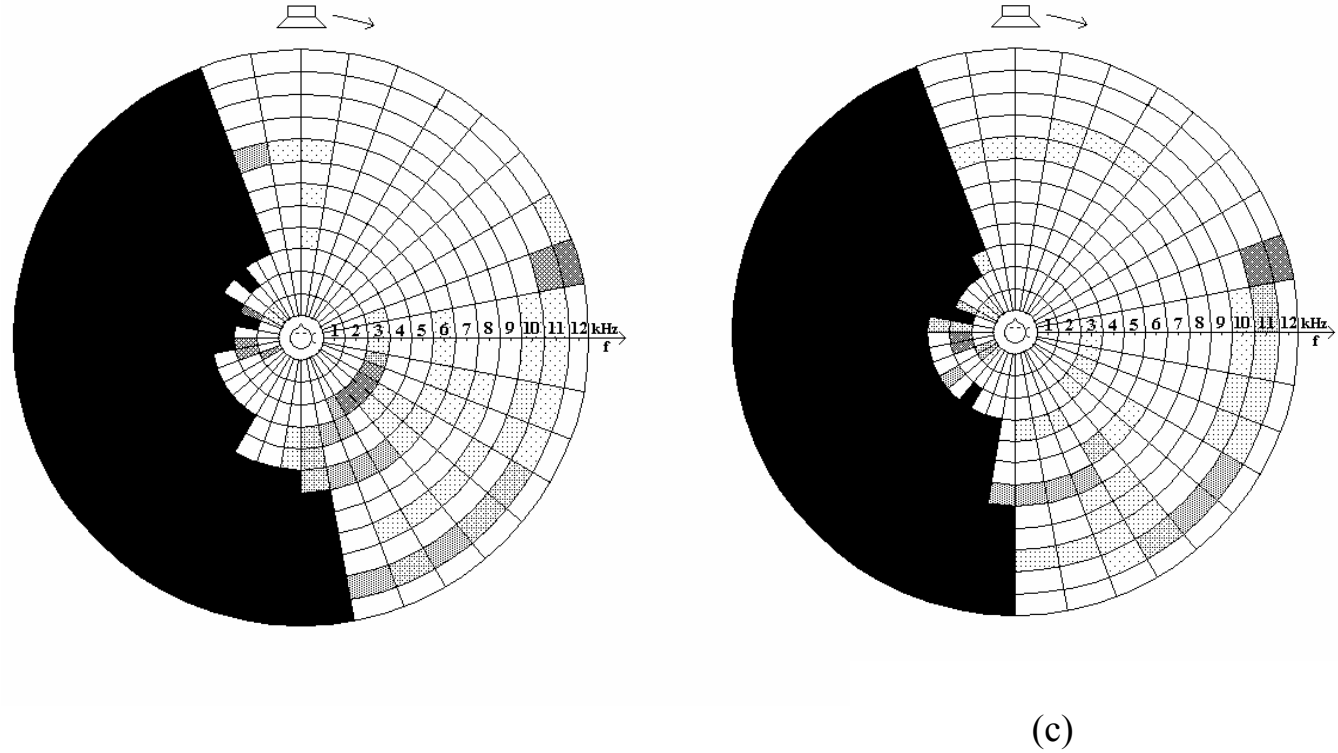
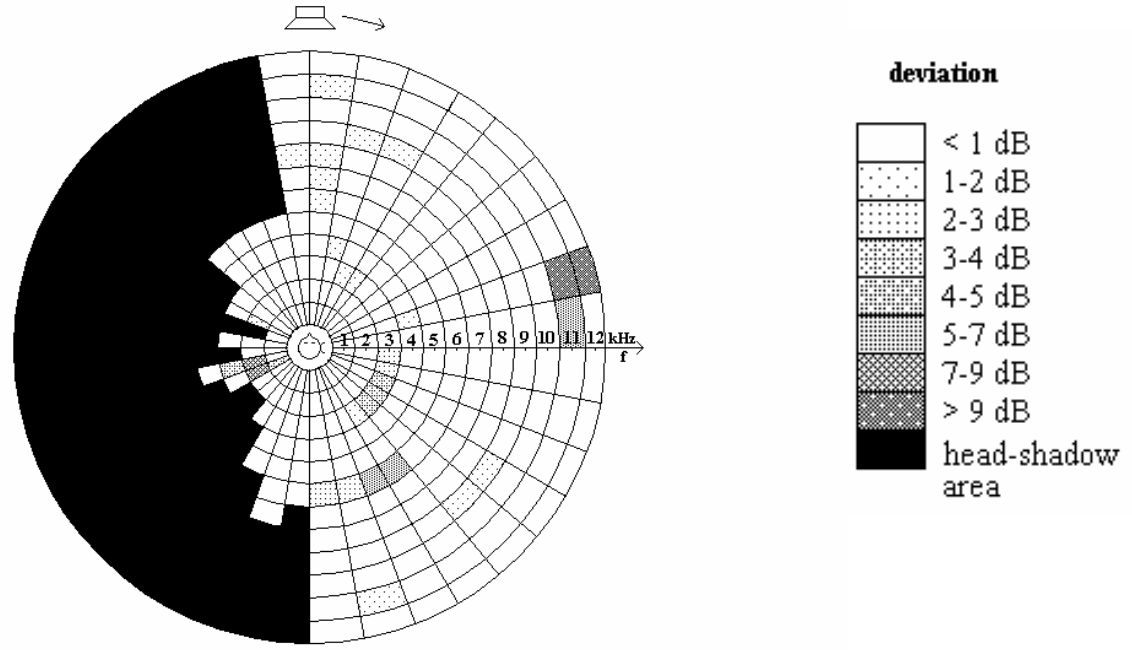


Fig.59. HRTFDs showing the effect of hair (a), baseball cap (b) and glasses (c) from the elevation 0° (horizontal plane).

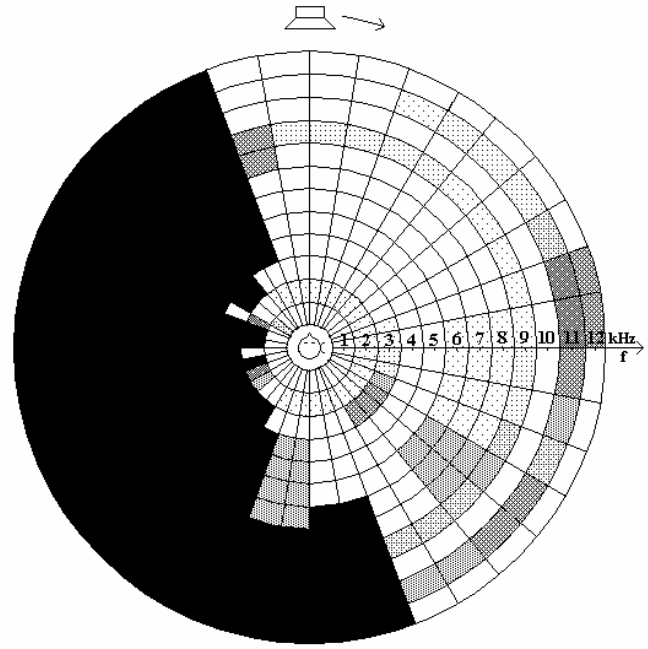


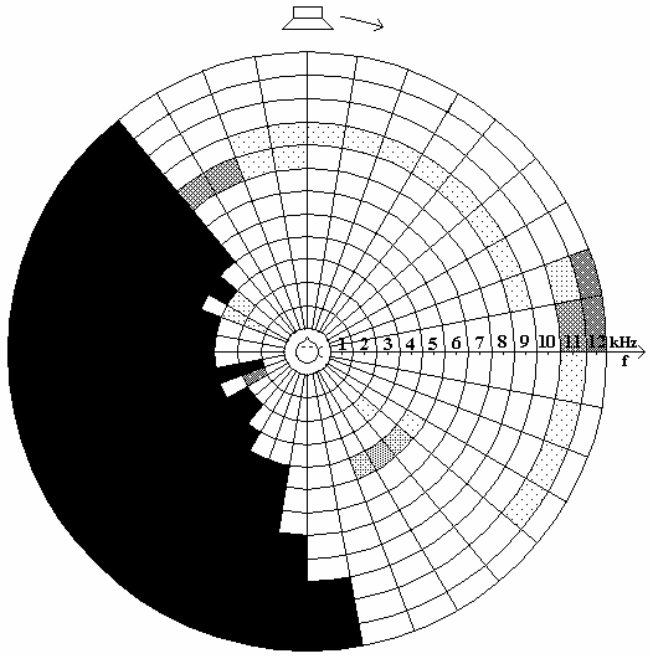
(c)



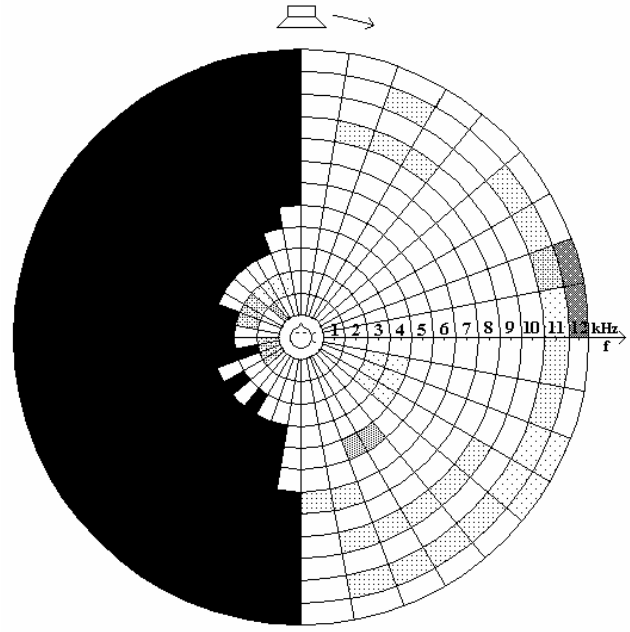
(b)

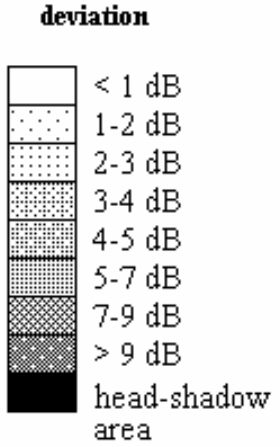
Fig.60. HRTFDs showing the effect of hair (a), baseball cap (b) and glasses (c) from the elevation +5°.





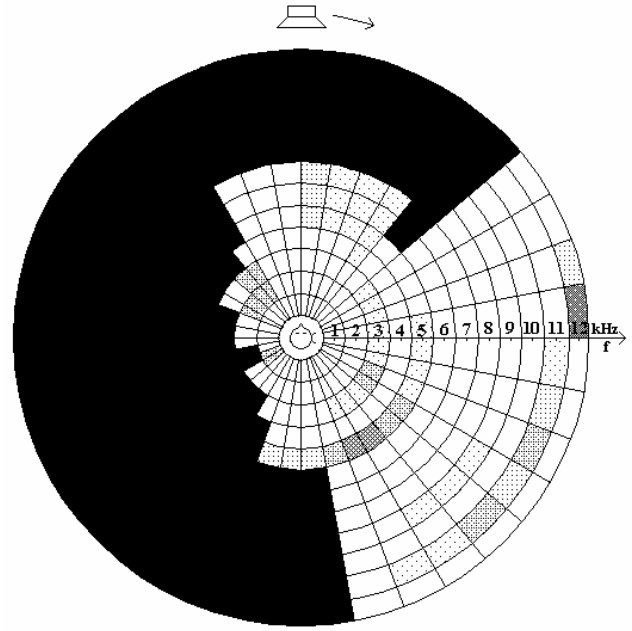
(c)

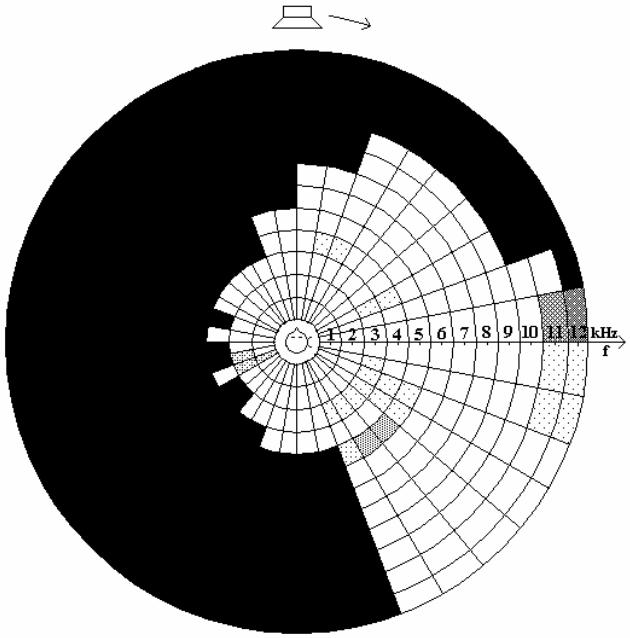




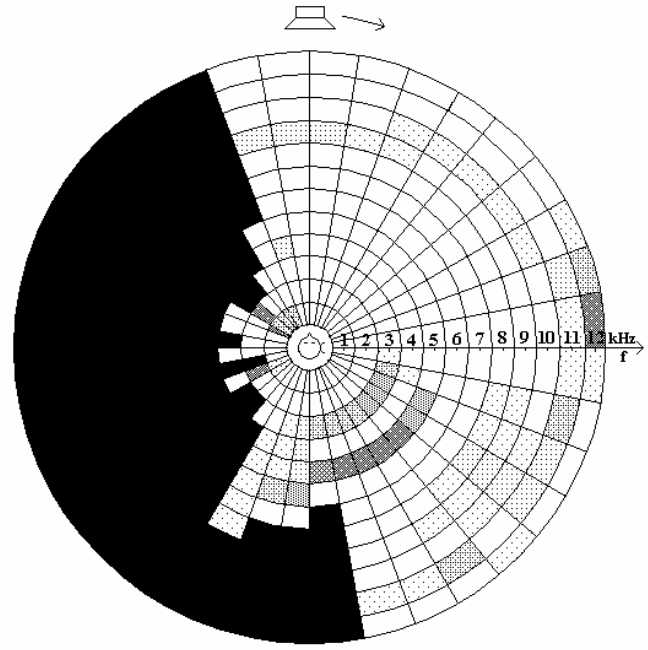
(b)

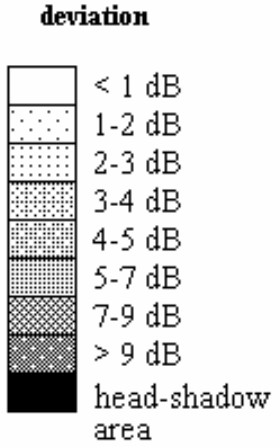
Fig.61. HRTFDs showing the effect of hair (a), baseball cap (b) and glasses (c) from the elevation +10°.





(c)





(b)

Fig.62. HRTFDs showing the effect of hair (a), baseball cap (b) and glasses (c) from the elevation +15°.

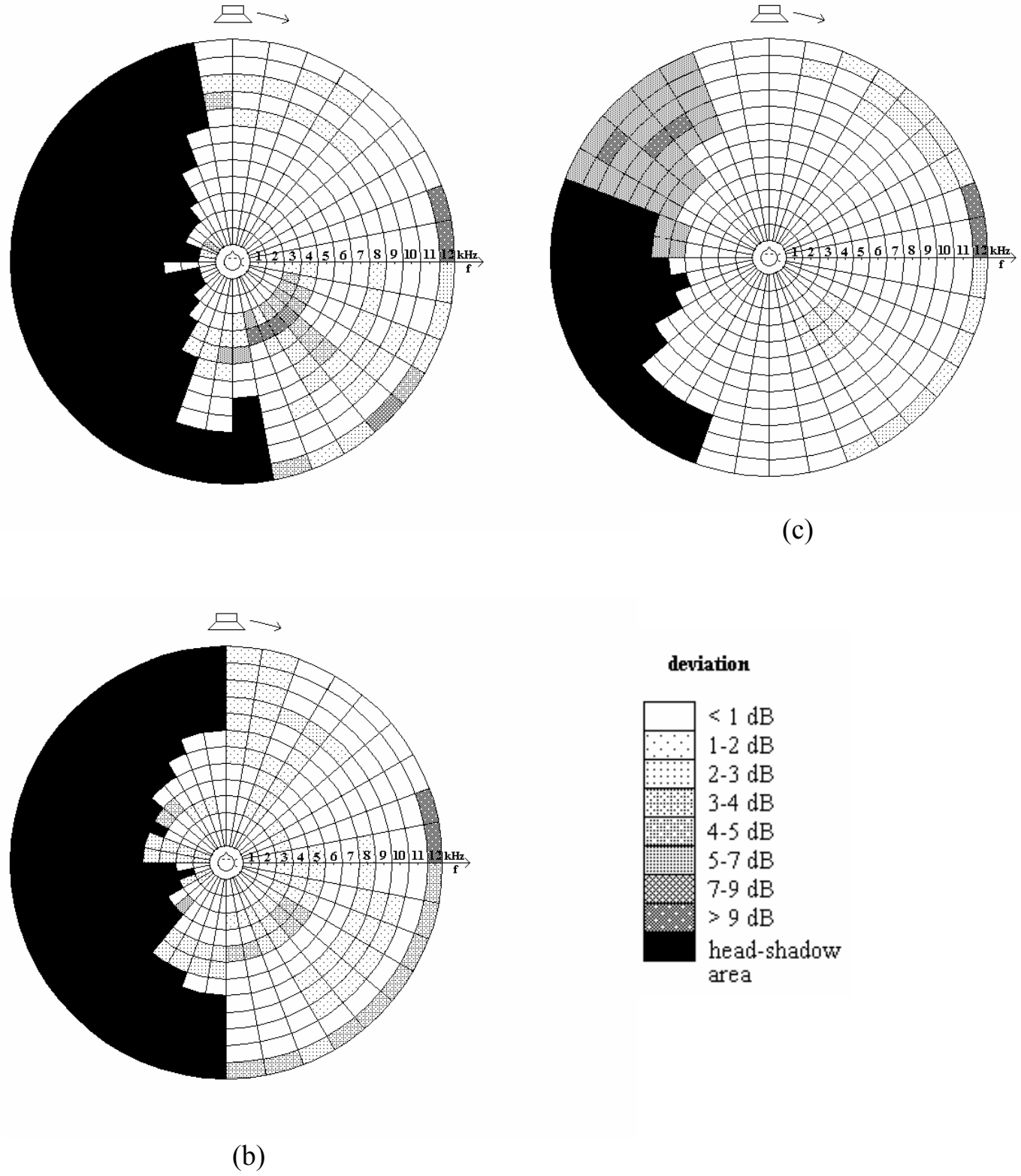
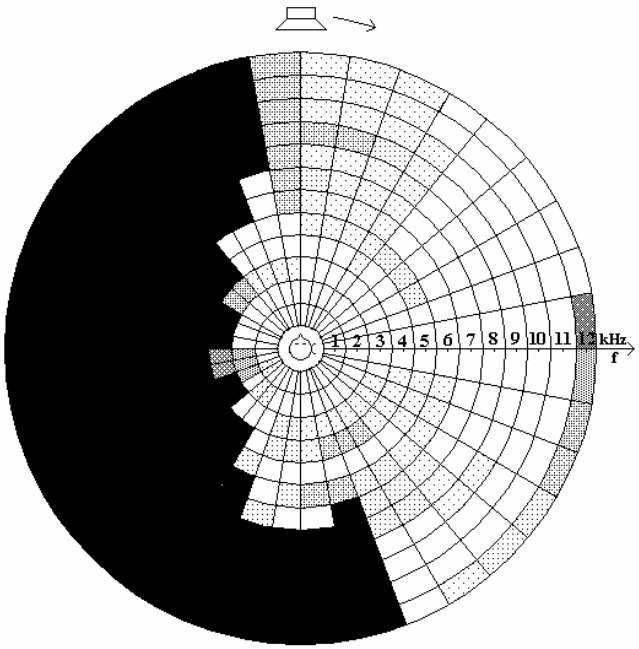
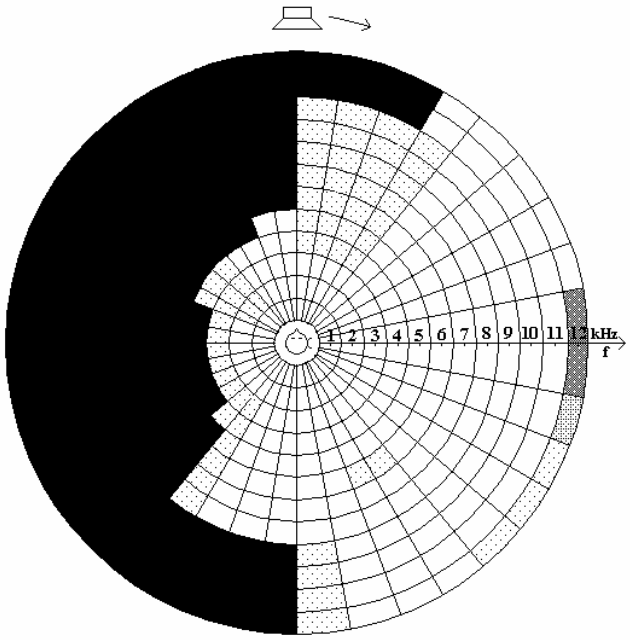
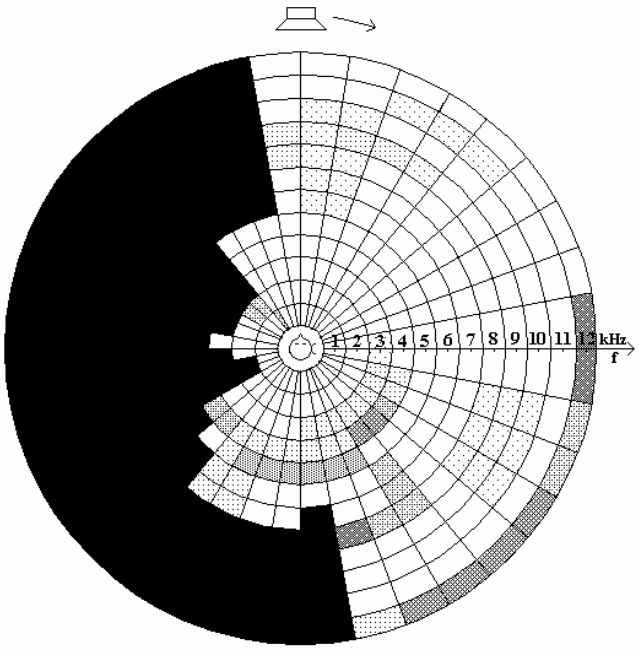


Fig.63. HRTFDs showing the effect of hair (a), baseball cap (b) and glasses (c) from the elevation +20°.





(c)



(b)

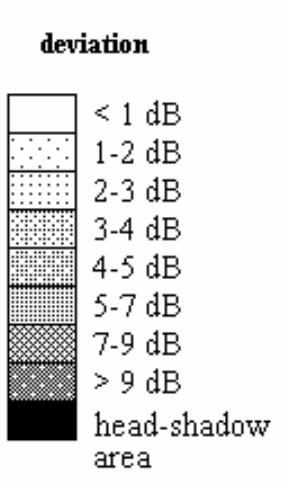
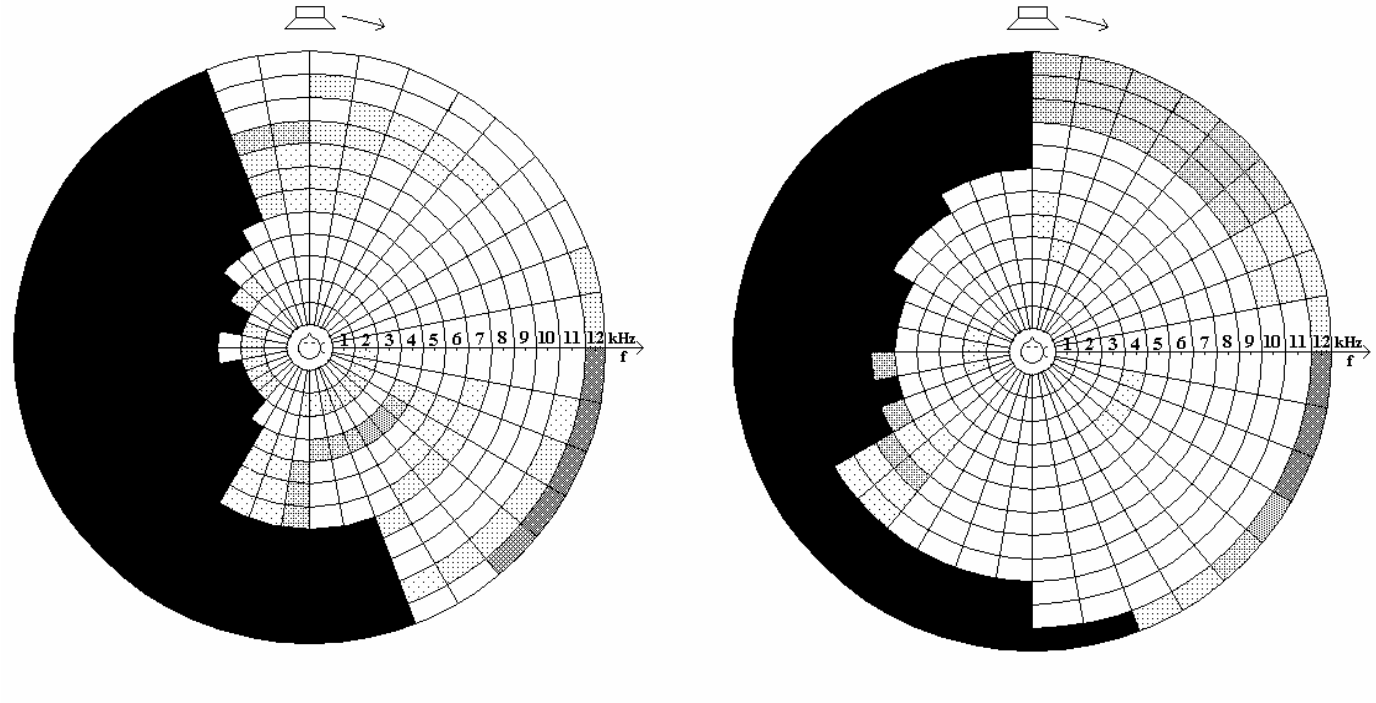
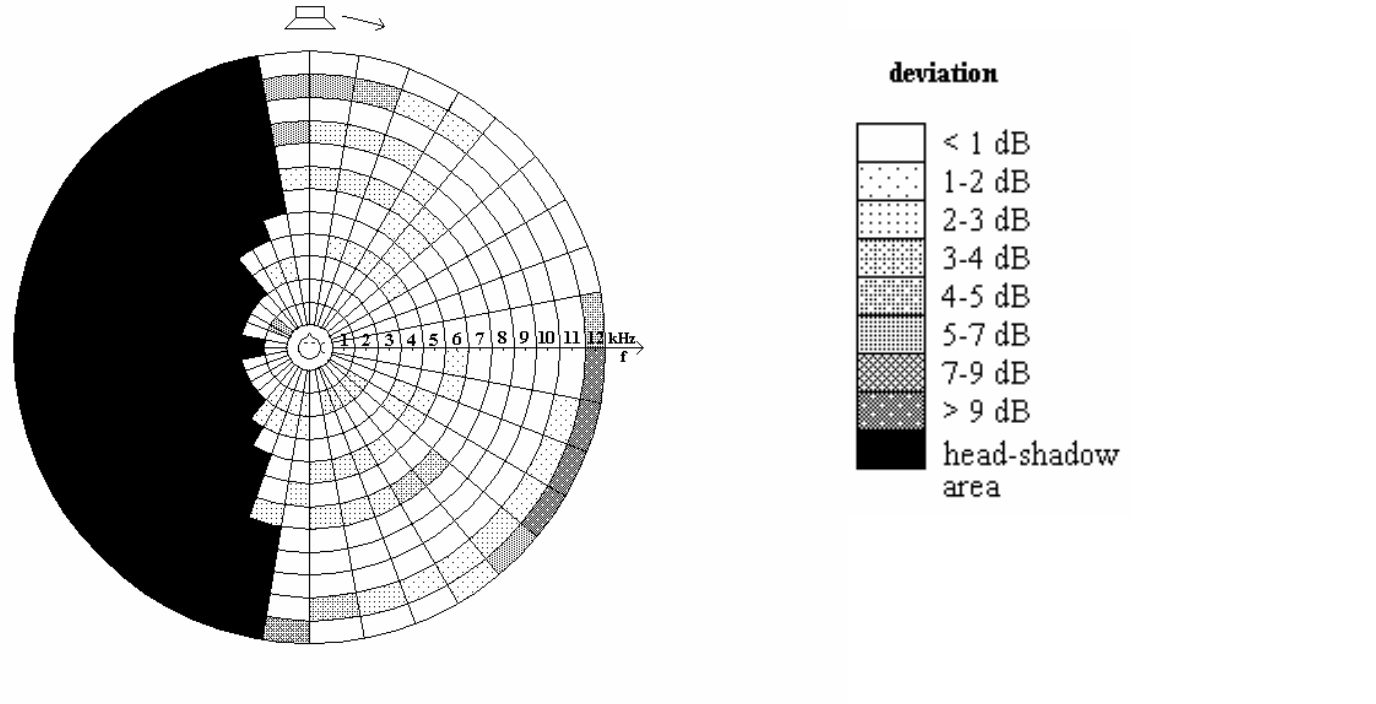


Fig.64. HRTFDs showing the effect of hair (a), baseball cap (b) and glasses (c) from the elevation +25°.

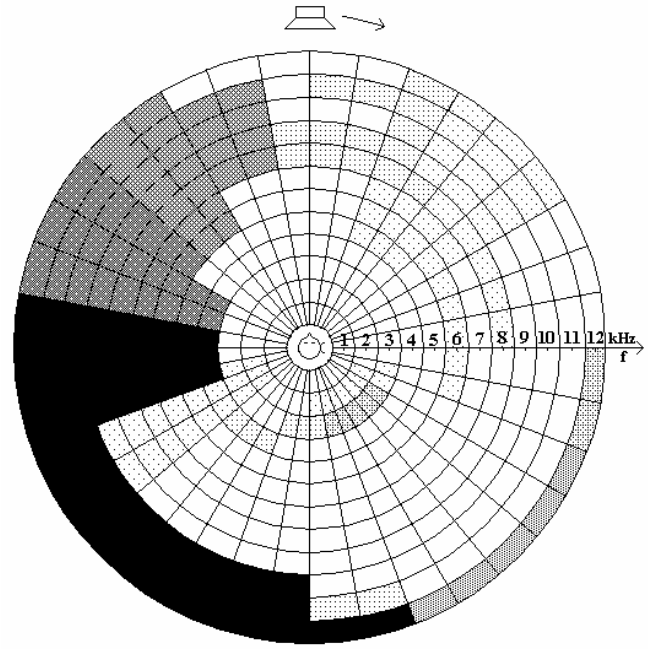


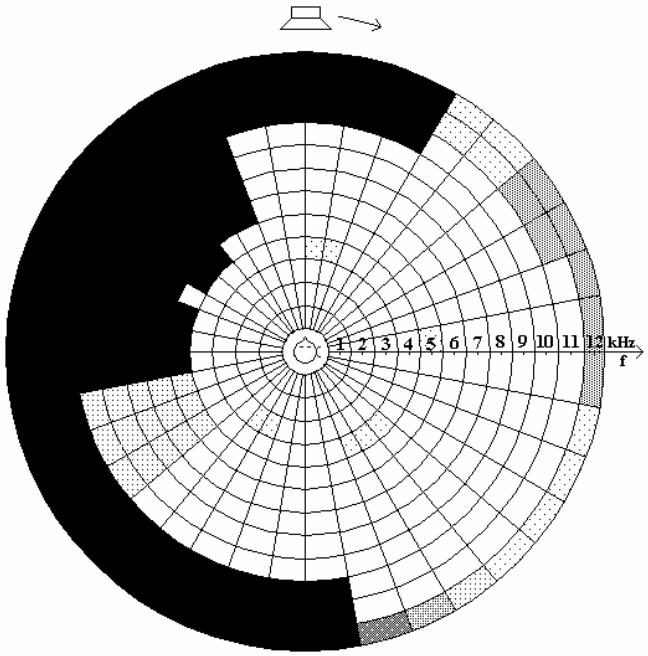
(c)



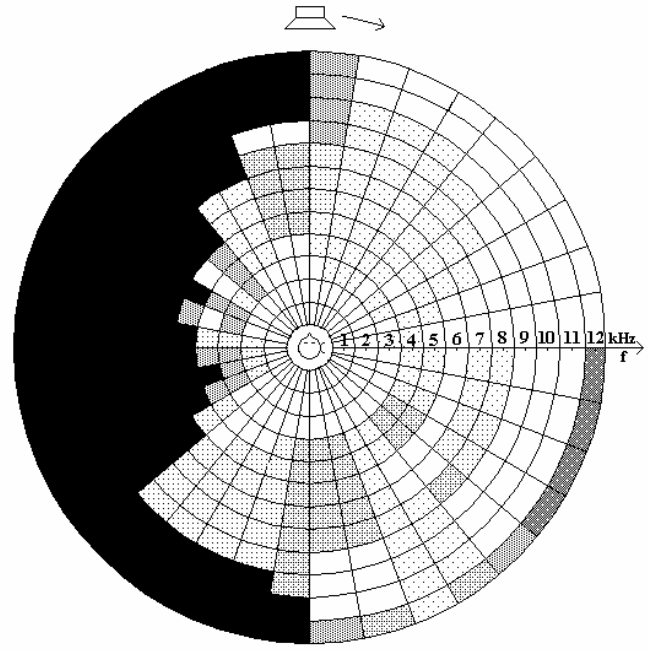
(b)

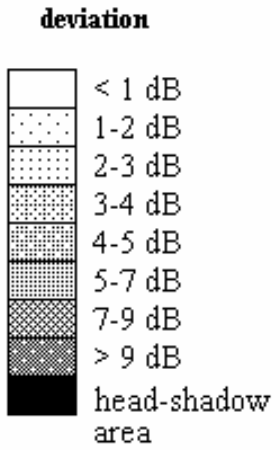
Fig.65. HRTFDs showing the effect of hair (a), baseball cap (b) and glasses (c) from the elevation +30°.





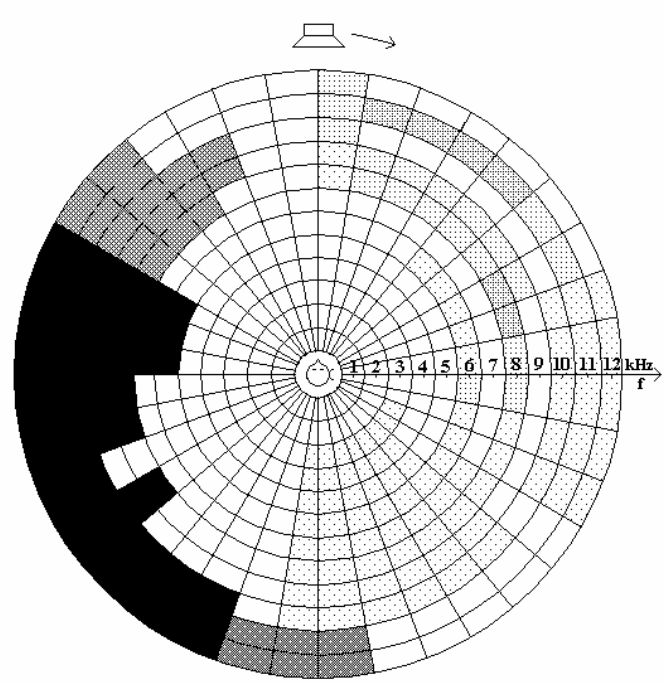
(c)



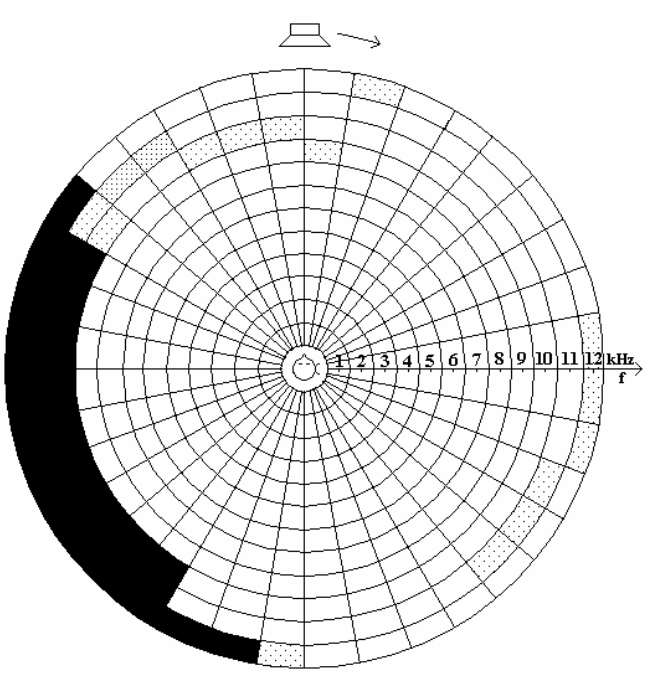


(b)

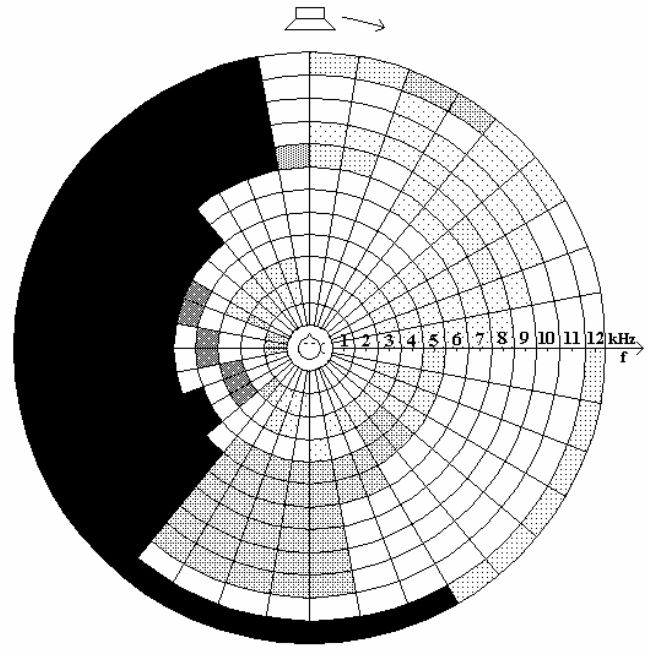
Fig.66. HRTFDs showing the effect of hair (a), baseball cap (b) and glasses (c) from the elevation +45°.



(a)



(c)



(b)

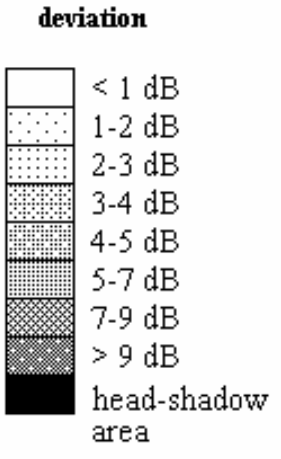


Fig.67. HRTFDs showing the effect of hair (a), baseball cap (b) and glasses (c) from the elevation +60°.

5.4.1 Hair

Toupees can be difficult to place onto the head symmetrically to the median plane. The evaluation was based on HRTFDs from 11 elevational positions, 1° horizontal steps, two channel results and using three different haircuts.
The monaural sensitivity domain is from –20º to +90 degrees. Hair produces a broadband and significant effect, mostly at 9, 10, and 11 kHz. The most important domain is between 4-5 kHz, where the differences are large and permanent as the source is moving in the horizontal plane independent of the elevational position. At lower elevations (up to 20°) the 3,5 kHz components, at higher elevations (above 20°) the 2,5 and the 2,8 kHz components are influenced as well. Amplifying of hair at 10 kHz was also found in [69].
The head shadow area is extended to 200°-340° and differences up to 10 dB appear at 1,8 and 2,2 kHz. Above +30° elevation this effect is less significant. At +45° and +60° we cannot find a clear monaural sensitivity domain. The deviations and the expansion of the shadowed domain decrease (Fig. 68a-68c). *Shaw* suggested that hair produce differences between 3-5 dB [6].

5.4.2 Glasses

The differences appearing at 9, 10 and 11 kHz and the changes at 4-5 kHz are less significant than by the cap and hair. Glasses are small, thin objects, they may influence the HRTFDs at higher frequencies. The elevational-depended components at 3.2 and 3.5 kHz are reduced and decreased as the elevation increases (Fig. 69a-69c).

5.4.3 Baseball cap

The same high frequency components are mainly influenced (9, 10, 11 kHz) in the monaural sensitivity domain. Up to +15° elevation the differences at 3.5-5 kHz are the most significant. Frequency shift of the peaks and valleys occur depending on azimuth. Above +15° elevation the affected regions are divided into separate domains: 3-3.2 kHz and around 5 kHz. Because of the shadowing effect of the visor above +20° elevation, the HRTFs vary too rapidly and random to evaluate components above 8 kHz. The domains at 3, 4-6, 7, 9, and 12 kHz are mainly disturbed, but the head-shadow area is not influenced very

much. The affected low frequency components are 1.6, 1.8, 2.2 and 2.5 kHz (Fig. 70a-70c).

5.4.4 Clothing

Clothing has a common damping effect due to sound absorption. A thin T-shirt does not influence the transmission, but a thick shirt or coat has a damping up to 2-3 dB at 2-4 kHz, 3 dB at 8 kHz and 2 dB at 11 kHz. In the head shadow area the low frequency components at 1.5, 1.8, and 2.5 kHz show +2 and +4 dB amplification. *Tarnóczy* observed the same effect: damping of 2-4 dB at 90° and 6 dB at 180°, amplifying at 6 kHz [74]. The body below 10 kHz is not significant, but clothing have influence above 1500 Hz but mostly above 5 kHz [75]. Clothing has smoothing effects and cause ±3 dB at 1200 Hz [69]. Both the ITD and ILD show differences between measurements made with the bare torso and those with a clothed torso and it seems to be not possible to generalize about the effect of clothing [51].

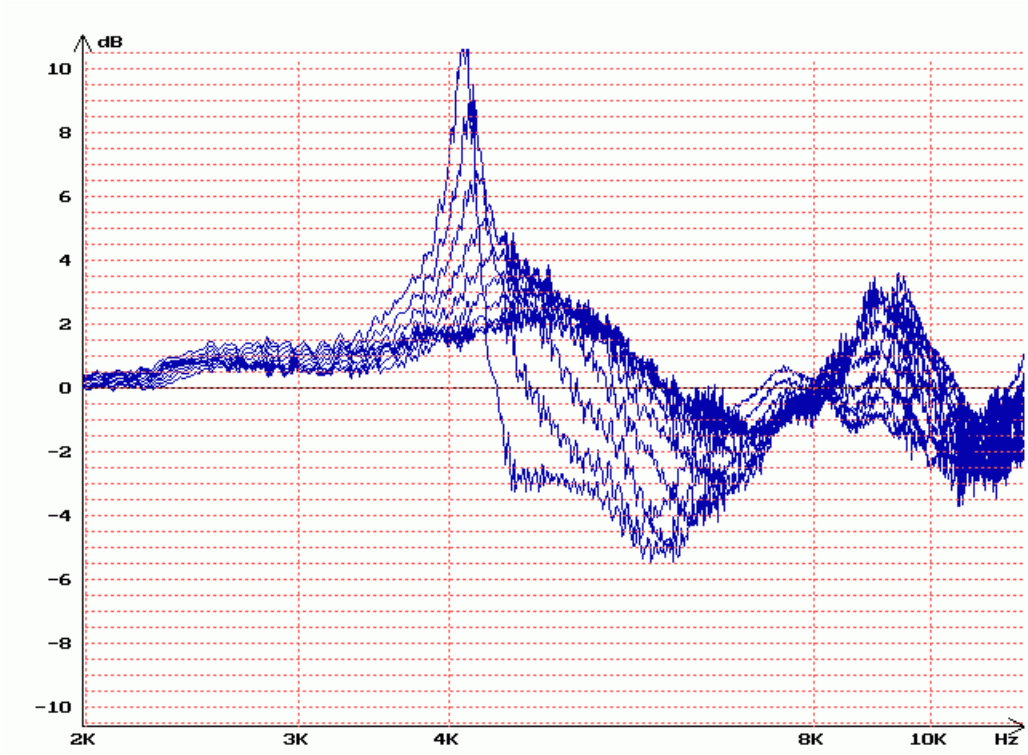


Fig.63a. Horizontal plane HRTFDs as the function of frequency between φ=150°-195° in 5° steps using hair according to Table 14. Note the different scaling of the axes before comparing.

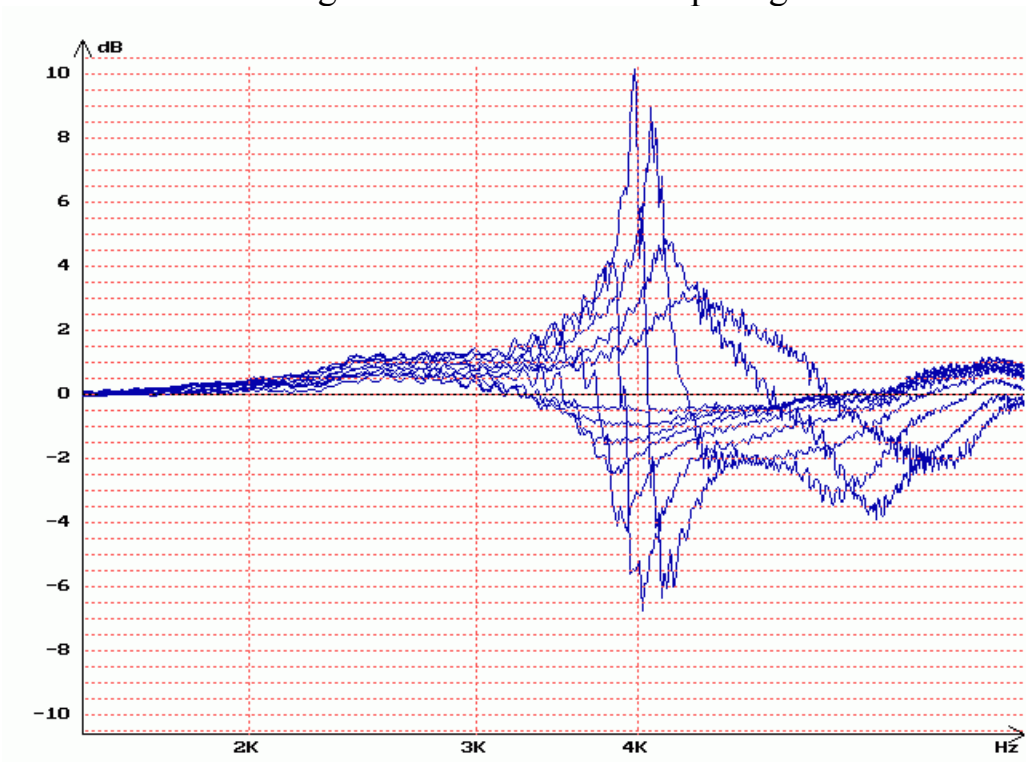


Fig.63b. Horizontal plane HRTFDs as the function of frequency between φ=80°-170° in 10° steps using hair according to Table 14. Note the different scaling of the axes before comparing.

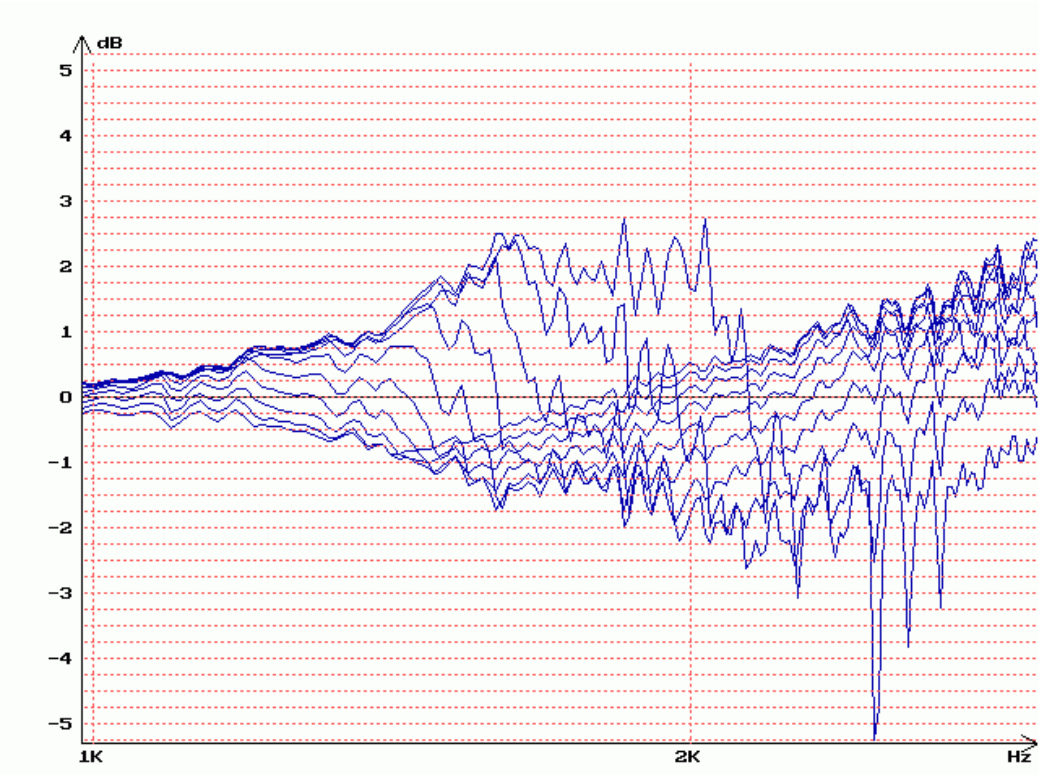


Fig.63c. Horizontal plane HRTFDs as the function of frequency between φ=297°-321° in 3° steps using hair according to Table 14. Note the different scaling of the axes before comparing.

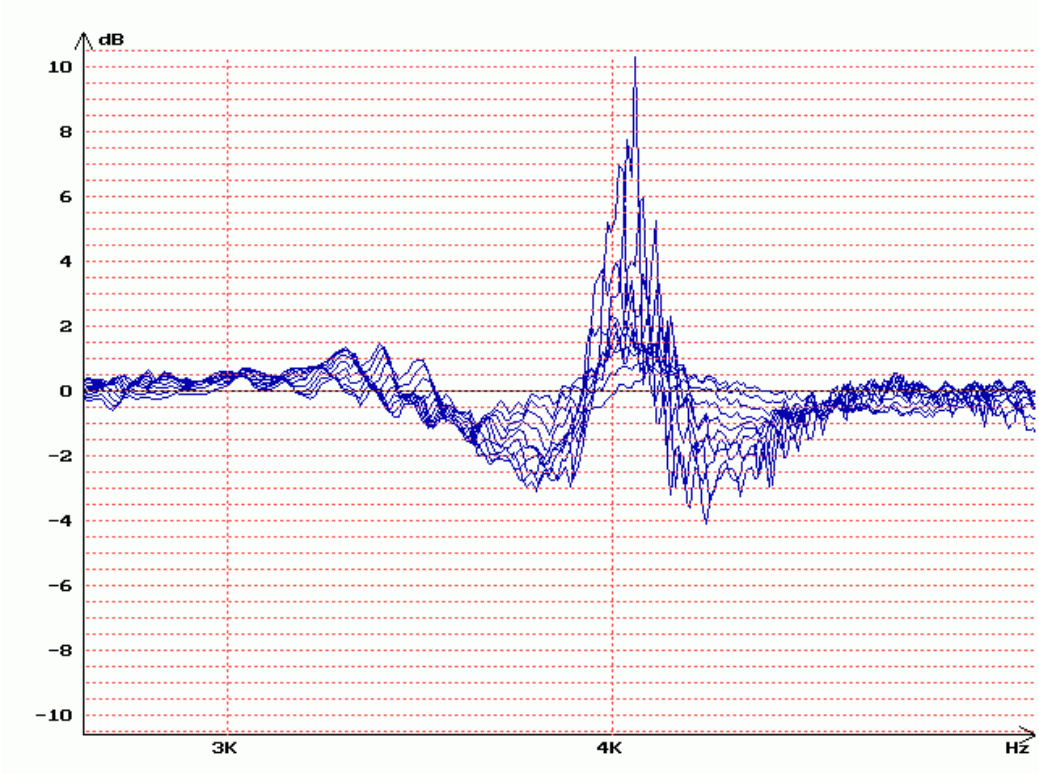


Fig.64a. Horizontal plane HRTFDs as the function of frequency between φ=105°-150° in 5° steps using glasses according to Table 14. Note the different scaling of the axes before comparing.

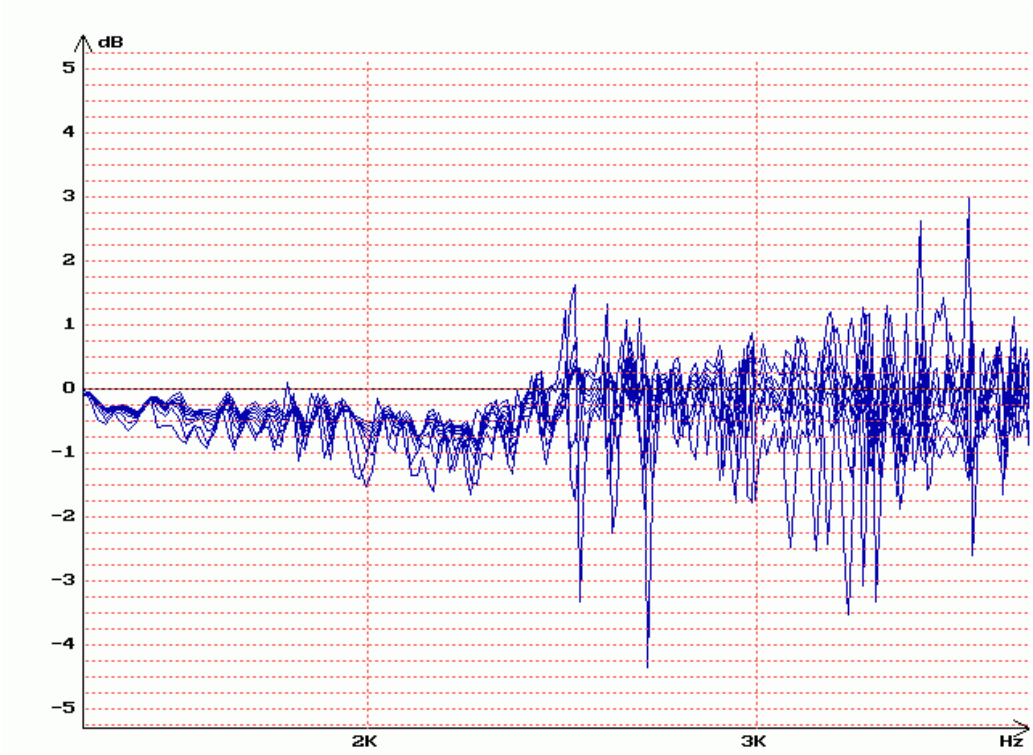


Fig.64b. Horizontal plane HRTFDs as the function of frequency between φ=265°-300° in 5° steps using glasses according to Table 14. Note the different scaling of the axes before comparing.

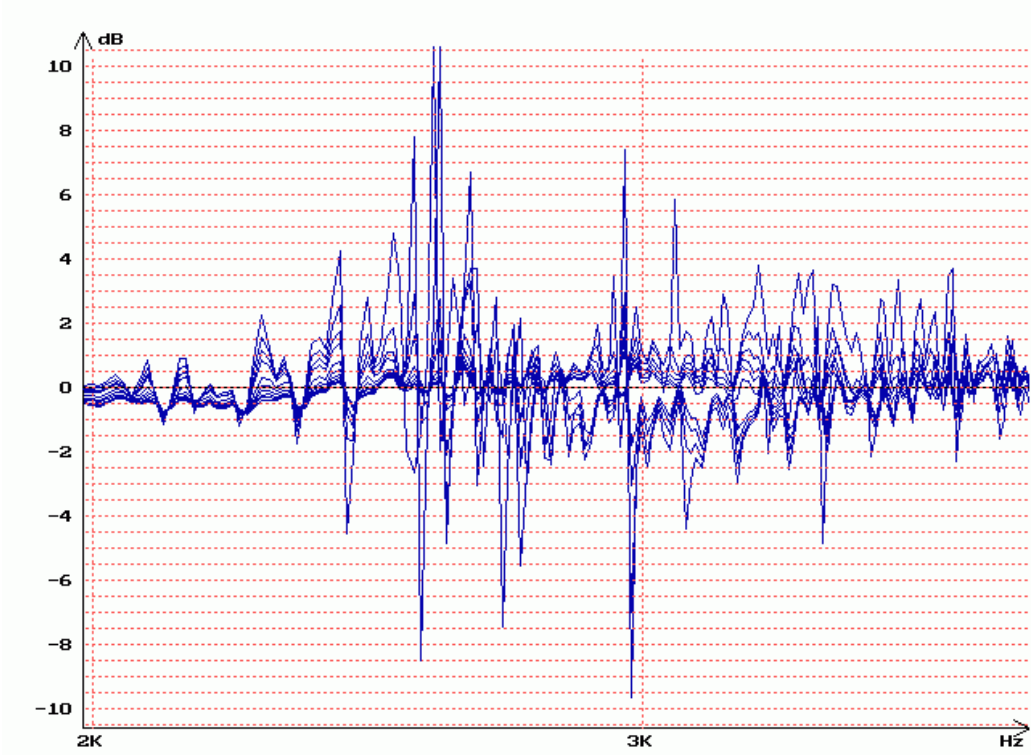


Fig.64c. Horizontal plane HRTFDs as the function of frequency between φ=250°-270° in 2° steps using glasses according to Table 14. Note the different scaling of the axes before comparing.

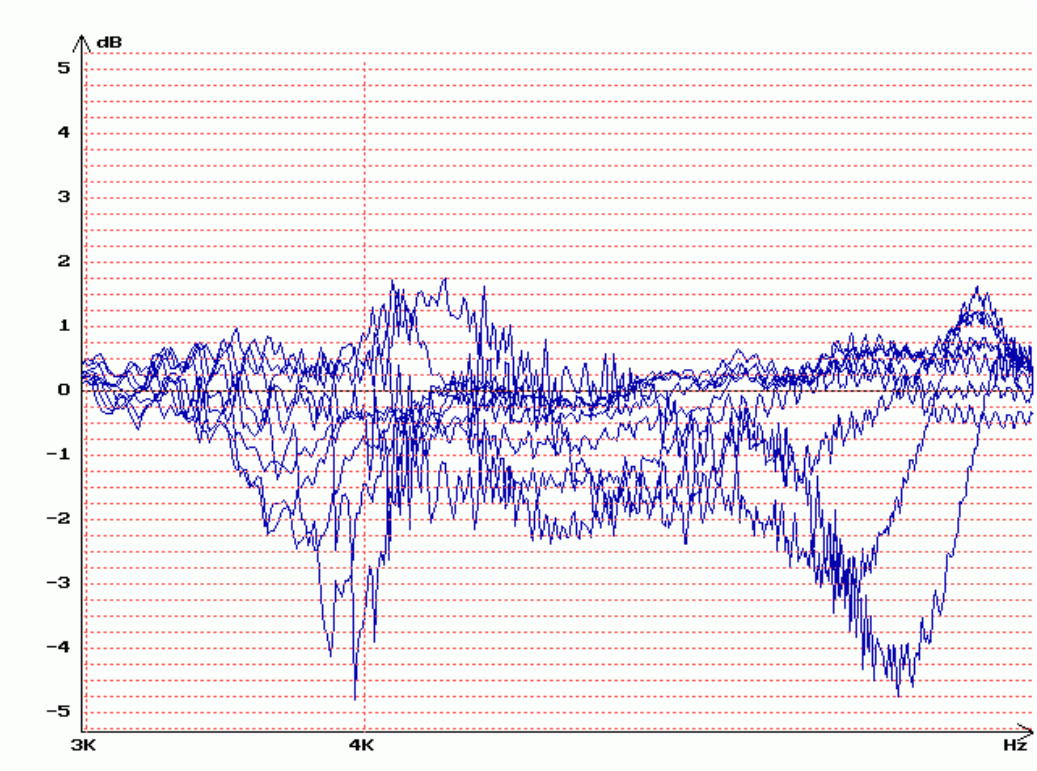


Fig.65a. Horizontal plane HRTFDs as the function of frequency between φ=90°-170° in 10° steps using baseball cap according to Table 14. Note the different scaling of the axes before comparing.

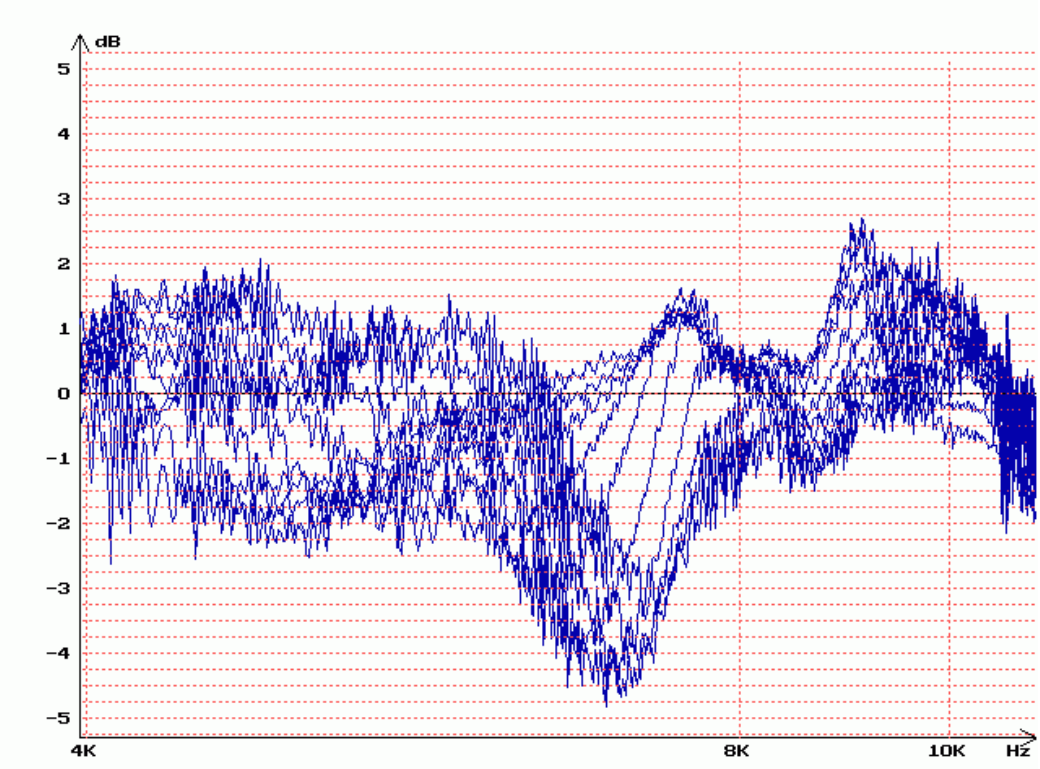


Fig.65b. Horizontal plane HRTFDs as the function of frequency between φ=140°-185° in 4° steps using baseball cap according to Table 14. Note the different scaling of the axes before comparing.

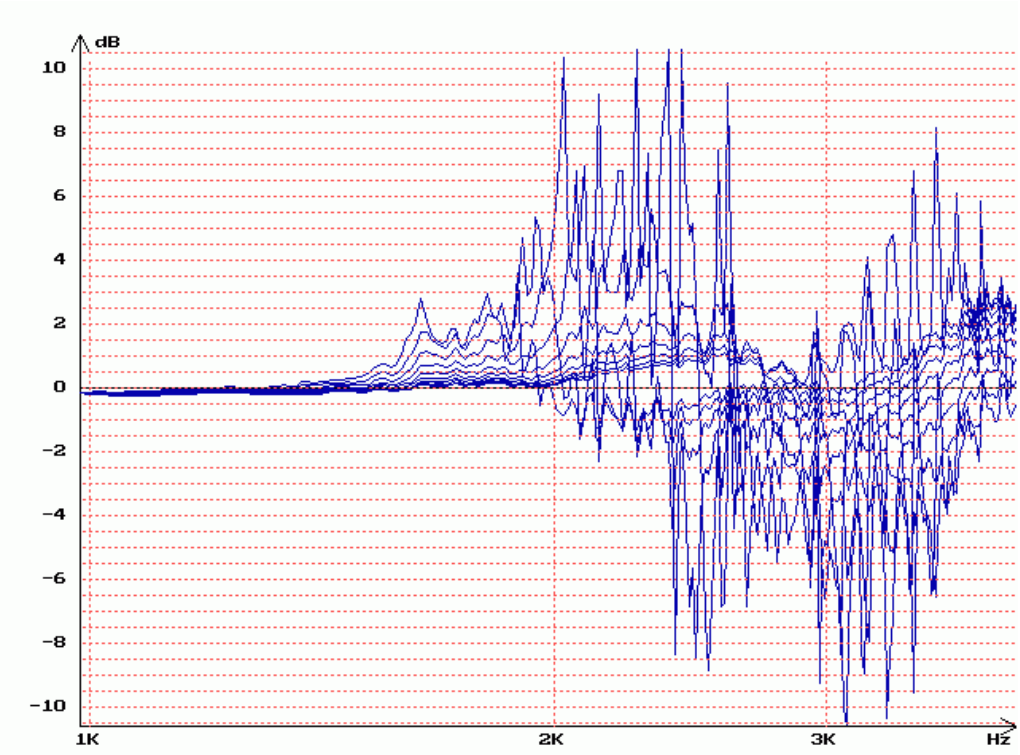


Fig.65c. Horizontal plane HRTFDs as the function of frequency between φ=230°-260° in 2° steps using baseball cap according to Table 14. Note the different scaling of the axes before comparing.

6 Discussion

6.1.1 The sensitivity domains

We have already seen that the most accurate and sensitive domain of the hearing system is where the overall HRTF level is the highest. The lateral ear has more signal power and better SNR than the contralateral ear [157]. Furthermore, the lateral HRTFs do not vary significantly neither in the frequency nor by azimuthal movements of the sound source. This is the domain between –20°-90° showing some symmetry to the φ=45°-axe, which direction can be declared as *absolute maximum* of the *monaural sensitivity* as the source is moving toward the ear. The localization cues are the interaural time and intensity differences. Our measurement shows that the monaural sensitivity region can be recognized significantly only between an elevation of -10° to +30°. It is quite logical to term high gain regions as sensitivity regions, like by an antenna, but even if monaural listening benefits from high gain regions there is no reason to believe that these allow e.g. good localization binaurally.
Local increase of the HRTFs is at φ=180°. A *local* monaural sensitivity domain can be identified ±20° away from the direction "back". Superior localization acuity for rear locations compared with lateral locations was reported in [160]. This is not general, but it could be due to the local increase of the monaural sensitivity near to the median plane.
The monaural sensitivity is very unsymmetrical to the median plane: it goes to φ=90° but begins only at φ=–20°. The overlapping domain of the two ears is only ±20° left and right from the median plane (Fig.50.) called *binaural sensitivity domain*. This assumes that the interaural and complex auditory sensitivity is not based only on the monaural sensitivity of the HRTFs. Humans try to face the sound sources for the best localization and use the interaural differences and the binaural fusion. In the median plane no interaural differences appear and only the HTRF should deliver all localization cues. In real-life situations head movements are very useful and important to find the source. If they are not present, front-back confusion and poor localization performance appear.
This kind of symmetry can be observed at the minimum of the sensitivity. The local and absolute minimum is at ca. 250°-260° in the head-shadow area. Local minimum at –90° in the ILD was also found in and modeled by rigid sphere [52].

6.1.2 Frequency limits in the lateral-contralateral evaluation

There are different limited areas in the frequency domain partitioned by "cut-off" frequencies during the elevation of the sound source information.
The limit at 1500-1600 Hz is well known from the literature [5, 12, 51, 99, 101]. The HRTF has five major resonant points: 3, 5, 9, 11 and 13 kHz but there are large individual differences. The high frequency components are responsible for the localization: the sensation is more correlated with the real source direction if the signal has components above 5 kHz. Above 1600 Hz the lateralisation is made based on the envelope. The constant rise of the edges in the HRTFs may suggest that the possibilities of the envelope evaluation are limited and this phenomenon has an optimum. Lateralisation below 1600 Hz is based on ITDs [161]. Interaural Intensity Differences are present from 20 Hz-20 kHz but they become important above 500 Hz. Monaural spectral features of the pinnae appear above 3-3,5 kHz, primary for elevation cues [24]. Low frequency elevation cues are not due to the pinnae but to the torso below 3 kHz [88]. We can support this observation, as we did not observe any effects or deviations below 1600 Hz in the HRTFDs.
As it was previously shown, head shadowing causes random incidence. This means, the HRTFs of the contralateral ear vary too rapidly and randomly to evaluate and decode high frequency information and the SNR is less, than on the lateral side. The test with the baseball cap supports the finding that shadowing and diffraction effects are responsible for the large high frequency deviations in the HRTFs. The frequency, from where these effects will be effective, depends on the azimuth (marked as black areas on Fig.57-67), on the elevation and on the environment as well. The variations of this "cut-off frequency" are shown on Fig.71. as functions of azimuth. This averaged result is calculated from –10° up to +60° elevation for all objects for the right ear. The lowest value of 3 kHz is in the area of the minimum monaural sensitivity supporting the findings in [24, 157, 162].

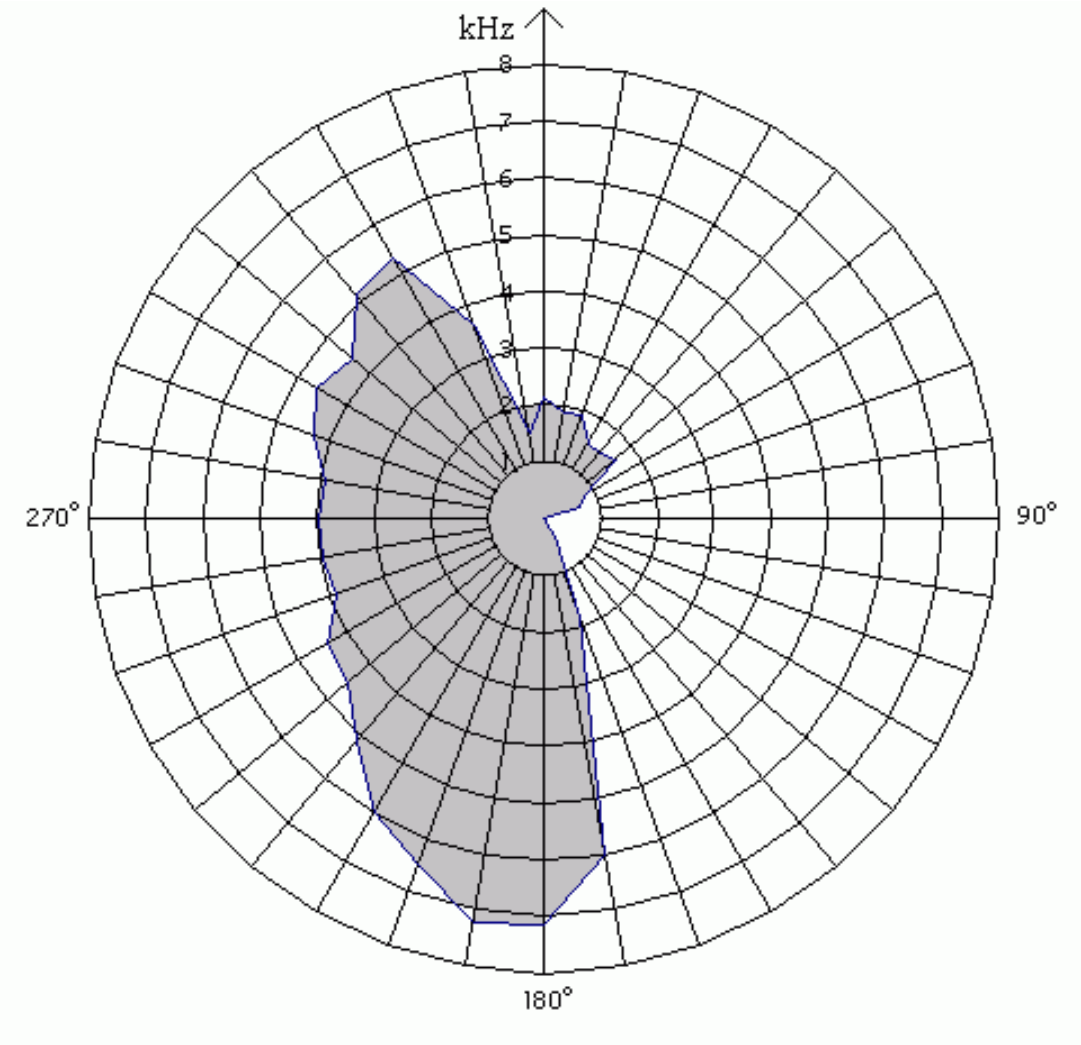


Fig.71. Frequency limit of the head-shadow area as the function of azimuth. Results are averaged over all HRTFDs. The lowest "cut-off frequency" of 3-3500 Hz is at the minimum of the monaural sensitivity (250-290°).

Although this frequency limit depends on azimuth we can define a stationary value around 3500 Hz. Near to this frequency component changes and differences in the HRTFDs appear both by the closer and by the contralateral ear. In the shadowed area only some low frequency components will be affected at 1600, 1800, 2200 and 2500 Hz. These so called "bright spots" were found by *Shaw* e.g. at 1,9 and 2,4 kHz [52, 162]. On the other hand, the closer ear will be affected at high frequencies: 9, 11, 4-5 kHz, and only seldom under 3 kHz. Special is the 8 kHz component where the most significant differences appear. At 3500 Hz (speech!) there is evaluation on both the lateral and contralateral side. This is the domain, where neither phase nor intensity differences provide an effective cue (at intermediate frequencies) [51].

6.2 Binaural evaluation

How could it be, that the hearing system is able to localize in the real life if someone is wearing glasses or a cap (these can be handled as non-individual HRTFs!) but in virtual audio environments it needs individual, "good quality" HRTFs, and allows only smaller deviations and changes of the HRTFs? Generally, headphone playback has decreased localisation performance. We suggest that equalized headphones and proper reproduction of HRTFs only allow a correct electro-acoustical transmission without distortions but the transmission of the "directional information" may fail.
The HRTFs are strongly influenced by the objects near to the listener's head. With accurate measurements we proved that even small changes in the environment cause large deviations in the entire frequency and spatial domain. Thus, the HRTFs can be declared as helpful and basic cue but not as a satisfying element of the localization-decoding procedure. As in every other "information decoding system" they represent a pre-filtering algorithm for higher processing levels but as stand-alone filters they cannot explain the whole decoding method.
Objects near to the head have different effects on frequency regions on the lateral side and on the contralateral side. High frequency components were affected by the closer ear and low frequency "bright spots" at the shadowed ear. Glasses have the smallest effects, because they are thin and cause rather high frequency responses. On the other hand, hair always has influence and caps only in the region where shadowing effects occur (due to the visor). We assume that the most undesired effect for the hearing system is the extending of the shadowed area both in frequency and azimuth, because this can lead to localization errors by losing high frequency information.
Our hearing system seems to be having the ability to "overcome" and disregard some effects appearing in the magnitude responses of the HRTFs without decreasing the localization performance. This feature is deactivated in case of using non-individualized HRTFs and/or headphone playback. The headphone-environment seems to be too "unnatural". This suggests that this "overcome function" of the higher processing system is only active when basic localization requirements are fulfilled. Furthermore, the localization is based not primarily on the magnitude of the HRTFs but on the phase information and higher processing. For the hearing system it is more important to have non-virtual environment for a good localization than differences and variations of the HRTFs.
This supports the efforts to increase the artificial recording and binaural playback systems but this is not the same as trying to get better and more accurate HRTFs [70-72]. It was shown that a binaural signal filtered with

complex HRTFs delivers different results depending on the fact as to which part would be modified. If only the magnitude response is modified, the localization can be fulfilled, but modifying the phase information disturbs the localization.
The HRTFDs confirm the important role of the interaural differences. If the source is in the monaural sensitivity region of one ear, the differences and changes due to the environment appear in the high frequency regions. At the same time, the HRTF of the contralateral ear will be influenced at lower frequencies and this results in an increased ILD. We do not find that frequency components vary in the way to decrease the ILD. Diffraction of low frequency components results in amplification on the contralateral side. Head and pinnae reflections are responsible for detection and evaluation on the lateral side. The closer ear in the high frequency regions in the monaural sensitivity domain will evaluate the information encoded in the sound waves. The contralateral ear makes evaluation of some low frequency elements where no high frequency information is available (low-pass filtering). Shadowing-effects affect the localisation: it causes random incidence, secondary sound paths, diffuse-like sound field and no primary wave front. Head shadow is the natural reason for that, but caps can also shadow.

6.2.1 Localisation performance in binaural playback systems

All binaural playback systems need HRTF reproduction as the first step. Their quality determines the localization performance [60, 120]. The goal is to get more precise, individual and "better" HRTFs to achieve the best localization. Little deviations of these HRTFs during headphone playback could result in decreased localization.
Let us consider the different binaural playback situations in order to declare "quality levels" in reference to the localization performance.

1. The best localization can be achieved with our natural hearing without headphones. We use (little) head-movements, individual complex HRTFs, room reverberations and reflections from the ground [164]. Our localization is influenced by visual information.
2. The basic situation for listening tests and HRTF measurements is the free-field environment. In real life, this can be on a high mountaintop or on the sea. Artificial solution is the anechoic room. Reflections in general decrease the localization performance, but not always [64]. They deliver information about the size and the form of the

listening room, covering materials, and the overall sound quality (reflections are essential for a source to "sound good"). In artificial free-field environment we have a finite number of sound sources. We can set only a limited number of loudspeakers and source positions [9, 10, 13, 71]. The localization is determined by the spatial distribution of the sources, by the properties of the signal (loudness, bandwidth, length, SNR, distortions etc.) and by the loudspeakers. The hearing system was able to overcome the changes of the HRTFs so far.

3. The next experimental setup is basically different. The use of headphones and HRTF reproduction decrease the localization performance [5, 46, 60] by losing the natural environment. The reproduced HRTFs have decreased spatial resolution, they are measured at several hundred different locations. This listening situation seems is be too difficult for the hearing system to overcome the changes in the HRTFs. We cannot simulate and reproduce the perfect individual HRTFs and we can hardly realize a listening room without headphones but using the HRTFs of other subjects (changing or removing of the pinnae, e.g. with a swimming cap [24]). We can assume that this effect appears if we put on a cap or glasses by the head and torso simulator. This "changing of the HRTFs" we make every day without decreasing the localization. The best localization performance in a virtual simulation can be achieved with complex, individual HRTF filtering through good quality equalized headphones. Fixing the head is not required but recommended.
4. Avoiding the reproduction of the phase information results in decreased localization.
5. Non-individual HRTFs can be upgraded with simple tools, like measuring the head and pinnae size or scaling in the frequency [4, 46, 116, 118]. The other solution is to get the HRTFs from a "good localizer".
6. Decreased spatial resolution of the HRTFs results in decreased localization performance. This leads to an increased number of interpolated, calculated HRTFs [165].
7. Further decrease is due to the use of HRTFs from a "randomly selected human", from an "averaged human

head" or from a dummy-head regardless of the quality [4, 70, 72].

8. Any calculated or averaged set of HRTFs leads to insufficient localization. HRTFs from rigid spherical head models or generalized, theoretical HRTF sets do not contain as many secondary peaks as measured data [12, 51, 157, 128]. This clearly indicates the role of fine structure of the HRTFs [95]. Simplified modeling of the head and torso (ellipsoid [24]) may help us understand some localization cues by complementing the results obtained with measured data.
9. The use of low quality headphones and/or the lack of equalization are unacceptable for scientific experiments. But we have to take into account that commercial users neither have diffuse-field equalized (FEC) good quality headphones nor the possibility to measure its transfer function and make a perfect equalization. Hardware rendering for the calculations and equalization is recommended.

The main break in the localization performance is after point 2. By using headphones with partly reproduced HRTFs, the hearing system is no longer able to overcome disturbances in the HRTFs. The substitution of the "air conduction" to headphones is responsible for the well-known errors [34]. In a hearing model the higher processing algorithms are not only active after the inner ear, but they have effects over the HRTFs as well.

By headphone playback we can calculate with losing the air conduction and with the fact that the pinnae do contribute a little bit by the transmission and measurements [77]. A comparison between headphone and loudspeaker playback can be found in [166]. The 25% error by 1 m-distance loudspeaker is increased up to 34% by a distance of 3 m, in contrast to the 24% error of the headphone playback.

Gardner found that loudspeaker seems to be a superior choice to headphones. Headphones have great difficulty reproducing frontal images using non-individualized HRTFs and externalization is clearly better using loudspeakers. Spectral maxima in HRTFs determine the location in the median plane but they are also potentially useful for horizontal localization. Out-of-head localization can be achieved using individual HRTFs, reverberation cues, visual cues and head movements (dynamic localization cues). Non-individual HRTFs can be used to generate externally images only when other cues are present [179].

Reverberation, even if it's only early reflections or attenuated, delayed versions of the direct sound (the non-minimum-phase method), is maybe sufficient to

produce external images [180]. Stimuli including reverberation yield lower azimuth errors and higher externalization, but decrease the elevational accuracy. Early reflections out to 80 ms are sufficient to provide externalization, but full auralization out to 1.5 s is not necessary. It was surprising that even the azimuth judgements were improved by approximately 5 degrees. Except for the interaction of head tracking and HRTFs for azimuth error, there is no clear advantage to using individualized HRTFs for improving localization accuracy, externalisation or reversal rates in a VAD of *speech*. Results showed no correlation between azimuth error and head size difference. These data may differ when noise stimuli or clicks are used. These effects in the magnitude response let us consider the phase information of the HRTFs to be important. *Time analysis* of the HRTFs has to be performed in the future to see the effect of small head-movements and rapid variations of the HRTFs. HRTFs are dynamic systems, their variations and differences *over time* deliver much more information than the simple magnitude response or the change of the magnitude response caused by changing the source location. Higher processing levels have more influence on the acoustic signal processing during decoding the directional information and are able to distinguish playback situations.

7 Results

This brief section summarizes the results and statements we achieved.
According to the binaural technique, the transmission of all the directional information is based on the signal pressure level at the eardrums. If we reproduced the same SPL artificially that we would have in a real-life listening situation, we would get a correct directional synthesis. This statement is neither necessary nor satisfactory.
If it would be satisfactory, we would not have all the headphone playback errors and decreased localization quality in a binaural synthesis even with the individual HRTFs (although it was reported by many authors to be not necessary the implication of individual HRTFs: HRTFs from a good localizer or a selected human could deliver similar results). Simulation of head-movements, reverberation or full auralization in a virtual reality, could improve the quality but it could also decrease the localisation blur [180]. It is still not clear what the reason is why the acoustical information transmission may fail. Even with the proper electroacoustic transmission the full simulation of a sound field seems to be insufficient, and using headphones the auditory system gets confused. The brain seems to be able to recognize and discriminate between the real-life environments and simulated sound fields.
The satisfactory application of non-individual HRTFs also supports that virtual simulation is not to be improved through more accurate simulation of HRTFs. Furthermore, if we are able to the same localization performance in a virtual simulation with individual and non-individual HRTFs, the importance of the accurate representation of the SPL at the eardrums could not be very significant. If the statement would be necessary, every little modification in the acoustical environment near to the head would lead to disturbances in perception. If we are listening exactly to the same sound source in a sound filed with or without wearing glasses – we do not recognize changes in the sensation at all. Nevertheless, the HRTFs vary rapidly. Thus, the SPL at the eardrums (or at any microphones due to Békésys' early observation) vary significantly without changing the localization ability or overall sound quality.
All these lead us to support the hypothesis of having a parallel-distributed auditory model, where higher processing is active even at the level of the outer ears and they do contribute in the evaluation of directional information. The brain is able sometimes to "neglect" the variation of the SPL and HRTFs by keeping the auditory and directional image and to distinguish between real life environments and virtual simulation independent of the SPL at the eardrums.

To prove this, we first needed to demonstrate the capability of an existing binaural playback system. The Beachtron system was already presented and used in the GUIB project. The ability for virtual simulation of sound sources for an application for blind users and the localization blur depending on the stimulus frequency as well as recommendation for further application can be found in Section 3.

- Nevertheless, all of the usual headphone playback errors occurred
- The average localization blur of about 10 degrees horizontal and 20 degrees vertical supports former measurement of other authors, but it is still inferior to real life localization.
- The system is capable for a GUIB application and future work with this system is suggested
- The test signals were chosen to assist the GUIB application with Earcons: full spectrum noise, high frequency and low frequency stimuli were used to determine the localization blur depending on stimulus frequency. Broadband and high frequency sounds are to localise the best. Furthermore, the short impulses are also similar to the tones that could be used for Earcons in the future.
- We found special localization behaviour and effects cited in the literature, like asymmetries on the left-right and up-down directions, missing locations, the variations and size of the "uncertainty domain" and the different localization blur in case of "incoming" and "outgoing" sound sources
- Three novel methods were used which are not common and seldom mentioned in the literature before. These parameters have been never evaluated together up to now.
    - The 3-categorie-forced choice is seldom used to determine the "uncertainty" of the subjects
    - The common application of the moving source in both direction allowed us to determine the localization blur together: the localization results presented are independent from the direction of the moving source
    - The application of a 2D simulation is a novel method, because commonly a constant source distance (around the head) is applied in the listening tests. The 2D surface is better acceptable for a GUIB application

Our main results are from the dummy-head HRTF measurement. For analysing accurately the HRTFs and the differences of the HRTFs caused by the environment near to the head a measurement system is needed with the following requirements:

- it must be a dummy-head system with the possibility of long-time measuring
- increased accuracy and reproducibility by the settings of azimuth and elevation
- increased spatial resolution both vertical and horizontal
- increased SNR and the possibility of reproducible measurements with the system.

We upgraded our former system. We got

- A full automatic, computer controlled system with the possibility to measure thousands of HRTFs in long-time measurements
- An accuracy of 0,77% and 1,14% by setting of elevational and horizontal positions respectively, which is better than commonly achieved accuracy
- A spatial resolution of 1 degree horizontal and 5 degrees vertical according to the best possible resolution of the auditory system
- An optimal 89 dB SNR or more based on the signal processing and the robust averaging system, which is much better than commonly achieved SNRs of about 50-70 dB.
- Furthermore, the numerical algorithm presented is capable of creating a special noisy-like excitation signal with a SNR independent of the frequency. The applied non-MLS pseudo-random noise stimuli can be used for every transfer function measurement with this system. The algorithm is able to generate easily and quickly the proper stimuli for every system other than ours.
- The deviation between measured transfer function is less than 0,5 dB using unidirectional microphones.

The use of this system allowed us to evaluate a huge database of recorded HRTFs and HRTFDs.

- the definition of HRTFD mathematically and the 2D representation of unsigned deviation in a polar histogram are well suited for the evaluation of small differences of about 1 dB
- These HRTFDs
  - can be easily calculated (complex division)
  - contain no individualism (it will be eliminated by the dividing)
  - and they can be measured with the system accurately in a huge amount.
- This kind of accuracy showed hidden effects, like the high frequency pinnae effect at 70-90 degrees of azimuth.

- Effects of clothing and everyday life objects were found to be significant and typical in some frequency and spatial domains
- Conceptions and new definition of the evaluation of directional information based on the HRTFs were given like
  - Monaural and binaural sensitivity domains
  - Frequency domains and “cut-off” frequencies of the lateral and contralateral evaluation
  - The significant effect of head-shadow and other shadowing effects in the evaluation of high frequency information delivered by the HRTFs and this was considered to be the most disturbing effect during the evaluation of directional information.

With the HRTFDs it has been proved that small differences in the acoustical environment near to the head influences the HRTFs and thus, the SPL at the eardrums without affecting the localization performance and the transmission of the acoustical information in real life environments.

On the other hand, virtual simulations even with an accurate presentation of the SPL at the eardrums often are insufficient and inferior to real life situations. Improvement of the binaural playback systems are suggested more by using alternative headphone designs and by simulating head movements and reverberation than by applying more accurate or individual HRTFs.

The extension of the binaural statements and physiological models by introducing the activity of the higher processing at the level of localization and evaluation of outer ear information is suggested.

8 Conclusions

This work focused on the role and effect of the HRTFs in the decoding of the acoustical information. This includes first of all the directional information (the location of the sound sources).
For the analysis we presented first the localization blur, discrimination capabilities, spatial resolution and the well-known headphone playback errors in a binaural system. The Beachtron system is not a state-of-the-art solution for binaural playback and simulation of virtual environments. It is however suited for listening tests and for low-cost solutions for everyday users: it offers real-time filtering of HRTFs, user-friendly applications and programming, headphone equalization and even individual settings of the HRTFs through the measurement of the head diameter. We found this system suitable for GUIB applications. On the other hand, in-the-head localization and front-back confusions are present. It seems, that headphone equalization and HRTF processing can be made in order to get a proper electroacoustic transmission to the eardrums, but it is not always suited for transmitting all the "directional information". *Plenge* reported that in-the-head and out-of-head localization does not depend on any kind of electroacoustic transmission [167]. The listening test delivered expected results supporting former results from the literature.
The statements and observations of the binaural technique lead us to investigate the role and first of all, the *fine structure* of the HRTFs in the localization. It is obvious that for this investigation we need increased measurement accuracy and accurate settings of source positions. For long-term measurements only dummy-heads are suited using broadband noise stimulus.
Dummy-heads and their HRTFs were often tested in listening tests and declared as insufficient solution for binaural playback due to large localization errors (see *Møller et al.*). It is assumed that the reason for this is the "standardized" shape of the head and torso, and they represent the average human free from individual properties. It is also suggested that improving of the quality of HATs can be made in the way to make them more detailed, thus, with more accurate HRTFs. Our measurement showed that HRTFs are not the critical point of the localization under free-field conditions and even in a virtual environment the headphone errors (the playback medium itself) are more significant than the HRTFs.
The everyday life objects near to the head affect the environment and have clear and large influence on the HRTFs (over 10 dB), although we do not recognize any differences and decrease of the localization performance in real-life

situations. The hearing system is able to extract directional and decode the acoustical information from the sound waves even if the HRTFs vary randomly and rapidly. HRTFs seem to be important only in reducing the ambiguity as a basic pre-filtering effect, first of all in the horizontal plane where interaural differences are the basic cues for the localization. In the median plane HRTFs are the basic cues, but small head movements are significant to avoid in-the-head localization and front-back confusions. However, the basic considerations of the binaural technique should be revised and extended by the evaluation of the parallel processing of the auditory cortex.

8.1 Future works and application notes

On the basis and results of this work there are several ways to continue.
The measured **dummy-head HRTF data** are suited for listening tests. All these data are burned on CD ROMs and are available for a binaural playback system.

- It would be an interesting point to find out how they work in the virtual audio environment in comparison with the Beachtron system. This seems to be a difficult task, because the two systems do not have a common structure of data storage. The Beachtron includes only 72 HRTFs in a 75-point FIR-filter format in the time-domain in an unknown file format. Our database contains ca. 4000 pieces of 4096-points HRTFs in the frequency domain, using different sampling frequency.
- Basic listening tests can be fulfilled using the HRTFs of the "bare torso" as generic, non-individual HRTFs. The HRTFs' spatial resolution of 1° allows the most accurate synthesis of the virtual audio sources corresponding to the best conceivable spatial resolution of the human hearing system. So we can avoid interpolated HRTFs at the speed of our signal processing hardware.
- A comparison can be made between a VAD using interpolated HRTFs and a simulation with the full amount of HRTF data.
- The most significant question is, whether subjects have better localization performance using the "dressed" torsos' HRTFs or not. It is expected that the use of dummy-head HRTFs with glasses or hair would not result in increased localization performance, because they would evaluated as another particular generic HRTF set.
- In addition, the set of the bare HRTFs could have been adjusted parametrically in the frequency domain to model the measured effects.
- Individual HRTF measurements on real humans are recommended using the same system as used for the dummy-head measurements. SNR, accuracy of the measured HRTFs can be compared.
- Using the individual HRTFs, listening test could provide the efficiency of accurate but non-individual generic HRTFs.
- It was suggested that alternative headphone designs may lead to better playback quality by avoiding or reducing in-the-head localization or elevation shift. Headphones can be purchased, e.g.

from Ultrasone GmbH. They produce headphones with displaced membrane in order to cancel elevation shift.

- Simulation of the importance of the small head movements can be made with the dummy-head HRTFs by adding a random "head movement" effect. E.g. the direction "front" corresponds to the HRTFs from –1, 0 and +1 degrees with a uniform distribution. This could model small head movements.
- The simulation of the lost air conduction in a headphone playback could support or revise the suggestion that this is also a reason for decreased localization and errors.
- Measurements handling with all other parameters of localization would be interesting. Simulation of reflections, room reverberation, head movements etc. is cited in the literature significant.
- Full auralization and creating of a virtual environment with head-tracking device is the final step for the equipment both in 2D and 3D.
- The system is suitable for transfer function measurements with a accuracy of about 1 dB. Noise analyzers, microphones, headphones, loudspeakers etc. can be measured using international standards (IEEE, IEC, DIN).

The results from section 3 offered **new material for the GUIB Project** as well. We have the average and worst-case resolution achieved by the Beachtron system as the function of stimulus frequency. Suggested listening tests are:

- Evaluation of the "averaged" spatial resolution. Using the same excitation signals, source locations have to be presented based on Figure 29, and subjects have to determine and to localize the sources. We can determine how much of the subjects are able to discriminate sources in this resolution indeed. It is expected that subjects will not have a 0% error ratio using an "average" distribution of sound sources on the VAD (e.g. every 10 degrees).
- The same investigation has to be made using the "worst case" resolution obtained from the listening tests. An optimal limit has to be determined where 90% of the users are able to use the VAD.
- It is essential to test these averaged, worst case and optimal resolutions using the Earcons. The Earcons were created based on the blind persons' comments, and they are available in wave-files.
- Results about pure tones, sinusoidal test signals are still missing with this system. It is not known whether high frequency noise bursts or pure high frequency tones are better to localize. This

question is related to the bandwidth of the test signal and to the efficiency of localization to tonal or noisy sound events.

- As a special test, vertical localization can be tested using timbre and/or pitch modulation of stimulus tones to create acoustic images “above” and “below”. It is expected that one third of the subjects will not be able to localize in the median plane.
- Special test can be made for investigating left-right and/or up-down asymmetries in connection with right –and left-handed persons.
- Further listening test in anechoic rooms with existing multi-channel loudspeaker setups (e.g. 5.1 home theater systems) or with common used PC-speaker sets could deliver information about low-cost solutions for blind persons or multimedia applications without using headphones.

Dummy-head production could benefit of the results presented. HATs and their HRTFs are still inferior to HRTFs measured on real humans. It is suggested that the outer geometry, shape and material are responsible for this. Because even large HRTF deviations are acceptable in free-field listening, improvement of dummy-head recording systems does not fail on the accuracy of their HRTFs. Moreover, binaural playback systems do not have to be improved by adding more accurate or individual HRTFs, but by simulating other parameters and/or by using alternative headphone designs.

HRTF filtering and the use of virtual audio environments are seldom investigated in connection with speech recognition, intelligibility and synthesis. Listening test could be made in order to determine measurable parameters of speech evaluation.

Further investigation of the role of the playback method and binaural playback media will extend and complete the (binaural) hearing models and they could be the basis for improving headphones and virtual audio solutions for the visually impaired. Higher processing algorithms and evaluation mechanisms are activated at the level of outer ears. The extension of the actual hearing models by the objective measurable parameters, which are used for localization in case of rapid variable HRTFs is highly suggested. To determine the physical parameters of the sound waves responsible for localization cues in an environment, where HRTFs are disturbed or missing, listening tests have to be performed without any kind of HRTF reproduction and focusing only on other parameters, which are still there to carry directional dependent information.

9 References

See Appendix A for alphabetical order and number of appereance.

[1] Gy. Békésy: Introduction; Handbook of Sensory Physiology. Volume V/1. Springer Verlag Berlin, Heidelberg, New York, 1974.
[2] Gy. Békésy: Experiments in hearing. New York, McGraw-Hill book Co., 1960.
[3] Gy. Békésy: Sensory Inhibition. Princetion, NJ, Princton University Press, 1967.
[4] P. Minnaar, S. K. Olesen, F. Christensen, H. Møller: Localization with Binaural Recordings from Artificial and Human Heads. *J. Audio Eng. Soc.* **49(5)**, pp. 323-336, 2001.
[5] J. Blauert: Spatial Hearing. The MIT Press, MA, 1983.
[6] E. A. G. Shaw: Transformation of sound pressure level from the free-field to the eardrum in the horizontal plane. *J. Acoust. Soc. Am.* **56(6)**, pp. 1848-1861, 1974.
[7] S. Mehrgart, V. Mellert: Transformation characteristics of the external human ear. *J. Acoust. Soc. Am.* **61(6)**, pp. 1567-1576, 1977.
[8] D. Hammershøi, H. Møller: Free-field sound transmission to the external ear; a model and some measurement. *DAGA'91*, Bochum, pp. 473-476, 1991.
[9] C. B. Jensen, M. F. Sorensen, D. Hammershøi, H. Møller: Head-Related Transfer Functions: Measurements on 40 human subjects. *Proc. of 6th Int. FASE Conference*, Zürich, pp. 225-228, 1992.
[10] H. Møller, M. F. Sorensen, D. Hammershøi, C. B. Jensen: Head-Related Transfer Functions of human subjects. *J. Audio Eng. Soc.* **43(5)**, pp. 300-321, 1995.
[11] D. Hammershøi, H. Møller: Sound transmission to and within the human ear canal. *J. Acoust. Soc. Am.* **100(1)**, pp. 408-427, 1996.
[12] W. M. Hartmann: How we localize sound. *Physics Today*, pp. 24-29, 1999 November.
[13] H. Møller, M. F. Sorensen, C. B. Jensen, D. Hammershøi: Binaural Technique: Do We Need Individual Recordings? *J. Audio Eng. Soc.* **44(6)**, pp. 451-469, 1996.
[14] J. C. Middlebrooks: Narrow-band sound localization related to external ear acoustics. *J. Acoust. Soc. Am.* **92**, pp. 2607-2624, 1992.
[15] H. Fisher, S. J. Freedman: The role of the pinna in auditory localization. *J. Audiol. Research* **8**, pp. 15-26, 1968.

[16] J. Blauert: Localization and the law of the first wavefront in the median plane. *J. Acoust. Soc. Am.* **50**, pp. 466-470, 1971.
[17] J. Blauert: Untersuchungen zum Richtungshören in der Medianebene bei fixiertem Kopf. Dissertation, Techn. Hochschule Aachen, 1969.
[18] M. Morimoto, H. Aokata: Localization cues of sound sources in the upper hemisphere. *J.A.S. of Japan* **E 5**, pp. 165-173, 1984.
[19] A. J. Watkins: Psychoacoustical aspects of synthetized vertical locale cues. *J. Acoust. Soc. Am.* **63**, pp. 1152-1165, 1978.
[20] R. A. Butler, K. Belendiuk: Spectral cues utilized in the localization of sound in the median sagittal plane. *J. Acoust. Soc. Am.* **61**, pp. 1264-1269, 1977.
[21] S. K. Roffler, R. A. Butler: Factors that influence the localization of sound in the vertical plane. *J. Acoust. Soc. Am.* **43**, pp. 1255-1259, 1968.
[22] J. Blauert: Sound Localization in the median plane. *Acoustica* **22**, pp. 205-213, 1969/1970.
[23] E. A. G. Shaw: External ear response and sound localization. In Gatehouse: Sound Theory and Applications, Amphora, Groton, CT., pp. 30-41, 1982.
[24] V. R. Algazi, C. Avendano, R. O. Duda: Elevation localization and head-related transfer function analysis at low frequencies. *J. Acoust. Soc. Am.* **109**, pp. 1100-1122, 2001.
[25] M. Cohen, E. Wenzel: The design of Multidimensional Sound Intrefaces. in W. Barfield, T.A. Furness III (Editors) „Virtual Environments and Advanced Interface Design", pp. 291-346, Oxford Univ. Press, New York, Oxford, 1995.
[26] R. H. Domnitz, H. S. Colburn: Lateral position and interaural discrimination. *J. Acoust. Soc. Am.* **61**, pp. 1586-1598, 1977.
[27] L. F. Elfner, D. R. Perrott: Lateralization and intensity discrimination. J. Acoust. Soc. Am. 42, pp. 441-445, 1967.
[28] L. A. Jeffress, D. McFadden: Detection, lateralization and the phase angle alpha. *J. Acoust. Soc. Am.* **47**, pp. 130, 1970.
[29] A. W. Mills: Lateralization of high frequency tones. *J. Acoust. Soc. Am.* **32**, pp. 132-134, 1960.
[30] G. Moushegian, L. A. Jeffres: Role of interaural time and intensity differences in the lateralization of low-frequency tones. *J. Acoust. Soc. Am.* **31**, pp. 1441-1445, 1959.
[31] J. Blauert: Ein Beitrag zur Trägheit des Richtungshörens in der Horizontalebene. *Acoustica* **20**, pp. 200-206, 1968.
[32] S. V. Galginaitis: Dependence of localization on azimuth. *J. Acoust. Soc. Am.* **28**, pp. 153-154, 1956.
[33] L. A. Jeffress, D. McFadden: Differences of interaural phase and level in detection and lateralization. *J. Acoust. Soc. Am.* **49**, pp. 1169-1179, 1971.

[34] K. Genuit, H. J. Platte: Überlegungen zur Substitution des natürlichen Außenohres durch elektroakustische Mittel. *DAGA'80*, München, pp. 779-782, 1980.
[35] K. Hartung: Modellalgorithmen zum Richtungshören, basierend auf den Ergebnissen psychoakustischer und neurophysiologoscher Experimente mit virtuellen Scallquellen. Dissertation, Ruhr-Universität, Bochum, 1997. (Shaker Verlag, Aachen, 1999)
[36] P. Laws: Entfernungshören und das Problem der Im-Kopf-Lokalisiertheit von Hörerignissen. *Acoustica* **29**, pp. 243-259, 1973.
[37] F. E. Toole: In-head localization of acoustic images. *J. Acoust. Soc. Am.* **48**, pp. 943-949, 1969.
[38] P. Laws: Zum Problem des Entfernungshören und der Im-Kopf-Lokalisertheit von Hörerignissen. Dissertation, Techn. Hochschule Aachen. 1972.
[39] G. Plenge: Über das Problem der Im-Kopf-Lokalisation. *Acoustica* **26**, pp. 241-252, 1972.
[40] N. Sakamoto, T. Gotoh, Y. Kimura: On „out-of-head localization" in headphone listening. *J. Audio Eng. Soc.* **24**, pp. 710-716, 1976.
[41] W. Noble: Auditory localization in the vertical plane: Accuracy and constraint on bodily movement. *J. Acoust. Soc. Am.* **82**, pp. 1631-1636, 1987.
[42] L. R. Bernstein, C. Trahiotis, M. A. Akeroyd, K. Hartung: Sensitivity to brief changes of interaural time and interaural intensity. *J. Acoust. Soc. Am.* **109(4)**, pp. 1604-1616, 2001.
[43] D. McFadden, E. G. Pasanen: Lateralization at high frequencies based on interaural time differences. *J. Acoust. Soc. Am.* **59**, pp. 634-639, 1976.
[44] K. Genuit: Eine systemtheoretische Beschreibung des Aussenohres. *DAGA'85*, Stuttgart, pp. 459-462, 1985.
[45] P. Berényi, A. Illényi: What does it mean for an HRTF not to have the minimal phase property? *Proceedings of Inter-Noise 96*, Liverpool, pp. 2127-2130, 1996.
[46] E. M. Wenzel, M. Arruda, D. J. Kistler, F. L. Wightman: Localization using nonindividualized head-related transfer functions *J. Acoust. Soc. Am.* **94(1)**, pp. 111-123, 1993.
[47] S. H. Foster, E. M. Wenzel: Virtual Acoustic Environments: The Convolvotron. Demo system presentation at SIGGRAPH'91, *18th ACM Conference on Computer Graphics and Interactive Techniques*, Las Vegas, NV (ACM Press, New York), 1991.
[48] D. J. Kistler, F. L. Wightman: A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction. *J. Acoust. Soc. Am.* **91**, pp. 1637-1647, 1991.

[49] J. Sandvad, D. Hammershøi: Binaural auralization. Comparison of FIR and IIR filter representation of HIRs. *Proc.of 96th Convention of the Audio Eng. Soc.*, Preprint #3862, Amsterdam, 1994.
[50] B. G. Shinn-Cunningham, S. Santarelli, N. Kopco: Tori of confusion: Binaural localisation cues for sources within reach of a listener. *J. Acoust. Soc. Am.* **107(3)**, pp. 1627-1636, 2000.
[51] G. F. Kuhn: Model for the interaural time differences in the azimuthal plane. *J. Acoust. Soc. Am.* **62**, pp. 157-167, 1977.
[52] D. S. Brungart, W. M. Rabinowitz: Auditory localization of nearby sources. Head-related transfer functions. *J. Acoust. Soc. Am.* **106(3)**, pp. 1465-1479, 1999.
[53] D. W. Batteau: The role of the pinna in human localization. *Proc. Roy. Soc. London*, Series **B 168**, pp. 158-180, 1967.
[54] D. J. Haigh: Evidence for generation of multipath localisation cues by human pinnae. *Acustica* **84**, pp. 914-917, 1998.
[55] V. Mellert, K. F. Siebrasse, S. Mehrgardt: Determination of the transfer function of the external ear by an impulse response measurement. *J. Acoust. Soc. Am.* **56**, pp. 1913-1915, 1974.
[56] D. Wright, J. H. Hebrank, B. Wilson: Pinna reflections as cues for localization. *J. Acoust. Soc. Am.* **56**, pp. 957-962, 1974.
[57] M. B. Gardner, R. S. Gardner: Problem of localization in the median plane: effect of pinnae cavity occlusion. *J. Acoust. Soc. Am.* **53**, pp. 400-408, 1973.
[58] A. D. Musicant, R. A. Butler: The influence of pinnae-based spectral cues on sound localization. *J. Acoust. Soc. Am.* **75**, pp. 1195-1200, 1984.
[59] J. C. Middlebrooks, D. M. Green: Sound localization by human listeners. *Ann. Rev. Psychol.* **42**, pp. 135-159, 1991.
[60] H. Møller: Fundamentals of binaural technology. *Applied Acoustics* **36**, pp. 171-218, 1992.
[61] J. V. Hundeboll, K. A. Larsen, H. Møller, D. Hammershøi: Transfer characteristics of headphones. *Proc. of 6th Int. FASE Conference*, Zürich, pp. 161-164, 1992.
[62] H. Møller, D. Hammershøi, C. B. Jensen, M. F. Sorensen: Transfer Characteristics of Headphones Measured on Human Ears. *J. Audio Eng. Soc.* **43(4)**, pp. 203-216, 1995.
[63] J. Blauert, P. Laws: Verfahren zur orts- und klanggetrauen Simulation von Lautsprecherbeschallungen mit Hilfe von Kopfhörern. *Acustica* **29**, pp. 273-277, 1973.
[64] M. Bodden, G. Canavét, J. Grabke, K. Hartung, T. Takahashi: Räumliches Hören in komplexen akustischen Umgebungen. *DAGA'94*, pp. 1137-1140, 1994.

[65] J. Kawaura, Y. Suzuki, F. Asano, T. Sone: Sound localization in headphone reproduction by simulating transfer functions from the sound source to the external ear. *J. Acoust.Soc.Japan* **E 12**, pp. 203-215, 1991.
[66] P. Damaske, B. Wagener: Richtungshörversuche über einen nachgebildeten Kopf. *Acoustica* **21**, pp. 30-35, 1969.
[67] C. Jin, M. Schenkel, S. Carlile: Neural system identification model of human sound localization. *J. Acoust. Soc. Am.* **108(3)**, pp. 1215-1235, 2000.
[68] N. Cheung, S. Trautman, A. Horner: Head-Related Transfer Function Modeling in 3-D Sound Systems with Genetic Algorithms. *J. Audio Eng. Soc.* **46(6)**, pp. 531-539, 1998.
[69] M. D. Burkhard, R. M. Sachs: Anthropometric manikin for acoustic research. *J. Acoust. Soc. Am.*, **58(1)**, pp. 214-222, 1975.
[70] H. Møller: On the quality of artificial head recording systems. *Proceedings of Inter-Noise 97*, Budapest, pp. 1139-1142, 1997.
[71] P. Maijala: Better binaural recordings using the real human head,. *Proceedings of Inter-Noise 97*, Budapest, pp. 1135-1138, 1997.
[72] H. Møller, D. Hammershøi, C. B. Jensen, M. F. Sorensen: Evaluation of artificial heads in listening tests. *J. Acoust. Soc. Am.* **47(3)**, pp. 83-100, 1999.
[73] M. Kleiner: Problems in the design and use of „dummy heads". *Acoustica* **41**, pp. 183-193, 1978.
[74] T. Tarnóczy: Über den Verstärkerungs-Verminderungs-Effekt der Ohrmuschel und des Kopfes. *Proc. of 6th Int. FASE Conference*, Zürich, pp. 229-232, 1992.
[75] G. F. Kuhn: The pressure transformation from a diffuse sound field to the external ear and to the body and head surface. *J. Acoust. Soc. Am.* **65 (4)**, pp. 991-1000, 1979.
[76] K. Genuit, H. Sibinger: Kalibrierung einer Kunstkopf-Übertragungskette. *DAGA'94*, Dresden, pp. 685-688, 1994.
[77] P. Schöne: Zur Nutzung des Realisierungsspielraums in der kopfbezogenen Stereofonie. *Rundfunktechnische Mitteilungen*, Jahrgang 24, Heft 1, pp. 1-11, 1980.
[78] A. Schmitz: Diskussion verschiedener Verfahren zur Wiedergabe kopfbezogener Signale. *DAGA'94*, Dresden, pp. 277-280, 1994.
[79] P. A. Hill, P. A. Nelson, O. Kirkeby: Resolution of front-back confusion in virtual acoustic imaging systems. *J. Acoust. Soc. Am.* **108(6)**, pp. 2901-2910, 2000.
[80] R. Helle: Das Übertragungsmass von Kopfhörern an der Messpuppe KEMAR. *DAGA'80*, München, pp. 803-806, 1980.
[81] P. Schöne: Der Signalstörabstand bei Kunstköpfen. *DAGA'80*, Berlin, pp. 835-838, 1980.

[82] K. Genuit: Bestimmung strukturgemittelter Außenohr-übertragungsfunktionen. *DAGA'84* Darmstadt, pp. 667-670, 1984.
[83] K. Genuit: Ein kalibrierfähiges Kunstkopf Mess-System. *DAGA'84*, Darmstadt, pp. 279-282, 1984.
[84] R. L. Martin, K. I. McAnally, M. A. Senova: Free-Field Equivalent Localization of Virtual Audio, *J. Audio Eng. Soc.* **49(1/2)**, pp. 14-22, 2001.
[85] K. Genuit: Strukturbestimmende Merkmale von Außenohr-übertragungseigenschaften und deren Abhängigkeit von der Schalleinfallsrichtung. *DAGA'82*, Göttingen, pp. 1195-1198, 1982.
[86] K. Genuit: Ein Modell zur Beschreibung der Außenohreigenschaften. Dissertation, TH Aachen, 1984.
[87] Fukodome: Eqivalisation for dummy-head headphone system for reproduction directional information. *J.A.S. of Japan* **E 1 (1)**, pp. 59-67, 1980.
[88] K. Genuit, H. J. Platte: Untersuchungen zur Realisation einer richtungsgetreuen Übertragung mit elektroakustischen Mitteln. *DAGA'81*, Berlin, pp. 629-632, 1981.
[89] H. W. Gierlich, K. Genuit: Processing Artificial-Head Recordings. *J. Audio Eng. Soc.* **37(1/2)**, pp. 34-39, 1989.
[90] K. Genuit, M. Burkhard: Artificial head measurement systems for subjective Evaluation of sound Quality. *Sound and Vibration*, pp. 18, 1992 March.
[91] K. Genuit: Simulation des Freifeldes über Kopfhörer zur Untersuchung des Richtungshörens und der Selektionsfähigkeit. *Audiologische Akustik*, Jahrgang 27, Heft 6, pp. 206-221, 1988.
[92] H. W. Gierlich, K. Genuit: Entwurf eines mikroprozessorgesteuerten Aussenohrsimulators. *DAGA'84*, Darmstadt, pp. 671-674, 1984.
[93] K. Genuit: Untersuchungen zur Bedeutung von einzelnen Strukturen der Außenohrübertagungs-funktion und das räumliche Hören, *DAGA'86*, Oldenburg, pp. 485-488, 1986.
[94] J. Blauert: Psychoakustik des binauralen Hörens. *DAGA'84*, Darmstadt, invited plenary paper, pp. 117-128, 1984.
[95] S. Carlile, D. Pralong: The location-dependent nature of perceptually salient features of the human head-related transfer functions. *J. Acoust. Soc. Am.* **95**, pp. 3445-3459, 1994.
[96] F. Asano, Y. Suzuki, T. Sone: Role of spectral cues in median plane localization. *J. Acoust. Soc. Am.* **88**, pp. 159-168, 1990.
[97] M. Bodden: Binaurale Signalverarbeitung: Modellierung der Richtungserkennung und des Cocktail-Party-Effektes. VDI Fortschrittberichte, Reihe 17, Biotechnik, Nr.85, VDI Verlag, Düsseldorf, 1992.

[98] W. Gaik: Untersuchungen zur binauralen Verarbeitung Kopfbezogener Signale. VDI Fortschrittberichte, Reihe 17, Biotechnik, Nr. 63, VDI Verlag, Düsseldorf, 1990.
[99] T. T. Sandel, D. C. Teas, W. E. Feddersen, L. A. Jeffress: Localization of sound from single and paired sources. *J. Acoust. Soc. Am.* **27**, pp. 842-852, 1955.
[100] W. M. Hartmann, B. Rakerd: On the minimum audible angle – A decision theory approach. *J. Acoust. Soc. Am.* **85**, pp. 2031-2041, 1989.
[101] W. Mills: On the minimum audible angle. *J. Acoust. Soc. Am.* **30**, pp. 237-246, 1958.
[102] T. Z. Strybel, C. L. Manlingas, D. R. Perrott: Minimum Audible Movement Angle as a function of azimuth and elevation of the source. *Human Factors* **34(3)**, pp. 267-275. 1992.
[103] D. R. Perrott, A. D. Musicant: Minimum auditory movement angle: binaural localization of moving sources. *J. Acoust. Soc. Am.* **62**, pp. 1463-1466, 1977.
[104] J. Zwislocki, R. S. Feldman: Just noticeable differences in dichotic phase. *J. Acoust. Soc. Am.* **28**, pp. 860-864. 1956.
[105] P. A. Campbell: Just noticeable differences of changes of interaural time differences as a function of interaural time differences. *J. Acoust. Soc. Am.* **31**, pp. 917-922, 1959.
[106] M. Kinkel, B. Kollmeier: Diskrimination interauraler Parameter bei Schmalbandrauschen. *DAGA'87*, Aachen, pp. 537-540, 1987.
[107] J. L. Hall: Minimum detectable change in interaural time or intensity difference for brief impulsive stimuli. *J. Acoust. Soc. Am.* **36**, pp. 2411-2413, 1964.
[108] D. R. Perrott, J. Tucker: Minimum Audible Movement angle as a function of signal frequency and the velocity of the source. *J. Acoust. Soc. Am.* **83**, pp. 1522-1527, 1988.
[109] J. M. Chowning: The simulation of Moving Sound Sources. *J. Audio Eng. Soc.* **19**, pp. 2-6, 1971.
[110] D. W. Grantham: Detection and discrimination of simulated motion of auditory targets in the horizontal plane. *J. Acoust. Soc. Am.* **79**, pp. 1939-1949, 1986.
[111] Crystal River Engineering, Inc. : BEACHTRON – Technical Manual, Rev.C., 1993.
[112] K. Crispien, H. Petrie: Providing Access to GUI's Using Multimedia System – Based on Spatial Audio Representation. *J. Audio Eng. Soc. 95th Convention Preprint*, New York, 1993.
[113] M. Kleiner, B. I. Dalenbäck, P. Svensson: Auralization – an overview. *J. Audio Eng. Soc.* **41**, pp. 861-875, 1993.

[114] K. D. Jacob, M. Jorgensen, C. B. Ickler: Verifying the accuracy of audible simulation (auralization) systems. *J. Acoust. Soc. Am.* **92**, pp. 2395, 1992.
[115] C. Tan, W. Gan: Direct concha excitation for the introduction of individualized hearing cues. *J. Audio Eng. Soc.* **48(7-8)**, pp. 642-653, 2000.
[116] J. C. Middlebrooks: Individual differences in external-ear transfer functions reduced by scaling in frequency. *J. Acoust. Soc. Am.* **106(3)**, pp. 1480-1491, 1999.
[117] J. C. Makous, and J. C. Middlebrooks: Two-dimensional sound localization by human listeners. *J. Acoust. Soc. Am.* **87(5)**, pp. 2188-2200, 1990.
[118] J. C. Middlebrooks: Virtual localisation improved by scaling nonindividualized external-ear transfer function in frequency. *J. Acoust. Soc. Am.* **106(3)**, pp. 1493-1510, 1999.
[119] D. J. Kistler, F. L. Wightman: Principal Component Analysis of Head-Related Transfer Functions. *J. Acoust. Soc. Am.* **88**, pp. 98, 1990.
[120] F. L. Wightman, D. J. Kistler: Headphone Simulation of Free-Field Listening I.-II. *J. Acoust. Soc. Am.* **85**, pp. 858-878, 1989.
[121] J. Blauert, H. Lehnert, J. Sahrhage, H. Strauss: An Interactive Virtul-environment Generator for Psychoacoustic Research I: Architecture and Implementation. *Acoustica* **86**, pp. 94-102, 2000.
[122] R. L. McKinley, M. A. Ericson: Digital synthesis of binaural auditory localization azimuth cues using headphones. *J. Acoust. Soc. Am.* **83**, S18, 1988.
[123] J. F. Burger: Front-back discrimination of the hearing system. *Acustica* **8**, pp. 301-302, 1958.
[124] D. Burger, C. Mazurier, S. Cesarano, J. Sagot: The design of interactive auditory learning tools. *Non-visual Human-Computer Interactions* **228**, pp. 97-114, 1993.
[125] M. M. Blattner, D. A. Sumikawa, R. M. Greenberg: Earcons and Icons: their structure and common design principles. *Human-Computer Interaction* **4(1)**, pp. 11-44, 1989.
[126] K. Brinkmann, U. Richter: Zur Messunsicherheit bei psychoakustischen Messungen. *DAGA'87*, Aachen, pp. 593-596, 1987.
[127] G. Awad: Ein Beitrag zur Mensch-Maschine-Kommunikation für Blinde und Hochgradig Sehbehinderte. Dissertation, TU Berlin, Berlin, 1986.
[128] V. R. Algazi, C. Avendano, R. O. Duda: Estimation of a spherical-head model from anthropometry. *J. Audio Eng. Soc.* **49(6)**, pp. 472-479, 2001.
[129] E. Zwicker, R. Feldtkeller: Das Ohr als Nachrichtenempfänger. S. Hirzel Verlag, Stuttgart, pp.181, 1967.
[130] R. A. Butler, R. F. Naunton: Role of stimulus frequency and duration in the phenomenon of localization shifts. *J. Acoust. Soc. Am.* **36(5)**, pp. 917-922, 1964.

[131] Gy. Wersényi: Acoustic Signal Processing for Listening Tests in Virtual Audio. *2001 Polish-Czech-Hungarian Workshop on Circuit Theory, Signal Processing, and Telecommunication Networks*, Budapest, pp. 175-181, 2001.
[132] G. Boerger, G. Laws, J. Blauert: Stereophonic headphone reproduction with variation of various transfer factors by means of rotational head movements. *Acoustica* **39**, pp. 22-26, 1977.
[133] D. R. Begault: 3-D Sound for Virtual Reality and Multimedia. Academic Press, London, UK, 1994.
[134] A. W. Bronkhorst: Localization of real and virtual sources. *J. Acoust. Soc. Am.* **98**, pp. 2542-2552, 1995.
[135] S. R. Oldfield, S. P. A. Parker: Acuity of sound localisation: a topography of auditory space I-II. *Perception* **13**, pp. 581-617, 1984.
[136] P. Scherer: Inversionsversuch zur Vorne-Hinten-Ortung mit Sinustönen. *DAGA'84*, Darmstadt, pp. 743-746, 1984.
[137] Gy. Wersényi, A. Illényi: Evaluation of Differences in Dummy-Head HRTFs Caused by the Acoustical Environment Near to the Head. (Submitted to the acta acoustica).
[138] F. M. König: Headphone Reinforcement and Accompying Psychoacoustic Effects. *Proceedings of the International Békésy Centenary Conference on hearing and related sciences*, Budapest, pp. 166-171, 1999.
[139] F. M. König: A new supra-aural dynamic headphone system for in-front localization and surround reproduction of sound, *J. Audio Eng. Soc. Convention Preprint 4495*, München, 1997.
[140] F. König: Über die Notwendigkeit, ein- bis dreidimensional-räumliche Hörereignisse von variirenden Kopfhörerbeschallungstechniken zu beschreiben. Teil I.-II. *DAGA'96*, Bonn, pp. 384-387, 1996.
[141] F. König: Beschreibung von Klangfarbenveränderungen beim 3D-räumlich varianten Hören mittels bestimmter Kopfhörer-Beschallungsverfahren. *DAGA'97*, Kiel, pp. 598-600, 1997.
[142] S. M. Abel, C. Giguere, A. Consoli, B. C. Papsin: Front/Back Mirror Image Reversal Errors and Left/Right Asymmetry in Sound Localization. *Acoustica* **85**, pp. 378-389, 1999.
[143] K. Burke, A. Letsos, R. Butler: Asymmetric performances in binaural localization of sound in space. *Neuropsychologia* **32**, pp. 1409-1417, 1994.
[144] Gy. Wersényi: Measurement system upgrading for more precise measuring of the Head-Related Transfer Functions. *Proceedings of Inter-Noise 2000*, Nice, pp. 1173-1176, 2000.
[145] Gy. Wersényi, A. Illényi: Measurement Accuracy of a Dummy-Head Measurement System. (revised by the J. Audio Eng. Soc.)
[146] T. Behrens, H. Prante, C. Maschke: Untersuchungen zur Summenlokaliation in der Medianebene. *DAGA'94*, pp. 1157-1160, 1994.

[147] J. Borish, J. B. Angell: An Efficient Algorithm for Measuring the Impulse Response Using Pseudorandom Noise. *J. Audio Eng. Soc.* **31(7)**, pp. 478-488, 1983.
[148] K. Genuit, N. Xiang: Measurements of Artificial Head Transfer Functions for Auralization and Virtual Audio Environment. *Proceedings of ICA'95*, Trondheim, pp. 469-472, 1995.
[149] D. D. Rife, J. Vanderkooy: Transfer-function measurements with maximum-length sequences. *J. Audio Eng. Soc.* **37**, pp. 419-444, 1989.
[150] U. P. Svensson, J. H. Nielsen: Errors in MLS Measurements Caused by Time Variance in Acoustic Systems. *J. Audio Eng. Soc.* **47(11)**, pp. 907-926, 1999.
[151] B. Zhou, D. M. Green, J. C. Middlebrooks: Characterization of external ear impulse responses using Golay codes. *J. Acoust. Soc. Am.* **92**, pp. 1169-1171, 1992.
[152] S. Foster: Impulse response measurements using Golay codes. *IEEE Conf. Acoustics Speech Sig.Proc.* **2**, pp. 929-932, 1986.
[153] S. Müller, P. Massarani: Transfer-Function Measurement with sweeps. *J. Audio Eng. Soc.* **49(6)**, pp. 443-471, 2001.
[154] E. Terhardt, W. Aures: Wahrnehmbarkeit der periodischen Wiederholung von Rauschsignalen. *DAGA'84*, Darmstadt, pp. 769-772, 1984.
[155] A. V. Oppenheim, R. W. Schafer: Digital Signal Processing. Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1975.
[156] U. Tietze, Ch. Schenk: Halbleiter-Schaltungstechnik. Springer-Verlag, Berlin, 1999.
[157] C. I. Cheng, G. H. Wakefield: Introduction to Head-Related Transfer Functions (HRTFs): Representations of HRTFs in Time, Frequency, and Space. *J. Audio Eng. Soc.* **49**, pp. 231-249, 2001.
[158] Gy. Wersényi, P. Tatai: Detection of reflections in free-field directional hearing by waveform analysis of accurate dummy-head HRTFs. *Proceedings of IEEE Instrumentation and Measurement Technology Conference*, Budapest, pp. 606-609, 2001.
[159] P. D. Hatziantoniu, J. N. Mourjopoulos: Generalized Fractional-Octave Smoothing of Audio and Acoustic Responses. *J. Audio Eng. Soc.* **48(4)**, pp. 259-278, 2000.
[160] S. E. Boehnke, D. P. Phillips: Azimuthal tuning of human perceptual channels for sound location. *J. Acoust. Soc. Am.* **106(3)**, pp. 1948-1956, 1999.
[161] F. L. Wightman, D. J. Kistler: The dominant role of low-frequency interaural time differences in sound localization. *J. Acoust. Soc. Am.* **91**, pp. 1648-1661, 1992.
[162] E. A. G. Shaw: The external ear. *Handbook of Sensory Physiology* **1**, Auditory System, Anatomy Physiology Ear, Springer, New York, 1974.

[163] A. Illényi, Gy. Wersényi: Discrepancy in binaural tests and in measurements of sound field parameters. *Proceedings of the International Békésy Centenary Conference on hearing and related sciences*, Budapest, pp. 160-165, 1999.
[164] H. Wilkens: Mehrdimensionale Beschreibung subjektiver Beurteilungen der Akustik von Konzertsälen. *Acoustica* **38(1)**, pp. 10-23, 1977.
[165] E. M. Wenzel, S. H. Foster: Perceptual consequences of interpolating head-related transfer functions during spatial synthesis. *Proceedings of the ASSP Workshop on Applications of Signal Processing to Audio and Acoustics*, New York, USA, 1993.
[166] A. Ripka, G. Theile: Die Beurteilung verschiedener Stereofoner Wiedergaberichtungen bezüglich aber der Abbildungsschärfe. *DAGA'87*, Aachen, pp. 585-588, 1987.
[167] G. Plenge: On the difference between localization and lateralization. *J. Acoust. Soc. Am.* **56**, pp. 944-951, 1974.
[168] B. G. Haustein, W. Schirmer: Messeinrichtung zur Untersuchung des Richtungslokalisations-vermögens. *Hochfrequenztech. und Elektroakustik* **79**, pp. 96-101, 1970.
[169] R. S. Heffner, H. E. Heffner: Sound localization acuity in the cat: Effect of azimuth, signal duration and test procedure. *Hear. Res.* **36**, pp. 221-232, 1988.
[170] R. Y. Litovsky, D. H. Ashmed: Development of Binaural and Spatial Hearing in Infants and Children. in Binaural and Spatial Hearing in Real and Virtual Environments (edited by R.H. Gilkey and T.R. Anderson), Lawrence Erlbaum Ass., Mahwah, New Jersey, pp. 571-592, 1997.
[171] http://www.dasp.uni-wuppertal.de/audite/psychoak/psychoak26.htm
[172] S. R. Oldfield, S. P. A. Parker, „Acuity of sound localisation: a topography of auditory space III.", *Perception* **15**, pp. 67-81, 1986.
[173] J. C. Middlebrooks: Spectral Shape Cues for Sound Localization. Binaural and Spatial Hearing in Real and Virtual Environments, Lawrence Erlbaum Ass., Mahwah, New Jersey, pp. 77-97, 1997.
[174] R. L. McKinley, M. A. Ericson: Flight Demonstration of a 3-D Auditory Display. in Binaural and Spatial Hearing in Real and Virtual Environments (edited by R.H. Gilkey and T.R. Anderson), Lawrence Erlbaum Ass., Mahwah, New Jersey, pp. 683-699, 1997.
[175] R. O. Duda: Elevation Dependence of the Interaural Transfer Function. in Binaural and Spatial Hearing in Real and Virtual Environments (edited by R.H. Gilkey and T.R. Anderson), Lawrence Erlbaum Ass., Mahwah, New Jersey, pp. 49-75, 1997.
[176] W. Lindemann: Extension of binaural cross-correlation by contralateral inhibition I.-II. *J. Acoust. Soc. Am.* **80(6)**, pp. 1608-1630, 1986.

[177] R. Wettschurek: Die absoluten Unterschiedsschwellen der Richtungswahrnehmung in der Medianebene beim natürlichen Hören sowie beim Hören über ein Kunstkopf-Übertragungssystem. *Acoustica* **28**, pp. 197-208, 1973.
[178] R. Wettschurek: Über Unterschiedsschwellen beim Richtungshören in der Medianebene. *Gemeinschaftstagung für Akustik und Schwingungstechnik*, Berlin, pp. 385-388, VDI-Verlag, Düsseldorf, 1970.
[179] W. G. Gardner: 3-D Audio Using Loudspeakers. Kluwer Academic Publ., Boston, 1998.
[180] D. R. Begault, E. Wenzel, M. Anderson: Direct Comparison of the Impact of Head Tracking Reverberation, and Individualized Head-Related Transfer Functions on the Spatial Perception of a Virtual Speech Source. *J. Audio Eng. Soc.* **49(10)**, pp. 904-917, 2001.
[181] J. O. Pickles: An Introduction to the Physiology of Hearing. Academic Press, London, 1982.
[182] K. Crispien, K. Fellbaum: Use of Acoustic Information in Screen Reader Programs for Blind Computer Users: Results from the TIDE Project GUIB. *The European Context for Assistive Technology* (I. Porrero, R. Bellacasa), IOS Press Amsterdam, 1995.

10 German Abstract

HRTFs in der menschlichen Lokalisation: Messungen, spektrale Auswertung und praktische Anwendungen in virtueller akustischer Umgebung

AUSZUG DER DISSERTATION
Wersényi György

Brandenburgische Technische Universität, Cottbus
Technische Universität Budapest

I. Einleitung und Hintergründe

Die Rolle der Außenohr-Übertragungsfunktionen (Head-Related Transfer Functions: HRTFs) ist oft das Ziel der Untersuchungen der menschlichen Lokalisation. Es ist bereits bekannt, daß dieser Filtereffekt der Ohren und des Körpers der erste und einer der wichtigsten Schritte bei der räumlichen Orientierung für das Gehör ist.
In dieser Arbeit werden die HRTFs näher analysiert. Es wird gezeigt wie und unter welchen Umständen sie einen bedeutungsvollen Einfluss haben. Die Auswertung stützt sich auf drei Säulen: Messung, spektrale Auswertung und psychoakustische Bewertung.
Zuerst wird die Lokalisationsunschärfe bzw. die Diskriminationsfähigkeit mit 40 Probanden in einer zweidimensionalen virtuellen akustischen Umgebung in zwei Schritten untersucht. Im ersten Teil haben wir nach Fehlern gesucht, die

oft bei der Kopfhörerwiedergabe vorkommen. Im zweiten Teil wurde die Diskriminationsfähigkeit und die Möglichkeiten des ganzen Systems (inklusive die HRTF Simulation) festgestellt.
Um den Effekt und die Rolle der HRTFs näher zu analysieren, haben wir ein spezielles Meßsystem und ein Meßverfahren vorbereitet, mit dem man die HRTFs eines Kunstkopfes mit erhöhtem Signal-Rausch-Abstand präzise und reproduzierbar messen kann. Die Übertragungsfunktionen des Kunstkopfes wurden mit einem Rauschsignal im schalltoten Raum gemessen und in der Frequenzdomäne analysiert.
Im letzten Abschnitt wurden deren speziellen Eigenschaften mit Hilfe der sogenannten HRTFDs ermittelt (HRTF Differences). Sie sind als der Quotient von zwei HRTFs definiert. Ganz besonders ist der Effekt in der näheren Umgebung des Kopfes untersucht worden. Wie ändern sich die HRTFs durch kleine Veränderungen der Umgebung und welchen Einfluss hat das auf die Lokalisation, sowohl in Freifeldübertragung als auch bei Kopfhörerwiedergabe?

II. HRTFs in Hörtests: Lokalisationsunschärfe in einem 2D Virtual Audio System

Die HRTFs sind Teil der Aufarbeitung der akustischen Information in der menschlichen Lokalisation. Während einer Freifeldübertragung (z.B. Lautsprecherwiedergabe) ist dieser natürliche Filtereffekt von Ohrmuschel, Kopf und Oberkörper vorhanden. Bei einer Schallfeldsimulation durch Kopfhörerwiedergabe müssen diese Filterfunktionen elektronisch nachgebildet werden. Aus der Literatur kennen wir bereits zahlreiche Ergebnisse der Messungen der Lokalisationsunschärfe in einer virtuellen Umgebung. Es wurde gezeigt, daß die besten Ergebnisse, d.h. die beste räumliche Auflösung, durch die Anwendung von individuellen HRTFs erzielt werden können. Da dies in der Praxis unpraktisch und umständlich ist, wurden einige andere Möglichkeiten für HRTF Aufnahmen, wie z.B. Kunstkopf-HRTFs, HRTFs von einem „guten Lokalisator" usw untersucht.
Um die Möglichkeiten der HRTF-Synthese in der Wiedergabe zu analysieren, haben wir das BEACHTRON System angewendet. Dieses ist heutzutage wegen seiner Rechnerkapazität und Rechnergeschwindigkeit nicht mehr auf dem Stand der aktuellen Technik. Jedoch ist es eine kostengünstige, benutzerfreundliche und adäquate Realisierung einer virtuellen akustischen Umgebung.
Virtuelle akustische Umgebungen (VAD: Virtual Acoustic Display) haben ein breites Anwendungsgebiet. Für uns war der Ausgangspunkt das frühere GUIB-Projekt (Graphical User Interface for Blind Persons), wobei blinde Rechnerbenutzer die Umwandlung von visuellen Bildschirminformationen in akustische Signale ausgewertet haben. Das Ziel war, die räumliche Aufteilung eines Bildschirms akustisch darstellen zu können. Nach Lautsprechersimulationen muß auch die Anwendungsmöglichkeit der Kopfhörerwiedergabe geprüft werden.
Im ersten Teil haben 25 Probanden das System getestet. Dabei hatten sie die Aufgabe, Schallquellen in der Medianebene zu lokalisieren und auf die Medianabene symmetrische Bewegungen zu identifizieren. Durch die gezielten Fragen an den Versuchspersonen konnten wir die Existenz typischer Fehler der Kopfhörerwiedergabe, wie z.B. Elevationsverlegung, Vorne-Hinten-Vertauschung und Im-Kopf-Lokalisation bestimmen.
Im zweiten Teil ist dann die Diskriminationsfähigkeit (die räumliche Auflösung) festgestellt worden. 40 ungeübte Versuchspersonen haben ein Hörtest unter den folgenden Umständen und mit dem folgenden Vorgaben durchgeführt:

1. Minimum Audible Angle Untersuchung (MAA) in der Horizontal –und Medianebene, um die Diskriminationsfähigkeit der Schallquellen zu bestimmen. Dabei müßen sie den minimalen räumlichen Unterschied zwischen einer nicht beweglichen und einer beweglichen Schallquelle bestimmen und differenzieren können.
2. Bei der Beantwortung der Lokalisationsfragen wählen die Versuchspersonen eine der drei möglichen Antworten aus dem so genannten „3-categorie-forced-choice" aus.
3. Ein zweidimensionales „bildschirmartiges" VAD mit nicht konstanter Schallquellen-entfernung wird angewendet.
4. Die Versuchspersonen haben spezielle Erregersignale, wie weißes Rauschen, Tief -und Hochpaß-gefiltertes Rauschen, je in 300 ms Burts-Paare zu lokalisieren.
5. Die Untersuchungen werden in zwei Richtungen vorgenommen: Einmal entfernt sich die bewegliche Quelle, das andere mal nähert sich im Verhältnis zur Referenzquelle.

Die Ergebnisse wurden in Bezug auf die Lokalisation und auf die GUIB-Anwendung ausgewertet. Typische Eigenschaften der Lokalisation und die mögliche räumliche Aufteilung eines VADs werden für alle drei Signale präsentiert. Dazu gehören:

- das männliche/weibliche Lokalisationsvermögen
- die Durchschnitts –und Maximalwerte
- die Symmetrieeigenschaften
- die horizontale -und vertikale Auflösung im Vergleich
- usw.

III. Messungen der HRTFs

Die Ergebnisse der Hörtests zeigen, daß die Qualität der HRTFs und der HRTF Reproduktion während einer Kopfhörerwiedergabe kritisch ist. Kleine Änderungen in den HRTF-Sets (z.B. nicht individuelle HRTFs) können zu Lokalisationsfehlern, zur schlechteren räumlichen Auflösung und/oder zur Im-Kopf-Lokalisation führen. Deshalb sieht es so aus, als wenn im die HRTFs Gegensatz zur Freifeldübertragung eine wichtigere Rolle in einer virtuellen Simulation spielen. Um die Rolle der HRTFs und deren Feinstruktur zu analysieren brauchen wir ein genaues Meßsystem.
HRTFs kann man an Menschen oder mit einem Kunstkopf messen. Kunstkopfmeßsysteme haben den Vorteil, daß sie die HRTFs ganz genau messen zu können. Eine Messung kann bis zu mehreren Stunden dauern. Sie kann einen guten Signal-Rausch-Abstand (SNR) haben und mit erhöhter Präzision und Reproduzierbarkeit durchgeführt werden.
Wir stellen eine Meßeinrichtung für den schalltoten Raum vor, wo der erreichte SNR und die Präzision besser als bei früheren Untersuchungen ist. Der Signal-Rausch-Abstand kann bis zu 102 dB erreichen. Da die Speicherkapazität der DSP Karte und die Dauer der Messung begrenzt ist, führen wir die Versuche jedoch nur mit durchschnittlich 89 dB durch. Dieser Wert ist für unseren Zweck in jedem Falle ausreichend.
„Reproduzierbarkeit" ist in unserem Falle die Fähigkeit des Systems, die gleiche Schallquellenrichtung (Lautsprecherposition) immer wieder genau einstellen zu können, und in wiederholten Messungen die gleichen Ergebnisse zu liefern. Der Unterschied zwischen unseren gemessenen Übertragungsfunktionen liegt unabhängig von Elevation und Azimut unter 0,5 dB im gesamten Frequenzbereich.
Das benötigt eine umsichtige Kalibrierung des Systems. Die Einstellung des Azimuts erfolgt mit Hilfe eines computergesteuerten Schrittmotors und hat eine Präzision von 1,14%. Die Genauigkeit der Einstellung der Elevation mit einem Laser Pointer erreicht 0,77%. Diese Präzision war nötig um einige wichtige Effekte finden zu können.
Es wurden spezielle Verfahren entwickelt und realisiert, um diese Präzision erzielen zu können:

1. Das Erregersignal ist ein speziell modifiziertes weißes Rauschen: non-MLS pseudo-random noise, periodisch ein- und ausgeschaltet dargeboten. Es enthält alle erwünschten und vorteilhaften Eigenschaften des weißen Rauschens (breitbrandig, flaches Spektrum, „Zufalls-Phasen-Spektrum"). Es ist aber ein deterministisches Signal und kann exakt wiederholt werden. Das

erlaubt eine genaue periodische Wiederholung, und die Kalkulation von Durchschnittsergebnissen, welche den SNR erhöhen. Die Arbeit beschreibt die Herstellung, den Algorithmus, die Simulationen und die Eigenschaften dieses Signals (Frequenzunabhängige optimale SNR für alle Systeme).

2. Der Effekt des Netzbrummens wurde mit einer speziellen „Phasenverlegungsmethode“ um 18 dB reduziert
3. Wegen der Durchschnittsberechnung der Ergebnisse kann der SNR um bis zu +40 dB erhöht werden.

Die Richtungscharakteristik und die Übertragungsschwankungen der Übertragungskette wurden mit einem Referenzsignal eliminiert. Außerdem werden die Impulsantwort, die Rolle des schallabsorbierenden Materials und der Vergleich unserer Ergebnisse mit den Originalergebnissen des Kunstkopfes präsentiert.

IV. Auswertung von spektralen Differenzen in Kunstkopf HRTFs

Mit Hilfe des vorher genannten „genauen Meßsystems" wurden zahlreiche HRTF-Messungen durchgeführt und in der Frequenzdomäne ausgewertet. Dabei haben wir uns nur auf *Differenzen* beschränkt. Die HRTFDs, als der Quotient von zwei HRTFs, zeigen nur Effekte von Veränderungen, ohne die Notwendigkeit von individuellen Messungen zu haben. Durch die Teilung werden individuelle Unterschiede eliminiert und es können Kunstkopf-HRTFs analysiert werden. HRTFDs sind leicht und schnell zu berechnen, sind von individuellen Parametern befreit und können mit diesem System im großen Anzahl genau gemessen werden. Man kann Abweichungen von minimalen 0,5 dB zeigen, die im Zusammenhang mit existierenden physikalischen Phänomenen stehen. HRTFDs sind außerdem geeignet die Genauigkeit des Meßsystems zu analysieren und in einfachen Situationen Nachschall und primäre Reflexionen zu identifizieren.
Es wird eine neuartige 2D-Darstellung gezeigt, in der HRTFs und HRTFDs als Funktion der Frequenz und des Azimuts gleichzeitig abgebildet werden können. Einige Ergebnisse werden auch in einem doppeltlogarithmischen Achsensystem dargestellt.
Die Verarbeitung der Daten besteht aus zwei Teilen: Auswertung von „normalen" HRTFs des Kunstkopfes in der horizontal Ebene und die Auswertung von HRTFDs unter veränderten Hörraumbedingungen.
Wir präsentieren die lokalen und absoluten Maximum -und Minimumwerte der monauralen und binauralen Empfindlichkeit des Gehörs. Es werden auch typische „Grenzfrequenzen" vorgestellt, die in der Auswertung von räumlichen Informationen eine wichtige Rolle spielen. Drei verschiedene Domänen wurden festgestellt. Sie basieren auf der Auswertung von Hochfrequenzinformation auf der lateralen Seite und von tieffrequenten Komponenten auf der kontralateralen Seite. Dabei wurde der Effekt des Kopfschattens näher analysiert.
Mit der Hilfe der HRTFDs sind typische Eigenschaften und Effekte von „Objekten aus dem alltäglichen Leben" analysiert. Dies sind Brille, Haare, Mütze und einige Ergebnisse von Kleidung als Funktion von Azimut und Elevation. Es wurde festgestellt, daß die Charakteristik dieser alltäglichen Objekte die HRTFs mit mehr als 10 dB beeinflussen. Diese Effekte werden anhand der Abbildungen gezeigt. Das Gehör kann aber bei Freifeldbeschallung weder in der Lokalisation noch in der Qualität Änderung wahrnehmen. Wir hören vor oder nach dem Frisör, mit oder ohne Brille nicht anders. Daraus resultiert, daß die Feinstruktur der HRTFs in Wirklichkeit keinen großen

Einfluss auf die Aufarbeitung von Richtungsinformation hat. Das Gehör ist in der Lage sogar große Veränderungen zu „überbrücken”. Das steht aber im Widerspruch zur virtuellen akustischen Umgebungen, wo HRTFs mit der Hilfe von Kopfhörern simuliert werden. Hier können kleine Veränderungen der HRTFs zu einem erhöhten Lokalisationsfehler führen. In diesem Fall ist die Qualität der angewendeten HRTFs wichtig.

V. Zusammenfassung

Basierend auf dieser Entdeckung wird die Qualität von verschiedenen Wiedergabemöglichkeiten gezeigt und verglichen. Es wird darauf hingewiesen, daß mögliche Verbesserungen von binauralen Abspielmöglichkeiten kaum mit der „Verbesserung der HRTFs“ erzielt werden können. Das Problem liegt nämlich in der (Kopfhörer-)Wiedergabe. Das Gehör scheint in der Lage zu sein, bei der Auswertung von räumlichen Information den Hörraum zu erkennen: in Freifeldübertragung kann es die Veränderungen der HRTFs „ausgleichen”, während einer Kopfhörerwiedergabe jedoch nicht mehr.
Das Hauptanliegen dieser Arbeit ist, daß die akustische Information (vor allem die Richtungsinformation) die in den Schallwellen kodiert ist, und durch das Gehör in Freifeldübertragung trotz sehr variablen HRTFs dekodiert werden kann. Die akustische Umgebung in der Nähe des Kopfes „verzerrt” die Übertragung, und kann die HRTFs um mehr als 10 dB beeinflussen. Das Gehör bleibt trotzdem in der Lage alle akustisch relevanten Informationen zu bekommen. Auf der anderen Seite sind wir durch Kopfhörerwiedergabe in der Lage eine entzerrungsfreie Übertragung bis zum Trommelfell zu schaffen, allerdings mit beschränkter Informationsübertagung. In diesem Fall reagiert das Gehör sehr empfindlich auf die HRTFs, und oft bekommen wir nur eine schlechtere räumliche Auflösung und Lokalisationsfehler während der Wiedergabe.
Wir können die Schlussfolgerung ziehen, daß die Feinstruktur der HRTFs keine wichtige Rolle spielt. Stattdessen muß das Phasenspektrum und die zeitlichen Veränderungen der HRTFs in Betracht gezogen werden. Z.B. sind kleine Kopfbewegungen wichtig um Im-Kopf-Lokalisation zu vermeiden. Das Gehirn scheint eine wichtige Rolle nicht nur bei Aufarbeitung im Mittelohr, sondern sogar bei der Auswertung von HRTFs und Außenohrdaten zu spielen. Deswegen muß die Kopfhörerwiedergabe, als Wiedergabemöglichkeit näher untersucht werden. Die Qualität der binauralen Abspielsysteme kann durch die Verbesserung von Kopfhörern wahrscheinlich besser erzielt werden, als durch die Verbesserung des Kunstkopfes oder durch die Genauigkeit der HRTF-Sets.
Zukünftige Arbeit, Ausblick und weitere Untersuchungsmöglichkeiten sind angedeutet.

11 Appendix A

References in alphabetical order

Abel 142
Akeroyd 42
Anderson, M 180
Algazi 24 128
Angell 147
Aokata 18
Arruda 46
Asano 65 96
Ashmed 170
Aures 154
Avendano 24 128
Awad 127
Batteau 53
Begault 133 180
Behrens 146
Békésy 1 2 3
Belendiuk 20
Berényi 45
Bernstein 42
Blattner 125
Blauert 5 16 17 22 31 63 94 121 132
Bodden 64 97
Boehnke 160
Boerger 132
Borish 147
Brinkmann 126
Bronkhorst 134
Brungart 52
Burger 123 124
Burke 143
Burkhard 69 90
Butler 20 21 58 130 143
Campbell 105
Canavét 64

Carlile 67 95
Cesarano 124
Cheng 157
Cheung 68
Chowing 109
Christensen 4
Cohen 25
Colburn 26
Consoli 142
Crispien 112 182
Crystal River Eng. 111
Dalenbäck 113
Damaske 66
Domnitz 26
Duda 24 128 175
Elfner 27
Ericson 121 174
Feddersen 99
Feldman 104
Feldtkeller 129
Fellbaum 182
Fisher 15
Foster 47 152 165
Freedman 15
Fukodome 87
Furness 25
Gaik 98
Galginaitis 32
Gan 115
Gardner, M. B. 57
Gardner, R. S. 57
Gardner, W. G. 179
Genuit 34 44 76 82 83 85 86 88-93 148
Gierlich 89 92
Giguere 142
Gotoh 40
Grabke 64
Grantham 110

Green 59 151
Greenberg 125
Haigh 54
Hall 107
Hammershøi 8 9 10 11 13 49 61 62 72
Hartmann 12 100
Hartung 35 42 64
Hatziantoniu 159
Haustein 168
Hebrank 56
Heffner, H.E. 169
Heffner, R.S. 169
Helle 80
Hill 79
Horner 68
Hundeboll 61
Ickler 114
Illényi 45 145 163
Jacob 114
Jeffress 28 30 33 99
Jensen 9 10 13 62 72
Jin 67
Jorgensen 114
Kawaura 65
Kimura 40
Kinkel 106
Kirkeby 79
Kistler 46 48 119 120 161
Kleiner 73 113
Kollmeier 106
Kopco 50
König 138 139 140 141
Kuhn 51 75
Larsen 61
Laws 36 38 63 132
Lehnert 121
Letsos 143
Lindemann 176
Litovsky 170

Maijala 71
Makous 117
Manlingas 102
Martin 84
Maschke 146
Massarani 153
Mazurier 124
McAnally 84
McFadden 28 33 43
McKinley 122 174
Mehrgart 7 55
Mellert 7 55
Middlebrooks 14 59 116 117 118 151 173
Mills 29 101
Minnaar 4
Møller 4 8 9 10 11 13 60 61 62 70 72
Morimoto 18
Mourjopoulos 159
Moushegian 30
Musicant 58 103
Müller 153
Naunton 130
Nelson 79
Nielsen 150
Noble 41
Oldfield 135 172
Olesen 4
Oppenheim 155
Papsin 142
Parker 135 172
Pasanen 43
Perrott 27 102 103 108
Petrie 112
Phillips 160
Pickles 181
Platte 34 88
Plenge 39 167
Pralong 95
Prante 146

Rabinowitz 52
Rakerd 100
Richter 126
Rife 149
Ripka 166
Roffler 21
Sachs 69
Sagot 124
Sahrhage 121
Sakamoto 40
Sandel 99
Sandvad 49
Santarelli 50
Schafer 155
Schenk 156
Schenkel 67
Scherer 136
Schirmer 168
Schmitz 78
Schöne 77 81
Senova 84
Shaw 6 23 162
Shinn-Cunningham 50
Sibinger 76
Siebrasse 55
Sone 65 96
Sorensen 9 10 13 62 72
Strauss 121
Strybel 102
Sumikawa 125
Suzuki 65 96
Svensson 113 150
Takahashi 64
Tan 115
Tarnóczy 74
Tatai 158
Teas 99
Terhardt 154

Theile 166
Tietze 156
Toole 37
Trahiotis 42
Trautman 68
Tucker 108
Vanderkooy 149
Wagener 66
Wakefield 157
Watkins 19
Wenzel 25 46 47 165 180
Wersényi 131 137 144 145 158 163
Wettschurek 177 178
Wightman 46 48 120 161
Wilkens 164
Wilson 56
Wright 56
Xiang 148
Zhou 151
Zwicker 129
Zwislocki 104

12 Appendix B

AUTHOR	SIGNAL, REMARKS	RESULTS
Blauert [5]	free-field, noise signal, absolute minimum value	1°
	free-field, sound-klicks	0,75°-2°
	free-field, sinusoid signal (pure tones)	1°-4°
	free-field, narrow-band noise, Gauss-noise	1,4°-3,3°
	free-field , speech sample	0,9°-1,5°
Haustein, Schirmer [168]	free-field , broadband noise	3,2°
	free-field, 100 ms white noise impulses	±3,6° (front) ±9 to 10° (side)
Hartmann [12]	free-field, absolute minimum value	1°-2°
Heffner, Heffner [169]	free-field, noise signal, MAA value	1,3°-1,8° (front) 9°-10° (side)
Litovsky, Ashmed [170]	free-field, MAA value	1°
Kremer [171]	n. a.	3°-5°
Barfield, Furness [25]	free-field, MAA value	1°-5° (front) 5°-10° (side)
	free-field, absolute measurement	10°-20°
Begault [133]	free-field, MAA value, broadband signal, “optimal conditions”	1°
Oldfield, Parker [135, 172]	headphone, azimuthal mean value, absolute measurement	9°
	headphone, azimuth errors with HRTF filtering	4°-6°
	headphone, azimuth errors without HRTF filtering	11,9°

Middlebrooks [173]	free-field, avg. error of 150 ms broadband noise	5,8°
McKinley, Ericson [174]	avg. error, MAA value	5°
	avg. error of octave-band noise, absolute m.	4,4°-5,9°
	avg. error of pink noise, absolute meas.	6°-7°
	MAA for 500 Hz sinusoid signal	4°-5°
Makous, Middlebrooks [117]	free-field, absolute measurement, minimal avg. error for 150 ms signal bursts	2°
Middlebrooks [118]	headphone, avg. error, non-individual HRTFs (other-ear-condition) headphone, avg. error,, individual HRTFs (own-ear-condition)	17,1° 14,7°
Duda [175]	avg. error of a Maximum-likelihood assumption of dummy-head data avg. error (measured) with human HRTFs avg. error for broadband signals (12kHz)	5,1° 4,5° 3,4°
Gardner [179]	pink noise bursts of 250 ms, absolute measurement avg. angle error (headphones) avg. angle error (loudspeaker)	14,3° 12,1°
Begault, Wenzel [180]	avg. error (generic HRTF) avg. error (individual HRTF)	23° 20°

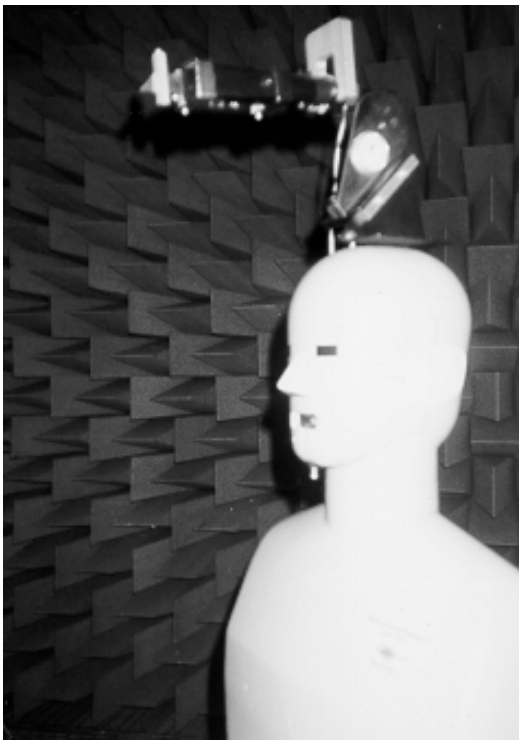
Table 15. Summary of localization results in the horizontal plane.

AUTHOR	SIGNAL, REMARKS	RESULTS
Blauert [5]	free-field, unknown speaker, 20 Subjects	17°
	free-field, known speaker, 7 Subjects	9° (front) ±10° (at δ=36°) ±13 to ±22° (above)
	free-field, white noise, 2 Subjects	4°
Wettschurek [177, 178]	free-field, MAA value, white noise (8-10° Standard Deviation)	±4° (front) ±10° (above)
	low-frequency noise (4 kHz cut-off freq.)	±8° (front) ±20° (above)
Oldfield, Parker [135, 172]	headphone, elev. mean value, absolute meas.	12°
	headphone, elev. error with HRTF filtering	6°-8°
	headphone, elev. error without HRTF filtering	21,9°
Wenzel, Foster [165]	free-field, avg. error, 16 subjects	ca. 25° (lower elevations,front) ca. 22° (side)
	headphone, non-individual HRTFs, 16 subjects	ca. 24° (lower elevations,front) ca. 23° (side)
Wightman, Kistler [120]	free-field, avg. error	ca. 20° (lower elevations,front) ca. 18° (side)

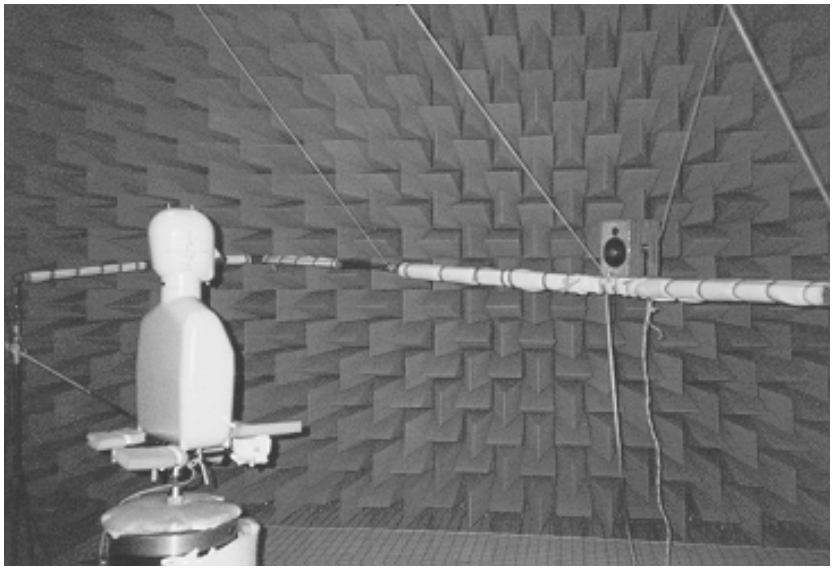
	headphone, avg. error	ca. 21° (lower elevations,front) ca. 20° (side)
Middlebrooks [173]	avg. error of 150 ms broadband noise	5,7°
McKinley, Ericson [174]	headphone, MAA value, dummy-head HRTF	30°-35°
Møller [70, 72]	headphone, relative localization error, using HRTFs of a random human headphone, relative localization error, using HRTFs of a dummy-head HRTFs	36% 55%
Makous, Middlebrooks [117]	free-field, absolute measurement, minimal avg. error for 150 ms signal bursts (94% of the subjects are within of 10° Standard Deviation)	3,5° (front) 20° (side)
Duda [175]	avg. error of a Maximum-likelihood assumption of dummy-head data avg. error (measured) with human HRTFs avg. error for broadband signals (12kHz)	12° 19,2° 17,2°
Gardner [179]	pink noise bursts of 250 ms, absolute measurement avg. angle error (headphones) avg. angle error (loudspeaker)	34,2° 32,4°
Begault, Wenzel [180]	avg. error (individual HRTF)	17-19°

Table 16. Summary of localization results in the median plane.

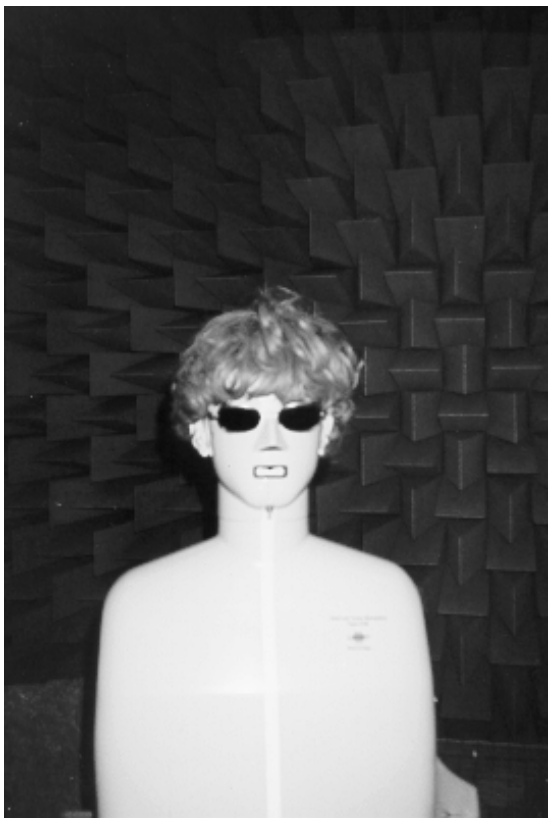
13 Appendix C



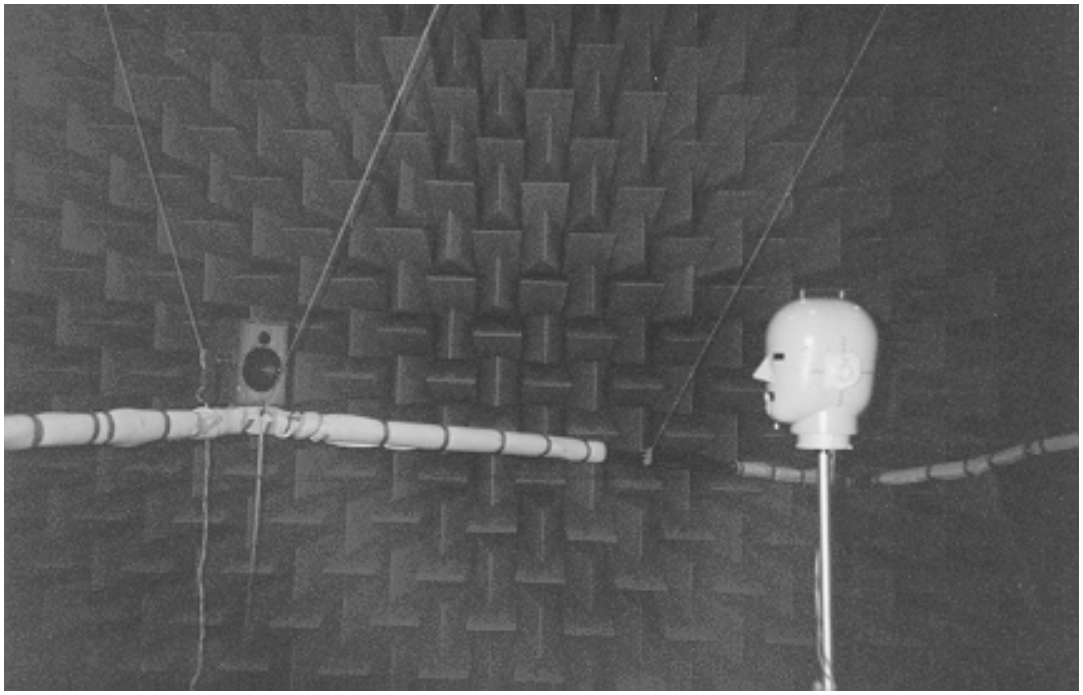
The laser targeting system.



Measuring the „bare“ torso in the anechoic room.

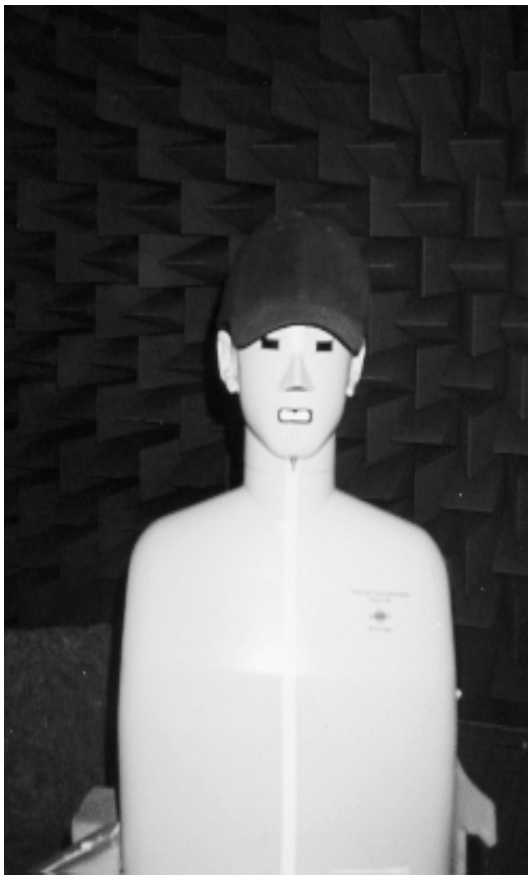


HRTFs of the dummy-head wearing hair and glasses.



Measurement without the torso.

The dressed torso using clothing, hair and baseball cap together.



Measurement using baseball cap.

14 Appendix E

Abbreviations

AVG	Average
BK	Brüel & Kjær
DHRTF	Derivated Head-Related Transfer Function
δ	Elevation degree in the head-related coordinate system
FEC	Free ear coupled (diffuse-field equalized headphone)
FFT	Fast Fourier Transform
FIR	Finite Impulse Response
φ	Azimuthal degree in the head-related coordinate system
GUIB	Graphical User Interface for Blind Persons (int. project)
HAT	Head and Torso Simulator (dummy-head)
HRIR	Head-Related Impulse Response
HRTF	Head-Related Transfer Function
HRTFD	Head-Related Transfer Function Differences
HVT	High Voltage Transformer
IFFT	Inverse Fourier Transform
ILD	Interaural Level Differences
ITD	Interaural Time Differences (Delays)
JND	Just Noticeable Difference
LTS	Laser Targeting System
MAA	Minimum Audible Angle
MAMA	Minimum Audible Movement Angle
MAX	Maximum, maximal
MIN	Minimum, minimal
MLS	Minimum Length Sequence
RES(FFT)	Resolution of the FFT (sample frequency/number of FFT)
RMS	Root Mean Square
SNR	Signal-to-Noise Ratio
SPL	Sound Pressure Level
TF	Transfer Function
TFD_i	Transfer Function Differences
VAD	Virtual Audio Display

15 Appendix F

Table 4 shows results from the median plane for all subjects and signals: signal A (above), signal B (middle) and signal C (below). Values are presented from the direction „up" (on the left) and „down" (on the right) as averaged, maximal and minimal values. Male and female subjects' results can be compared separately as well. The first value indicates the nearest point to the origin, where the subjects were able to discriminate the sources with certainty. The static reference signal is in the origin, the second impulse is moving first away, than toward the reference point. E.g., for signal A is the nearest discriminated sound source 16,7° higher than the origin (on average over every subject). This means a sound source further than 16,7° can be discriminated from the source in the origin independent of the direction of the moving source (away and toward the reference point). Maximal value of 32° (worst case) and best value of 8° was measured.
The second value shows, that the next nearest location of a sound source is another 16,6° higher from this point. In the right column, the values are shown the same way, only in the direction „down". In both directions only two new sound locations can be determined maximal. As a final result, sound sources can be placed at -31,6°; -16°, 0°, +16,7° and +33,3° in the median plane (on average) for correct separation of the sound sources.
Figure 25 shows the average values from the median plane for all signals up and down as well. In *vertical* directions no significant differences appear between female and male subjects (Table 4). The average resolution for signal A is about 15-17°, 19-24° for signal B and 18-23° for signal C. The maximum values can reach the double of the average value; the minimum values could be 10-50% of the mean value.
Tables 5 to 7 show results the same way, but in the horizontal plane for signals A, B and C respectively. Due to the better resolution 6 possible source locations can be determined left and right from the origin, which results in a total number of 13 sound locations. Furthermore, the last row of the tables indicates the difference between the nearby average values to show the effect of averaging.

SIGNAL A

UP			**DOWN**		
AVG [DEG]	16,7	33,3	AVG [DEG]	16,0	31,6
MAX [DEG]	32	48	MAX [DEG]	29	53
MIN [DEG]	8	17	MIN [DEG]	3	11
Male only:					
AVG [DEG]	15,8	32,6	AVG [DEG]	16,4	32,6
MAX [DEG]	29	48	MAX [DEG]	29	53
MIN [DEG]	8	17	MIN [DEG]	3	11
Female only:					
AVG [DEG]	17,8	34,3	AVG [DEG]	15,5	30,5
MAX [DEG]	32	44	MAX [DEG]	28	45
MIN [DEG]	12	24	MIN [DEG]	10	20

SIGNAL B

UP			**DOWN**		
AVG [DEG]	24,0	37,4	AVG [DEG]	18,5	35,9
MAX [DEG]	49	53	MAX [DEG]	27	54
MIN [DEG]	13	26	MIN [DEG]	13	25
Male only:					
AVG [DEG]	26,4	39,5	AVG [DEG]	18,5	35,8
MAX [DEG]	49	52	MAX [DEG]	27	54
MIN [DEG]	13	27	MIN [DEG]	13	25
Female only:					
AVG [DEG]	21,5	35,1	AVG [DEG]	18,7	36,0
MAX [DEG]	38	53	MAX [DEG]	24	46
MIN [DEG]	13	26	MIN [DEG]	13	27

SIGNAL C					
UP			**DOWN**		
AVG [DEG]	17,8	35,2	AVG [DEG]	23,3	35,5
MAX [DEG]	34	55	MAX [DEG]	45	55
MIN [DEG]	5	16	MIN [DEG]	13	27
Male only:					
AVG [DEG]	18,0	35,9	AVG [DEG]	24,1	36,1
MAX [DEG]	34	55	MAX [DEG]	45	55
MIN [DEG]	5	16	MIN [DEG]	13	27
Female only:					
AVG [DEG]	17,7	34,3	AVG [DEG]	21,3	34,0
MAX [DEG]	27	45	MAX [DEG]	40	42
MIN [DEG]	13	22	MIN [DEG]	14	27

Table 4. AVG, MIN and MAX values in the median plane for all subjects and signals. Left column shows first reference points, right columns the second reference points.

SIGNAL A						
LEFT						
AVG [DEG]	9,4	18,9	28,3	36,3	44,0	47,9
MAX [DEG]	20	41	55	54	63	70
MIN [DEG]	4	8	13	18	24	31
DIFFERENCE [DEG]		9,5	9,4	8,0	7,7	3,9
Male only						
AVG [DEG]	8,8	18,0	27,2	33,9	39,4	45,3
MAX [DEG]	20	41	55	54	60	70
MIN [DEG]	5	11	16	21	27	32
DIFFERENCE [DEG]		9,3	9,2	6,7	5,5	5,9
Female only						
AVG [DEG]	10,1	19,9	29,5	38,9	48,3	51,2
MAX [DEG]	16	27	41	49	63	69
MIN [DEG]	7	14	21	28	34	42
DIFFERENCE [DEG]		9,8	9,6	9,4	9,4	2,9
RIGHT						
AVG [DEG]	7,6	15,8	24,9	34,0	42,5	48,6
MAX [DEG]	14	26	38	57	60	69
MIN [DEG]	3	7	12	18	24	29
DIFFERENCE [DEG]		8,1	9,2	9,0	8,6	6,1
Male only						
AVG [DEG]	6,5	13,9	22,9	31,1	40,3	45,2
MAX [DEG]	9	20	38	48	60	69
MIN [DEG]	3	7	12	18	24	29
DIFFERENCE [DEG]		7,4	9,0	8,2	9,2	5,0
Female only						
AVG [DEG]	8,9	17,9	27,3	37,3	45,3	52,6
MAX [DEG]	14	26	37	57	58	66
MIN [DEG]	5	11	18	24	31	40
DIFFERENCE [DEG]		9,0	9,4	10,0	8,0	7,3

Table 5. AVG, MIN and MAX values in the horizontal plane for all subjects (signal A). The differences between the nearby AVG values are also shown. Columns show reference points discriminated by the subjects from the origin to the sides.

SIGNAL B
LEFT

AVG [DEG]	10,5	21,0	31,7	40,1	46,6	51,5
MAX [DEG]	22	38	53	58	70	65
MIN [DEG]	5	9	14	19	25	32
DIFFERENCE [DEG]		10,5	10,7	8,4	6,4	5,0
Male only:						
AVG [DEG]	10,6	21,4	32,3	39,6	42,1	47,1
MAX [DEG]	22	38	53	58	62	65
MIN [DEG]	5	9	14	19	25	32
DIFFERENCE [DEG]		10,8	10,9	7,3	2,5	5,0
Female only: AVG [DEG]	10,3	20,5	30,7	40,9	51,0	56,5
MAX [DEG]	15	28	39	53	70	60
MIN [DEG]	8	16	25	34	43	53
DIFFERENCE [DEG]		10,2	10,2	10,2	10,1	5,5
RIGHT:						
AVG [DEG]	8,3	17,4	27,6	38,3	44,2	51,1
MAX [DEG]	16	27	46	64	60	68
MIN [DEG]	4	8	12	17	22	28
DIFFERENCE [DEG]		9,1	10,2	10,7	5,8	6,9
Male only:						
AVG [DEG]	7,4	16,1	25,9	36,2	41,9	46,5
MAX [DEG]	12	27	46	63	60	64
MIN [DEG]	4	8	12	17	22	28
DIFFERENCE [DEG]		8,6	9,8	10,3	5,7	4,6
Female only:						
AVG [DEG]	9,5	19,4	30,2	41,5	47,8	57,5
MAX [DEG]	16	27	44	64	57	68
MIN [DEG]	5	10	23	31	37	44
DIFFERENCE [DEG]		9,8	10,8	11,3	6,3	9,7

Table 6. AVG, MIN and MAX values in the horizontal plane for all subjects (signal B). The differences between the nearby AVG values are also shown. Columns show reference points discriminated by the subjects from the origin to the sides.

SIGNAL C						
LEFT						
AVG [DEG]	11,6	22,1	31,2	39,3	45,2	47,7
MAX [DEG]	27	45	56	63	63	63
MIN [DEG]	4	10	15	22	27	32
DIFFERENCE [DEG]		10,5	9,0	8,1	5,9	2,4
Male only:						
AVG [DEG]	10,6	21,0	29,3	38,3	42,7	44,7
MAX [DEG]	24	42	52	63	63	63
MIN [DEG]	4	10	15	22	27	32
DIFFERENCE [DEG]		10,4	8,3	9,0	4,4	2,0
Female only:						
AVG [DEG]	13,0	23,7	33,7	40,7	48,7	52,2
MAX [DEG]	27	45	56	51	61	56
MIN [DEG]	8	13	20	26	33	41
DIFFERENCE [DEG]		10,7	10,0	7,0	8,0	3,5
RIGHT						
AVG [DEG]	9,3	17,8	27,8	36,8	47,1	51,2
MAX [DEG]	18	31	54	53	64	66
MIN [DEG]	4	8	11	14	17	21
DIFFERENCE [DEG]		8,5	10,0	8,9	10,4	4,1
Male only:						
AVG [DEG]	8,8	17,0	27,8	36,0	46,6	48,2
MAX [DEG]	18	31	54	53	64	64
MIN [DEG]	4	8	11	14	17	21
DIFFERENCE [DEG]		8,3	10,8	8,3	10,6	1,7
Female only:						
AVG [DEG]	10,1	19,0	27,9	37,8	47,9	54,5
MAX [DEG]	15	26	37	48	59	66
MIN [DEG]	7	14	21	29	38	46
DIFFERENCE [DEG]		8,9	8,9	9,9	10,1	6,6

Table 7. AVG, MIN and MAX values in the horizontal plane for all subjects (signal C). The differences between the nearby AVG values are also shown. Columns show reference points discriminated by the subjects from the origin to the sides.

RESUME

Personal data:

First name: GYÖRGY
Family name: WERSÉNYI
Date of birth: 17 January, 1975
Place of birth: Győr, Hungary
Nationality: Hungarian

Mailing address:

Széchenyi István University
Department of Telecommunications
H-9024
Győr
Egyetem tér 1.
Hungary
e-mail: wersenyi@sparc.core.hu

Education:

1989-1993	Révai Miklós Gymnasium, Győr, Hungary
1993-1998	M.Sc. degree in Electrical Engineering at the Technical University of Budapest. Graduated at the Department for Telecommunications and Telematics, "Békésy György" Acoustical Research Laboratory
1998-2002	Full-time Ph.D. student at the University "Békésy György" Acoustical Research Laboratory (Hearing and Speech communication group)
1998-2002	Research associate and lecturer in studio technologies and technical acoustics at the Széchenyi István Technical University in Győr, Hungary.
2000-2001	DAAD scholarship - 1 year research activity Brandenburgische Technische Universität, Cottbus Lehrstuhl Kommunikationstechnik Cottbus, Germany

1997-	Loudspeaker and headphone calibration Standardization and calibration of noise measurement systems Acoustical measurements Audiology
2002	"Huszty Dénes"-Prize for new achievements in acoustics, hearing research and related sciences
2002	Ph.D. degree in electrical engineering
2003-	Research associate, lecturer Head of the multimedia-laboratory Széchenyi István Technical University, Gyõr, Hungary

Language skills:

Hungarian	Native language
German	Expert level
English	Expert level
Spanish	Novice level

Computer skills:
Basic programming in C++ language and Windows applications user routine.

Research activities:
Human spatial and directional hearing, localisation and resolution in virtual and real-life environments, decoding and transmission of acoustical information, hearing system modelling, dummy-head and acoustical measurement techniques, noise measurement, multi-channel audio applications, digital audio coding (lossy and lossless), studio technologies, hearing aids and auditory measurements (audiometers), virtual reality.

Hobbies, other: home theatre systems, heavy metal music, card games, football, body building, Internet, sauna, computer games, horror/fantasy movies & books, spice foods, travelling, Mexican and Italian food.

Membership:

1997-	Hungarian Optical, Acoustical and Filmtechnical Society	(OPAKFI)
2000-	Audio Engineering Society	(AES)

Budapest, 31. July, 2002 Wersényi György

PUBLICATIONS

[1] WERSÉNYI, GY., TÁTRAI, R., *Measuring and discussion of an HRTF measurement system using artificial head.* Scienfitic Student Conference, Technical University of Budapest Department of Telecommunications and Telematics, Budapest, 1997.

[2] WERSÉNYI, GY., *Hardware and software realization of a dummy-head measurement system for spatial hearing investigations.* Graduation thesis, Technical University of Budapest - Department of Telecommunications and Telematics, Budapest, 1998.

[3] WERSÉNYI, GY., *Spatial hearing.* Educational help (30 pages in Hungarian), http://www.ttt.bme.hu/Num7/terhall.htm

[4] WERSÉNYI, GY., BERÉNYI, P., *Measuring of the Head Related Transfer Functions.* Acoustical Review, Vol. IV., 1.-4., 1999, Budapest. /**35-41**/.

[5] ILLÉNYI, A., WERSÉNYI, GY., *Discrepancy in binaural tests and in measurements of sound field parameters.* Proceedings of the International Békésy Centenary Conference on hearing and related sciences, 1999, Budapest. /**160-165**/.

[6] WERSÉNYI, GY., ILLÉNYI, A., *Averaged speech signal samples generated by speech signal method.* Proceedings of the International Békésy Centenary Conference on hearing and related sciences, 1999, Budapest. /**115-120**/.

[7] WERSÉNYI, GY., *Measurement system upgrading for more precise measuring of the Head-Related Transfer Functions*, Proc. of Inter Noise 2000 International Conference, Vol.II., August 27-30, 2000, Nice, France. /**1173-1176**/.

[8] WERSÉNYI, GY., TATAI, P., *Detection of reflections in free-field directional hearing by waveform analysis of accurate dummy-head HRTFs.* IEEE Instrumentation and Measurement Technology Conference, Budapest, 2001. May 21-23. /**606-609**/.

[9] WERSÉNYI, GY., *Acoustic Signal Processing for Listening Tests in Virtual Audio*. 2001 Polish-Czech-Hungarian Workshop on Circuit Theory, Signal Processing, and Telecommunication Networks, Budapest, Hungary, 14-17 Sep. 2001. /**175-181**/.

[10] WERSÉNYI, GY., *HRTFs in Human Localization: Measurement, Spectral Evaluation and Practical Use in Virtaul Audio Environment*. Ph.D. doctoral thesis, Technical University of Brandenburg, Cottbus, Germany, 2002.

[11] WERSÉNYI, GY., ILLÉNYI, A., *Measurement Accuracy of a Dummy-Head Measurement System,* revised by the Journal of AES.

[12] WERSÉNYI, GY., ILLÉNYI, A., *Evaluation of Differences in Dummy-Head HRTFs Caused by the Acoustical Environment Near to the Head,* revised by the ACUSTICA united with acta acoustica.

[13] WERSÉNYI GY., *Modelling of the human auditory system and information decoding,* in preparation (Acoustical Review, in Hungarian).

Erklärung

Ich bestätige, dass ich die Dissertation selbständig verfasst und alle in Anspruch genommenen Hilfen in der Dissertation angegeben habe. Es wurden keine gleichzeitigen oder früheren Promotionsanträge in Zusammenhang mit dieser Dissertation gestellt. Mir ist die geltende Promotionsordnung bekannt.

22-03-2002, Budapest WERSÉNYI, György