Technická Univerzita v Košiciach
Fakulta Elektrotechniky a Informatiky
Katedra Kybernetiky a Umelej Inteligencie

Priestorové odmaskovanie širokospektrálnych stimulov vo virtuálnom sluchovom prostredí

Vedúci diplomovej práce:
Ing. Norbert Kopčo, PhD.

Konzultant diplomovej práce:
Ing. Norbert Kopčo, PhD.

Diplomant:
Branislav Benikovský

Košice 2004

Technical University of Košice
Faculty of Electrical Engineering and Informatics
Department of Cybernetics and Artificial Intelligence

Spatial unmasking of broadband stimuli in a virtual auditory environment

Thesis advisor:
Ing. Norbert Kopčo, PhD.

Consultant
Ing. Norbert Kopčo, PhD.

Author:
Branislav Benikovský

Košice 2004

Declaration

I hereby declare that the work described in this thesis has been done solely by myself and that all literature I used is listed in the references.

In Košice 10.05.2004

. . . .
signature

Acknowledgment

I would like to thank to my thesis advisor and consultant Ing. Norber Kopčo, PhD.
for his valuable advices and suggestions, which directed my work on this thesis.

My acknowledgment belongs to my family and to God for encouraging me during my study at the university. I also want to thank to all my friends who kept their fingers crossed and thought of me. Finally, I thank to all my roommates for helping me solve many problems I’ve encountered while working on this thesis.

Názov práce :	Priestorové odmaskovanie širokospekrálnych stimulov vo virtuálnom sluchovom prostredí.
Katedra :	Katedra Kybernetiky a Umelej Inteligencie, TU FEI Košice
Autor :	Branislav Benikovský
Vedúci DP :	Ing. Norbert Kopčo, PhD.
Konzultant DP :	Ing. Norbert Kopčo, PhD.
Dátum :	01.05.2004
Kľúčové slová :	Priestorové odmaskovanie, efekt koktailovej párty, priestorový sluch, binaurálny sluch, detekcia, modulované signály, periférny sluchový systém, centrálny sluchový systém
Anotácia :	Táto štúdia prezentuje výsledky psychofyzikálneho experimentu, ktorý meral priestorovú závislosť detektability širokospekrálnych stimulov maskovaných šumom vo virtuálnom sluchovom prostredí. Na analýzu výsledkov je použitý model Single-Best-Filter, ktorý modeluje spracovanie zvukových podnetov v periférnej sluchovej dráhe. Výsledky naznačujú, že binaurálny príspevok nie je zanedbateľný a podstatne napomáha pri priestorovom odmaskovaní o viac ako 5 dB. Všetky merané prahy pomocou S-B-F boli nižšie ako širokospektrálne prahy.

Thesis title :	Spatial unmasking of broadband stimuli in a virtual auditory environment
Department :	Department of Cybernetics and Artificial Intelligence, TU FEI Košice
Author :	Branislav Benikovský
Supervisor :	Ing. Norbert Kopčo, PhD.
Tutor :	Ing. Norbert Kopčo, PhD.
Date :	01.05.2004
Keywords :	Spatial unmasking, Cocktail party effect, binaural hearing, detection, modulated signals, peripheral auditory system, central auditory system
Annotation :	This study presents results of a psychophysical experiment that measured the influence of spatial position on the detectability of broadband stimuli masked by noise in virtual auditory environment. To analyze the results the Single-Best-Filter model was used, which models the processing of sound signals in peripheral auditory pathway. The results show that binaural contribution to spatial unmasking can be as much as 5 dB. The single best filtered thresholds were always lower than the broadband thresholds.

Summary

List of Abbreviations

BIN – binaural

BMLD – binaural masking-level difference

BRD – broadband

CF – central frequency

HF – high-frequency

HRIR – Head Related Impulse Responses

HRTF – head-related transfer function

ILD – interaural level difference

IPD – interaural phase difference

ITD – interaural time difference

LF – low-frequency

M – Masker

MLD – masking-level difference

N – Noise

RMS – root mean square

S – Signal

S-B-F – Single-Best-Filter

SNR – Signal Noise to Ratio

SRM – Spatial Release from Masking

T – Target

TDT – Tucker-Davis Technologies

1. Introduction

The human ability to understand speech in complicated (e.g., noisy) environment is very good. This phenomenon is also called the „cocktail-party effect". It's supposed that the cocktail-party effect is connected with the effect of „spatial unmasking" („spatial release from masking", SRM). SRM means that the human being can very precisely detect a tone in the presence of noise, when the source of the tone signal is spatially separated from the source of the noise [18].

The aim of this study is to characterize some attributes of the neural mechanisms that contribute to spatial unmasking.

Previous psychophysical studies of spatial unmasking showed, that SRM is determinated by two factors (see aso. Kopco and Shinn-Cunningham, 2000 for SRM of pure tones, or Good, Gilkey and Ball, 1997 for SRM wide-spectrum sound):

SRM of high-frequency (HF) stimuli is primary determinated by energetic factors, i.e., by the fact that when the signal source position is changed from the noise position, the signal-to-noise ratio (SNR) at one of the ear improves.

Besides energetic factors, unmasking of low-frequency stimulus is also caused by binaural factors, i.e., by the ability of the human auditory system to compare time and intensity, with which a sound has been received by the left and right ear, and on the basis of this comparison to continue in separating of sounds.

The present study extends results of a previous study [14] and tries to test whether information from peripheral auditory channels is combined in the central auditory system or not. Stimuli of sound are processed in the peripheral auditory pathway in parallel frequency channels, which generate distributed neural code of sound stimuli. There are two alternatives of how central auditory system can process information from peripheral channels when stimulated by a broadband sound:

1) Central auditory system elects only one peripheral auditory channel (the one with the most favorable SNR) and SNR in this narrow-band channel determines the SNR needed for detection of the broadband sound

2) Central auditory system combines information from several peripheral channels and this combination causes that threshold SNR needed for detection of a broadband sound is lower than the SNR needed for detection by a single channel.

Comparison of results of electrophysiological measurements of neural activity in the cat inferior colliculus with human behavioral data [8] suggests that the alternative of using the single-best-filter peripheral auditory channel is more probable.

The aim of this study is to verify the hypothesis about using the single-best-filter peripheral auditory system model for wide-band stimulus. In case the single-best-filter hypothesis is not confirmed, we would try to use an alternative model proposed by Dau [20] to model our data. The Dau model combines information from multiple channels to determine the overall performance.

1.1 Formulation of the Problem

The goal of this thesis is to study one aspect the human spatial auditory perception. Specifically, we perform an experimental study that looks on spatial unmasking of broadband "chirp" stimuli in simulated anechoic environment. The work on the thesis can be defined into several phases: 1) review of basic information about spatial hearing and detailed analysis of spatial unmasking and of methods used in experimental studies of spatial unmasking; 2) review and implementation of the models (Single-Best-Filter model and Dau's model) that are used to analyze the experimental data; 3) data collection and analysis of the results; and 4) discussion of the results obtained in the study. The goals of this thesis can be summarized in the following steps:

1. Prepare a review of the problem of spatial auditory perception with the emphasis on spatial unmasking, the cocktail party effect and on Dau's model of processing of sound signals in peripheral auditory pathway.

2. Prepare the procedure for experimental measurement of auditory thresholds of broadband sounds in MATLAB

3. Collect experimental data on four healthy voluntary human listeners

4. Analyze and to evaluate the experimental data in the context of Dau's model.

1.2 Description of Diploma Work

This diploma work is divided into 10 chapters:

Chapter 2. : ***Background*** describes problems of spatial auditory perception, spatial unmasking, and cocktail party problem. Also explained are the basic ideas, terminology, and neural mechanisms of spatial hearing and binaural masking.

Chapter 3. : ***Models*** Two models that can generate predictions of the experimental data are introduced by presenting basics schemas and main principles of their functioning.

Chapter 4. : ***Experiment: Motivation and Hypothesis*** describes the main questions addressed in this study and our hypotheses about the expected results of the study.

Chapter 5. : ***Experimental Methods*** describes the implementation of the experiment. Specifically, this includes the preparation of the experimental procedure, preparation of the experimental stimuli as well as of the experiment itself, a pilot study, as well as audiological measurements used to check hearing of the experimental subjects.

Chapter 6. : ***Results*** describes the results of the analysis of experimental data and their comparison with predictions of the models.

Chapter 7. : ***Conclusions and Summary*** evaluates completion of individual tasks defined for this study.

2. Background

2.1 Hearing

Sound constantly surrounds us and informs us about many objects in our world. Determining the sources of sounds is one our most important biological traits. Any animal's ability to locate food, avoid predators, find a mate, and communicate depends on being able to determine sound sources. In order to determine sound sources we need to know that an object exists, what it is, where it is, if it is moving, and so on.

When we listen, sounds from various sources are combined into one complex sound field and do not reach us as individual sounds. We do not have separate „pipelines“ to our senses for each object in our environment, as the ancient Greeks suggested. We receive one sound input that is made up of sounds from all sources in our environment. Figure 2.1 depicts this situation for several musical instruments. Whether we listen at concert (where all the instruments exists as sound sources) or to the radio (where only the radio loudspeaker is the source of sound), we can determine the various instruments in the band. In order to determine sound sources, the nervous system first processes the complex sound field by translating (transducing) the various physical aspects of sound into a neural code for the physical characteristics of the complex sound field. This neural code is then further processed to provide neural subsets that aid us in determining the sound sources. Finally, this auditory information is combined with that from other sensory systems and from that provided by experience to elicit appropriate behaviors in response to the presence of sound sources. The entire sequence of coding, processing, integration, and responding to sound defines hearing as depicted in Figure 2.2.

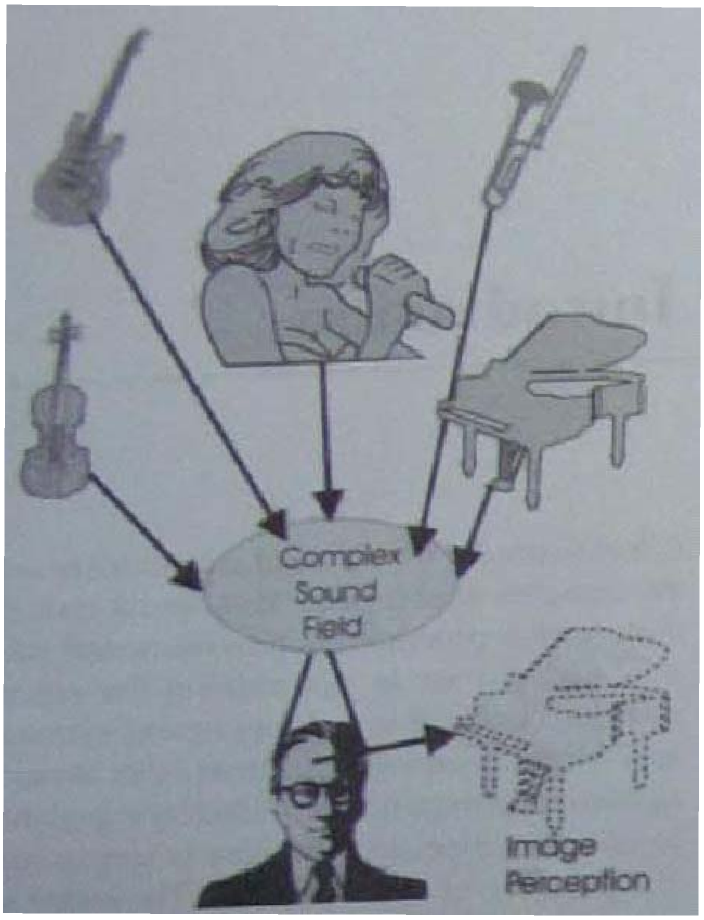


Figure 2.1 a schematics diagram indicating a number of objects (musical instruments) that could procedure sounds. The sounds from all these sources are combined into one complex sound field that is received by the listener. The auditory nervous system of the listener first provides a neural code of the basic physical attributes of the complex sound field, and this neural code is further processed to aid the listeners in determining the various sources. The listener perceives an auditory image of each sound source (e.g., the piano). (Yost, 2000)

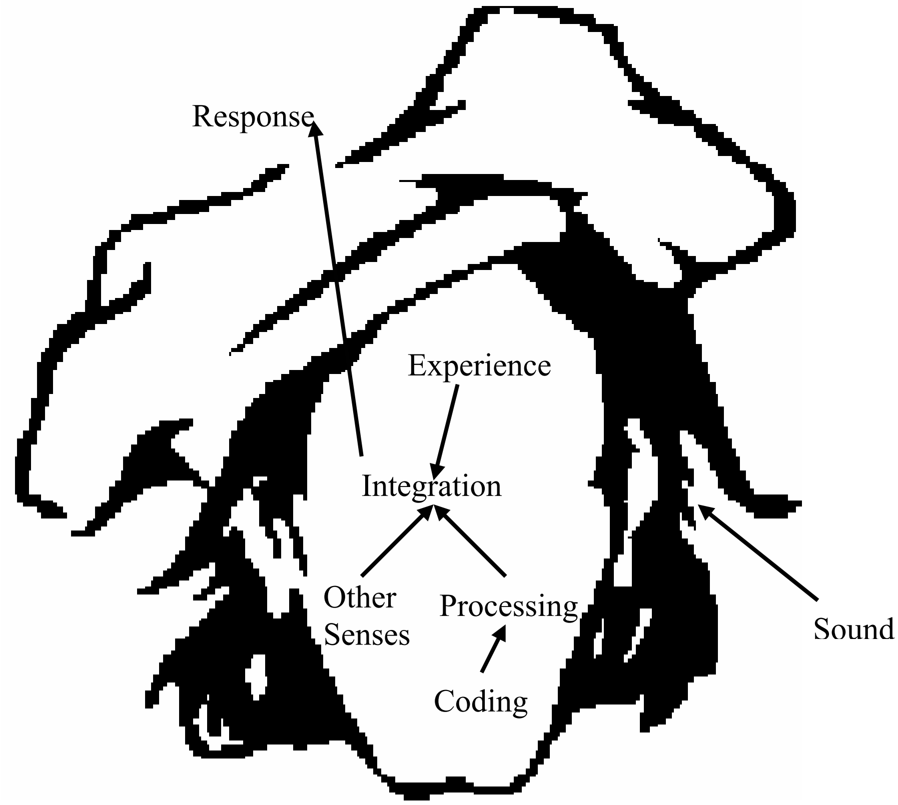


Figure 2.2 the stages of processing that lead to hearing. The physical attributes of sound (frequency, intensity, and time) are first coded by the peripheral auditory nervous system. This neural code is then processed by higher neural centers to help the listener determine the sources of sounds. This neural information is integrated with other sensory information and that based on experience, and all of this neural processing leads to behavioral responses.

2.2 Cocktail Party Problem

In 1953 Cherry wrote,

How do we recognize what one person is saying when others are speaking at the same time (the „cocktail party problem“)? On what logical bases could one design a machine („filter“) for carrying out such an operation? A few of the factors which give mental facility might be following:

- *The voice come from different directions*
- *Lip-reading, gestures, and the like.*
- *Different speaking voices, mean pitches, mean speeds, male and female, and so forth.*
- *Accents differing.*
- *Transition probabilities (subject matter, voice dynamics, and syntax).*

Figure 2.3 the cocktail party effect in real environment.

For many years, psychologists have wondered how we are able to follow the voice of one talker in the midst of a competition between different voices. An example to this situation would be a person at a cocktail party bombarded with babble coming from a multitude of people.

In 1953, Cherry appropriately dubbed this dilemma of separating one voice from another, the "cocktail party" problem. Similarly, infants must pay attention to one voice over others if they are to make sense of the babble around them. Just like our hypothetical individual at a cocktail party, infants must be able to reliably segment one speech stream, even in the face of other sounds and distractions. The question is how. [24]

The cocktail party effect can be analyzed as two related, but different, problems. The primary problem of interest has traditionally been that of *recognition*: how do humans separate speech sounds, and is it possible to build a machine to do this task. What cues in the signal are important for separating one voice from other conversations and background noise? Can, and should, a machine use the same cues for the task, or should it use other acoustical evidence that humans cannot detect?

The inverse problem is the *synthesis* of cues that can be used to enhance a listener's ability to separate one voice from another in an interactive speech system. In a user interface it may be desirable to present multiple digitized speech recordings simultaneously, providing browsing capabilities while circumventing the time bottleneck inherent in speech communication because of the serial nature of audio. Synthesis of perceptual cues by a machine for human listeners might allow an application to perceptually nudge the user, making it easier to attend to a particular voice, or suggest that a new voice come into focus.

In 1953, Cherry reported on objective experiments performed at MIT on the recognition of messages received by one and two ears. This appears to be the first technical work that directly addresses what the author called the ``cocktail party problem." Cherry proposed a few factors that may make the task of designing a ``filter," that could separate voices, easier: (see on the top of this topic).

All factors, except for the last, can be removed by recording two messages from the same talker on magnetic tape. The author stated that “the result is babble, but nevertheless, the messages may be separated.” In a Shannonesque analysis, Cherry suggested that humans have a vast memory of transition probabilities that make it easy for us to predict word sequences. [19]

2.3 Spatial hearing

When a sound is produced, it propagates from its source through the environment until it reaches the listener’s ears. The sound received at the ears is different from the sound produced by the source because it is modified by interactions within the listener’s body, head, and pinnae (Brungart and Rabinowitz, 1999; Shinn-Cunningham, Santarelli and Kopčo, 2000). In addition, if there are acoustically-reflective objects (for example walls) in the environment, acoustic reflections off these objects are received by the ears along with the “direct” sound. The basis of spatial hearing is in the listener’s auditory system extracting cues about the location of the sound source from the sounds received at the ears and the listeners using these cues to perform various tasks (localization, detection of sounds). [23]

2.4 Spatial Auditory Cues

There are many key attributes of the acoustic signals that reach the ears, including the time of arrival at the two ears, the level at the two ears, and the spectral profile. These attributes vary with the position of a sound source relative to a listener (Yost, 1994) and are used to locate and differentiate between multiple sound sources in their horizontal, or azimuthal, positions. The Interaural Level Difference (ILD; see Figure 1) occurs when a sound is presented on one side of a listener (Hartmann, 1999).

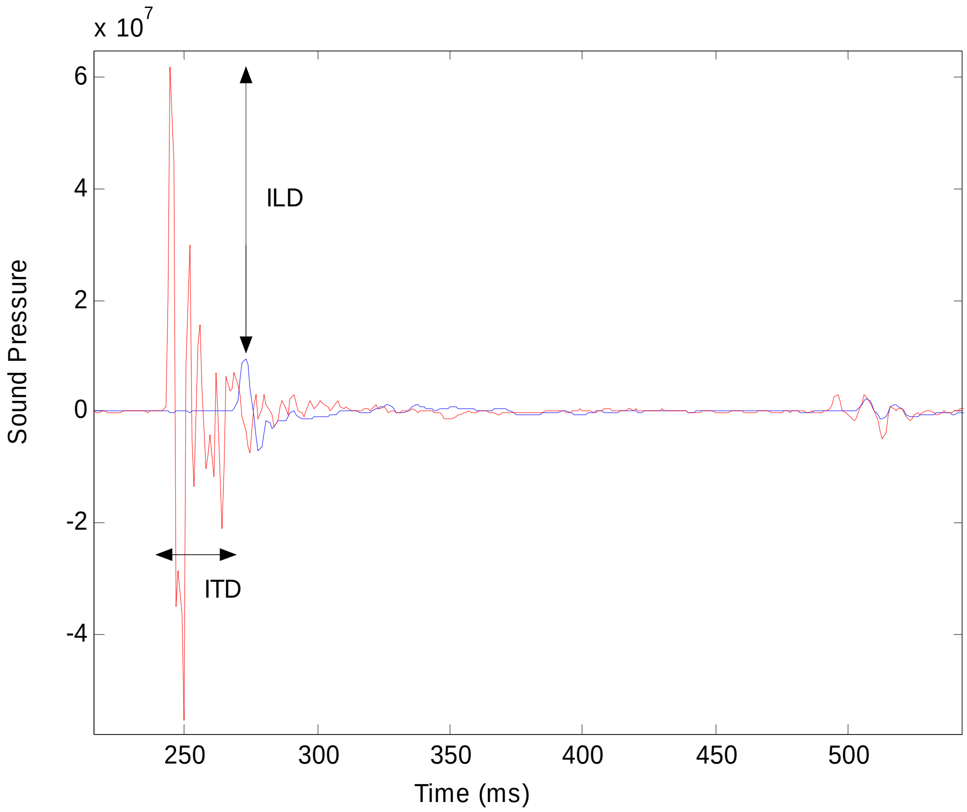


Figure 2.4: Interaural Level Differences (ILD) and Interaural Time Differences (ITD)

Figure 2.4 illustrates two pseudo-anechoic Head Related Impulse Responses (HRIRs; refer to pg. 9 for further explanation) for the right and left ear. The x-axis represents time (ms) and the y-axis represents sound pressure. The red waveform is the HRIR for the right ear and the blue waveform is the HRIR for the left ear. The waveforms show the pressure signals that reach the left and right ear canals when an impulse is played at an angle of 90^{o} on the azimuthal (horizontal) plane, 1 meter from the listener's head. The two impulse responses illustrate the concept of the interaural level difference (the difference in intensity of the sound signal between the left and right ears) and the interaural time difference (the difference in time it takes for a sound signal to reach the left and right ear). The ILD is shown by the difference in sound pressure (along the vertical axis) between the two waveforms. The ITD is shown by the difference in time (along the horizontal axis) between the waveforms.

The ILD value changes as the sound source moves to different spatial locations around the listener's head. There are two components to the ILD. The first component depends upon the relative distance from the sound source to each ear (Shinn-Cunningham et. al., 2000b). This ILD component increases as the sound moves along the interaural axis (the axis connecting listener's ears) toward the listener (see Figure 2.5) because intensity level is inversely proportional to the square of the distance from the sound source to the ear. This means that the closer the sound source is to the listener, the greater the intensity level will be, and (for lateral sources), the greater the difference in level at the two ears will be. This proportionality is especially important when the sound source is near the listener (< 1 m) because the ratio of the difference in distance between the left and right ears becomes much more significant. The ILD decreases when moving a sound source medially around the head from the interaural axis toward the median plane (the plane that divides the right and left side of the head [3]; see Figure 2.5). When the sound source reaches the median plane the difference in distance between the left and right ears is minimal, making the ILD value essentially zero [3]. The second component of the ILD is dependent on the frequency content of the signal. At very high frequencies (e.g., above 3 kHz), the effect of the "head shadow" can cause up to a 15 - 20 dB difference in the intensity reaching the different ears (e.g., Hartmann, 1999). However, for low frequencies (<1.5 kHz) the wavelength of the sound is larger than the head and the value of the ILD due to the head shadow becomes negligible (Middlebrooks and Green, 1991).

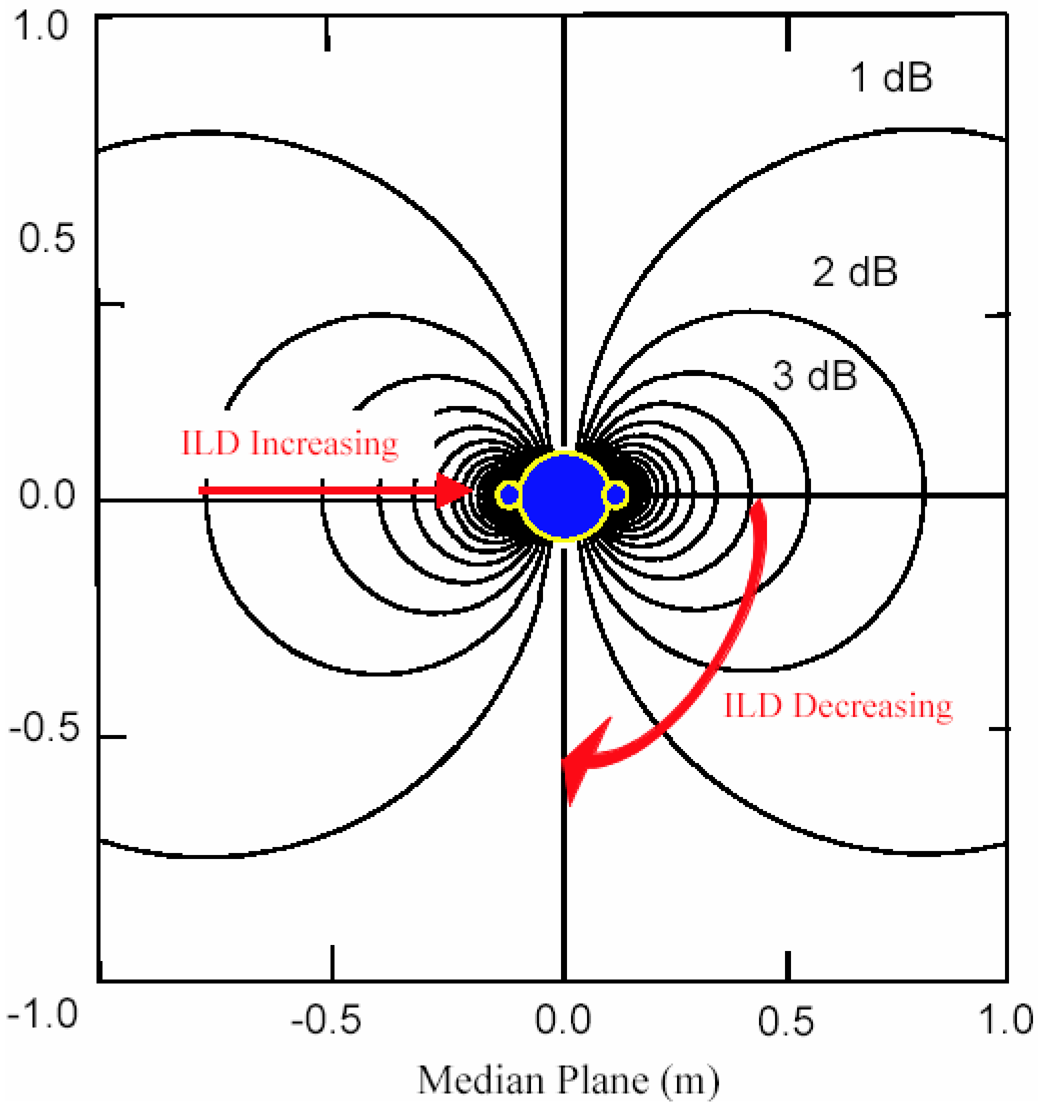


Figure 2.5: The Interaural Level Difference (ILD) on the Interaural Axis and Median Plane (Mraz, 1999)

This figure 2.5 illustrates how the distance component of the ILD changes when moving a sound source along the interaural axis or toward the median plane. This figure shows a top down view of the listener's head seen in blue. The contours show source positions from which the ILD would equal the values shown. ILD values grow in concentric circles (shown in steps of 1 dB) on each side of the listener. The ILD values are plotted as a function of the median plane (x-axis) and the interaural plane (y-axis). The ILD value equals 0 dB when the source is on the median plane, equidistant from the 2 ears. The ILD value increases when moving along the interaural axis toward the listener. The ILD value decreases when moving medially from the interaural axis toward the median plane. The arrows in the figure 2.5 illustrate these concepts.

Even for low-frequency signals that have a negligible ILD, the brain can still determine source laterality, confirming the presence of another localization cue, the interaural time difference (ITD; see Figure 2.4). The ITD is simply the time delay between the sound signals at the two ears. An ITD occurs because of the difference in travel time from the sound signal to the left and right ears. ITDs are important for determining sound location relative to the head. The ITD is maximal for sources along the interaural axis and decreases as the source approaches the median plane (Iannone, 2000).

The auditory system does not compute ITD directly, instead it uses the third spatial cue of interest, the interaural phase difference (IPD), in narrow bands of frequency (Shinn-Cunningham et. al, 2000b). The IPD is cyclical, varying from $-\pi$ to π, as a function of ITD (see Figure 2.6). If the ITD is constant, the IPD will change as a function of frequency according to the following equation:

IPD (cycles) = ITD (s) / $2\pi f$ (s/cycle)

(Where f is the frequency).

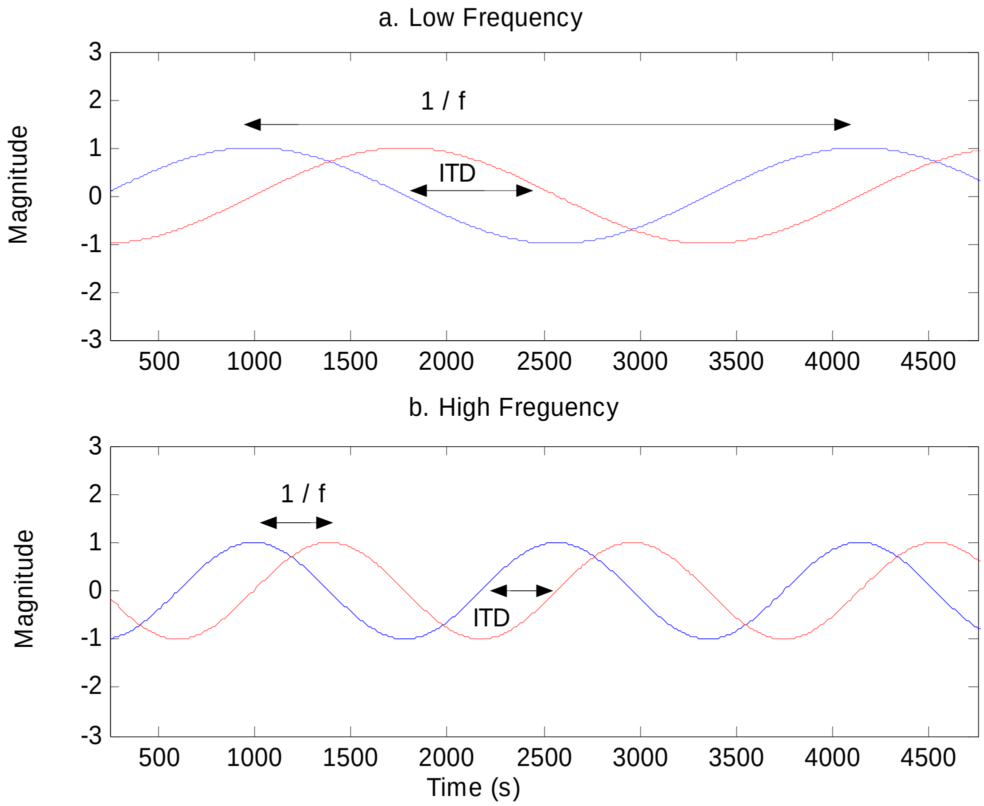


Figure 2.6: IPD vs. ITD

The figures above illustrate the difference between IPD and ITD. The blue waveforms represent the magnitude of sound signals for the left ear and the red waveforms represent the magnitude of sound signals for the right ear (y-axis) plotted as a function of time (x-axis). Part a shows a low frequency sinusoid and part b a high frequency sinusoid. In both instances, the ITD of the signals is about the same (shown by the arrows). However, the difference in frequency makes the same ITD result in different IPD values.

More than one sound location can result in the same ITD. This ambiguity causes “cones of confusion” (see Figure 2.7). All points on the same cone lead to essentially the same interaural time differences. Subjects make behavioral localization errors due to this phenomenon, particularly along the vertical and front/back dimensions (Middlebrooks and Green, 1991). For example, a listener may perceive a sound stimulus towards the rear even when it was played from the front, or vice versa. In other words, listeners can generally tell how far left versus right a source is located from the ITD, but they have more difficulty determining the source location in front of versus back or up versus down. [3]

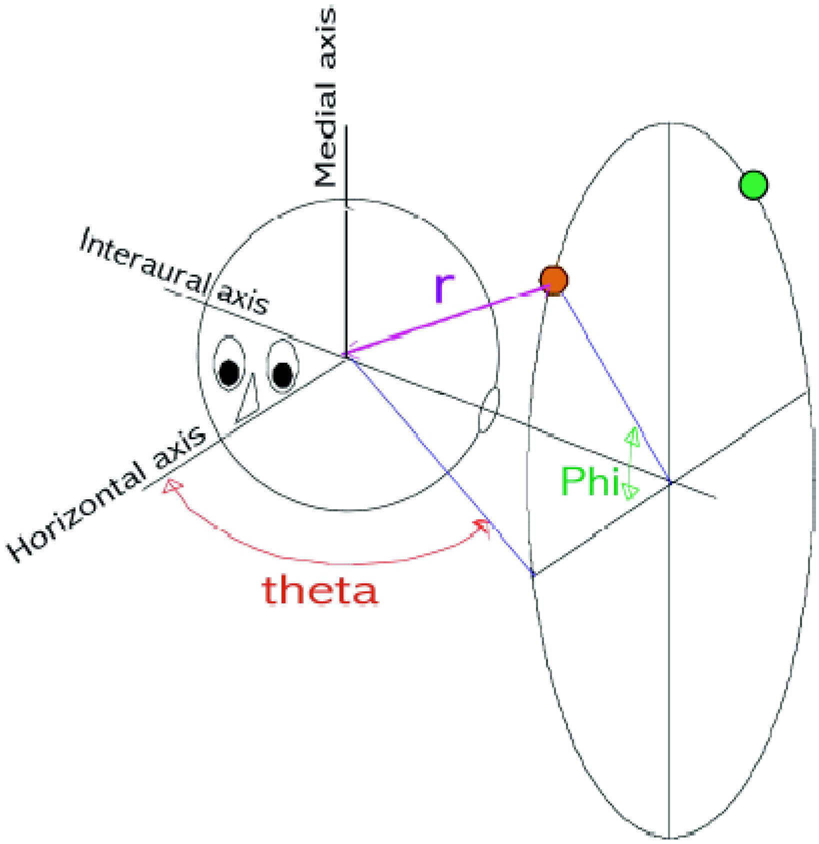


Figure 2.7: "Cones of Confusion" (Mraz, 1999)

This figure 2.7 shows an example of a "cone of confusion" in three dimensions. The large circle on the side of the listener's head around the interaural axis and parallel to the median plane represents all points at the same distance from the head and on the same "cone of confusion." The orange sphere represents a sound source lying on this "cone of confusion". This location is defined by the coordinates of space (r, θ, Φ), where r is the radial distance of the source from the head, phi (Φ) is the angle from the horizontal plane, and theta (θ; which defines the "cone of confusion") is the angle between the median plane and the line defined by the source and the center of the head. The green sphere shows where a listener may typically localize the original sound source due to the phenomenon caused by the "cones of confusion. [9]

2.5 Head Related Transfer Functions

Head Related Transfer Functions (HRTFs) describe how a sound signal is transformed as it moves through space and finally reaches the eardrum. For each ear, every position in space (r, θ, ϕ) is associated with a different HRTF. Also, HRTFs are specific to an individual because of each subject's unique physical attributes. HRTFs are measured using the following equations:

Figure 2.8 Scheme of making HRTF for L and R ear.

$$\mathbf{HRTF_{left}(\omega,r,\theta,\phi) = O_{left}(\omega) / I_{r,\theta,\phi}(\omega)}$$

$$\mathbf{HRTF_{right}(\omega,r,\theta,\phi) = O_{right}(\omega) / I_{r,\theta,\phi}(\omega)}$$

(where $O_{left}(\omega)$ and $O_{right}(\omega)$ are the sound signals at the left and right ears, respectively, $I_{r,\theta,\phi}(\omega)$ is the sound signal emitted by the sound source, and ω is frequency). HRTFs vary with source distance and direction. When represented in the time domain, HRTFs are referred to as Head Related Impulse Responses (HRIRs; see Figure 2.4). Essentially, HRTFs are filters that contain the ILD, ITD, IPD, and spectral cues that would arise if a source were at a particular location relative to the listener. Once measured, HRTFs can be used to accurately recreate over headphones what the listener would hear from a sound at a certain position in space. This creates a virtual auditory environment for the listener. It also allows the experimenter to control the signals reaching the listener during testing. [9]

The transformation of a sound from the source to the ear is constant for a fixed sound source and listener position. The sound source, the environment in which the sound propagates (including the listener, and all objects and walls in the environment), and the ear create a linear system that transforms the input signal (sound produced by the source) into an output signal (the sound received at the ear). This system can be mathematically characterized by its impulse response called the Head-Related Transfer Function (HRTF). The HRTF describes the signal reaching the ear when a broadband impulse is played from a specific source location. This impulse response is sufficient to predict how any sound coming from a specific location is altered as it travels to and impinges on the ear. Because the sound has to travel through a different path to each of the two ears, a pair of HRTFs (the left-ear and the right-ear HRTF) provides complete information about how any sound is received at the ears when originating from a given location.

There are two main applications of HRTFs in hearing research. First, HRTFs can be used to generate virtual auditory environment. That is, by convolving a sound with the HRTF one can simulate how the sound would be received at the listener’s ears if it was presented from any location around the listener in any environment. Second, HRTFs can be analyzed to determine what spatial auditory cues are available to the listener when a sound is presented from a specific location in a specific room. [3]

Individualized HRTF measurements used in experimental procedure were made in the Auditory Neuroscience Lab of the Boston University CNS Dept (measured by Kopco and Persico, 2002). The subjects seated in the center of a quiet classroom (rough dimensions of 5x9x3.5 m; broadband T_{60} of approximately 700 ms). Subjects were seated with their heads in a headrest so that their ears were approximately 1.5-m above the floor. Measurements were taken for sources in the right front horizontal plane (at ear height) for all azimuths between -90° to 90° by step of 5°. The distances relative to the center of the head were 120 cm.

The Maximum-Length-Sequence (MLS) technique (e.g., see Vanderkooy, 1994) was used to measure HRTFs. Two identical 32 767-long maximum length sequences were concatenated and presented through a small loudspeaker using a 44.1-kHz sampling rate (details regarding the equipment are described below). The response to the second sequence was recorded. This measurement was repeated ten times and the raw measurements averaged in the time domain. This average response was then used to estimate a 743-ms-long head related impulse response.

HRTFs were measured using a Tucker-Davis Technologies (TDT) signal processing system under computer control. For each measurement, the concatenated MLS sequence was read from a PC hard-drive and sent to a TDT D/A converter (TDT

PD1), which drove a Crown amplifier connected to a BOSE mini-cube loudspeaker. At the start of the measurement session, the subject was positioned so that the center of his/her head was at a location marked on the floor of the room. The subject's head position was read from a Polhemus FastTrak electromagnetic tracker worn on the head to ensure that the center of the head was within 1-cm of the correct location in the room, marked on the floor. The experimenter used other angular and distance markings on the floor to hand-position the loudspeaker to the appropriate azimuth and distance prior to each measurement. Miniature microphones (Knowles FG-3329c) mounted in earplugs and inserted into the entrance of the subjects' ear canals (to produce blocked-meats HRTF recordings) measured the raw acoustic responses to the MLS sequence. Microphone outputs drove a custom-built microphone amplifier that was connected to a TDT A/D converter (TDT PD1). These raw results were stored in digital form on the computer hard-drive for off-line processing to produce the estimated HRTFs.

HRTFs measured as described above include room echoes and reverberation. To eliminate room effects, time-domain impulse responses were multiplied by a 6-ms-long $\cos_2$ time window (rise/fall time of 1 ms) to exclude all of the reverberant energy while retaining all of the direct-sound energy. The resulting "pseudo-anechoic" HRTFs were used to simulate sources (and in all subsequent analysis). HRTFs were measured only for sources in the right hemi field. To simulate sources in the left hemi field, HRTFs from the corresponding right-hemi field positron were used, exchanging the left and right channels (i.e., left/right symmetry was assumed; given that only pure tone targets were simulated in the left hemi field, this approximation should introduce no significant perceptual artifacts in the simulated stimuli).[4]

By the preparing of Matlab codes for the experimental procedure we make a mistake and changed L and R HRTFs; however it shouldn't have an effect on experimental results. It therefore because the subjects don't hear sounds through their ear but through „foreign ear" (it means by helping of HRTFs). [3]

2.6 Spatial Unmasking

When listening for a target auditory signal in the presence of another simultaneous signal (a masker), a listeners' ability to perceive the target is influenced by the target and masker locations. In general it is easier to detect or recognize the target when it is spatially separated from the masking sound compared to the condition when the two sources are located at the same position [13, 14, and 15]. Three factors contribute to this *spatial unmasking* effect. First, the acoustic signal-to-noise ratio (SNR) at either ear changes with target or the masker location due to both head shadow effects and distance effects. Spatial separation of target and masker can either increase or decrease the SNR at a given ear, depending on the spatial locations of target and masker.

In addition to simple energetic effects due to changes in SNR at the ears, changes in source location lead to changes in the binaural cues due to that source. The auditory system can detect the presence of the target due to changes in the binaural cues in the target plus masker stimulus compared to the binaural cues in the masker alone. In general, the target influences IPD cues in the target plus masker most when the IPDs in the target and masker are most different; thus, target detection is easiest when target and masker IPDs differ by π. Similarly, detection of an in-phase target masked by an in-phase noise is easiest if the ILD of the masker is 0 and ILD of the target is ∞ (Durlach and Colburn, 1978).

Finally, informational masking can be influenced by the perceived spatial locations of target and masker. While there is no "standard" definition of informational masking, it is used to refer to influences that cannot be ascribed to simple acoustical parameters of the sounds reaching the two ears (e.g., attentional effects, cross-modality influences, etc.). Informational factors have been shown to play an especially important role for tasks involving high levels of uncertainty, e.g., when complex sounds are masked by complex sounds [15] or when speech is masked by speech [16].

For pure-tone targets, the role of informational masking is thought to be negligible; at threshold, pure-tone detection is determined by the subject's ability to detect subtle changes in the masker due to the presence of the target, not by the ability to "hear out" the target as a separate auditory event. In general, even for signal levels above threshold where the target is perceived as a separate object, its perceived location is strongly biased by the location of the masking noise.

Free-field masking of chirp-train targets has also been studied [1, 18], leading to spatial unmasking of up to 20 dB (similar to the results for pure-tone targets).

As previously noted, the signal levels and phases at the ears depend on source location. Similarly, the SNR at the ears depends on the spatial configurations of the target and noise sources. Different frequency bands have different target IPDs, and thus, the signal will be easier to hear in different frequency bands. In particular, if the signal and the masker IPD are 180^{o} out of phase in a particular frequency band, the increase in signal audibility is greatest. If the signal and masker IPD are identical, there is no benefit. Changes in the SNR at the ears and the IPD of the target and masker influence the intelligibility of a speech target in a noisy background and determine the amount of spatial unmasking [3]. When target and masker are in the same location, the SNR at the two ears is equal. However, if the target moves away from the masker, the SNR generally will improve at the ear towards which the target moves. As a result, the target is easier to understand. In addition to this benefit, when the target is displaced from the masker in a way that the ITD of the target and masker differ, there is an increase in intelligibility above and beyond the benefit due to the change in SNR at the "better" ear. This benefit can be predicted by analyzing the IPD of the target and masker in each frequency band, assuming that the increase in audibility in each band depends on the IPD difference (Zurek, 1993).

A number of recent studies of spatial unmasking used HRTFs to simulate sources (targets and maskers) at different directions over headphones. Results of these studies show that spatial unmasking arises from both these monaural (SNR) and binaural benefits (Drullman and Bronkhorst, 2000; Carlile and Wardman, 1995). Specifically, by moving the target and masker away from each other SNR can improve dramatically, causing changes in target threshold of as much as 15 dB. However, even after thanking this into account, thresholds can sometimes be lower than expected (by as much as 2 to 4 dB) due to differences in ITD (and IPD; Drullman and Bronkhorst, 2000). [23]

2.7 The “Better” Ear

The better ear is the ear with the higher SNR. A listener will use the signal from this ear to do most of the monaural signal processing. Conversely the worse ear is the ear with the lower SNR. This ear, however, is extremely important, because it establishes the localization cues.

2.8 Binaural masking

In the preceding section 2.3 we described the auditory system’s sensitivity to changes in interaural time and level, which are principally used to locate sound sources in the azimuth plane. Many experiments have shown that the threshold for detecting a signal masked by noise is lower when the noise and signal are presented in a particular way to both ears. In these experiment subjects first determined their masked thresholds when both the noise and tonal signal were presented equally to both ears. In one test experiment, the tonal signal was removed from one ear, such that the noise was at both ears and the signal at only one ear. In this case, the signal was easy to detect, and therefore the level of the tone had to be reduced to obtain masked thresholds. Subsequently, many investigators have studied the improvement in detection associated with presenting stimuli to both ears. A certain nomenclature has been developed to describe the various types of binaural configurations of signal and noise.

monotic: stimuli presented to only one ear

diotic: identical stimuli presented to both ears

dichotic: different stimuli presented to the two ears

Investigators have found that the masked threshold of a signal is the same when the stimuli are presented in a monotic or diotic condition. If the masker and signal are arranged in a dichotic situation, however, the signal has a lower threshold than in either the monotic or diotic conditions. There are several ways to present the S and M in a dichotic or diotic manner; again, a set of symbols is used to describe these stimulus conditions:

$\mathbf{S_0}$**:** signal presented binaurally with no interaural differences (diotic)

$\mathbf{M_0}$**:** masker presented binaurally with no interaural differences (diotic)

$\mathbf{S_m}$: signal presented to only one ear

$\mathbf{M_m}$: masker presented to only one ear

S_π: signal presented to one ear 180^o out of phase with the signal presented to the other ear

M_π: masker presented to one ear 180^o out of phase with the signal presented to the other ear

For the binaural conditions described above, the signal or masker is identical in all dimensions except that denoted by a subscript. Thus,

monotic: M_mS_m
diotic: M_0S_0
dichotic: M_0S_π, M_0S_m, $M_\pi S_\pi$, $M_\pi S_0$, $M_\pi S_m$

To compare detection on one binaural condition with in another, the data are usually presented as the difference between the signals levels required for detection (masked threshold) in a monotic condition. That is, the signal level required for detection in the relevant diotic or dichotic condition is subtracted from the signal level required for detection in the M_mS_m (monotic) conditions. Such a difference when expressed in decibels is called a *masking level difference* (MLD) or a *binaural masking-level difference* (BMLD).

Table 2.1 shows the type of improvement in detection provided by dichotic presentation of maskers and signals (MLD). These data represent approximately the maximum MLD obtain when the masker is a continuous, broadband white Gaussian noise and the signal is pulsed sinusoid of low frequency (below 1000 Hz) and long duration (grater than 100msec). [5]

Interaural condition Compared to MmSm	MLD (dB)
M_mS_m, M_0S_0, $M_\pi S_\pi$	0
$M_\pi S_m$	6
M_0S_m	9
$M_\pi S_0$	13
M_0S_π	15

Table 2.1: The MLD in dB for a Variety of Stimulus Conditions (Yost, 2000)

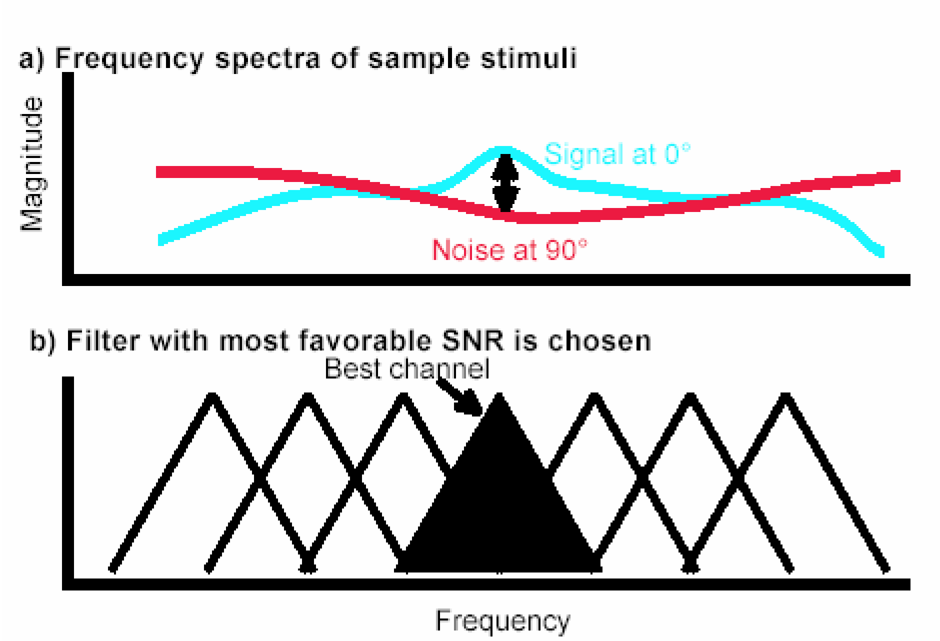
3. Models

In this chapter are described two models that can generate predictions of the experimental data. Here there are introduced by presenting basics schemas and main principles of their functioning.

3.1 Single-best filter model

S-B-F is a simple model which consists of a bank of 60 log-spaced GammaTone filters [2] for each ear. This model separately computes for each spatial configuration, the root-mean-squared energy at the output of every filter for the signal and noise. The SBF envisage about two filter widths: standard and narrow (scaling of 2,5). Shera et al. suggested these filter widths (2002). The model assumes that the filter with the largest SNR (over the set of 120) determines threshold, therefore is SNR computed for each filer. Predicated threshold is computed by -SNR – T_0 (T_0 is a model parameter)

The only free parameter in the model, the SNR yielding 79.4% correct performance, was fit to match the measured threshold when signal and noise were at the same location.



3.1 Schematic of the single-best-filter model. (Kopčo, 2003)

3.1.1 GammaTone

A gamma-tone is the product of a rising polynomial, a decaying exponential function, and a cosine wave. It can be described with the following formula:

$$gammaTone(t) = a * t^{\gamma-1} * e^{-2\pi * bandwidth * t} * \cos(2\pi * frequency * t + initialPhase)$$

where π determines the order of the gamma-tone.

The GammaTone function has a monotone carrier (the tone) with an envelope that is a gamma distribution function. The amplitude spectrum is essentially symmetric on a linear frequency scale. This function is used in some time-domain auditory models to simulate the spectral analysis performed by the basilar membrane. It was popularized in auditory modeling by [2]

This GammaTone filter has two free parameters the *Centre frequency*, and Bandwidth witch determine the passband of the filter.

3.2 Dau's model

In Dau (1992), Dau and Püschel (1993) and Dau et al. [21, 22] a model was proposed to describe the „effective“ signal processing in the auditory system. This model allows the prediction of masked thresholds in a variety of simultaneous and non-simultaneous conditions. The model was initially designed to describe temporal aspects of masking. There is no restriction as to the duration, spectral composition and statistical properties of the masker and the signal. The model combines several stages of preprocessing with a decision device that has the properties of an optimal detector. Figure 3.2 shows how the different processing stages in the auditory system are realized in the model. The frequency-place transformation on the basilar membrane is simulated by a linear basilar-membrane model (Schroeder, 1973; Strube, 1985). Only the channel tuned to the signal frequency is further examined. As long as broadband noise maskers are used, the use of off frequency information is not advantageous for the subjects. The signal at the output of the specific basilar-membrane segment is half-wave rectified and lowpass filtered at 1 kHz. This stage roughly simulates the transformation of the mechanical oscillations of the basilar membrane into receptor potentials in the inner hair cells. The lowpass filtering essentially preserves the envelope of the signal for high carrier frequencies.

Effects of adaptation are simulated by feedback loops (Püschel, 1988; Kohlrausch et al., 1992). The model tries to incorporate the adaptive properties of the auditory periphery. It was initially developed to describe forward masking data. Adaptation refers to dynamic changes in the transfer gain of a system in response to changes in the input level. The adaptation stage consists of a chain of five feedback loops in series, with different time constants. Within each single element, the lowpass filtered output is fed back to form the denominator of the dividing element. The divisor is the momentary charging state of the lowpass filter, determining the attenuation applied to the input. The time constants range from 5 to 500 ms. In a stationary condition, the output of each element is equal to the square root of the input. Due to the combination of five elements the stationary transformation has a compression characteristic which is close to the logarithm of the input. Fast actuations of the input are transformed more linearly. In the stage following the feedback loops, the signal is lowpass filtered with a time constant of 20 ms, corresponding to a cutoff frequency of nearly 8 Hz to account for effects of temporal integration. To model the limits of resolution an internal noise with a constant variance is added to the output of the preprocessing stages. The transformed signal after the addition of noise is called the internal representation of the signal. The auditory signal processing stages are followed by an optimal detector whose performance is limited by the nonlinear processing and the internal noise. [22]

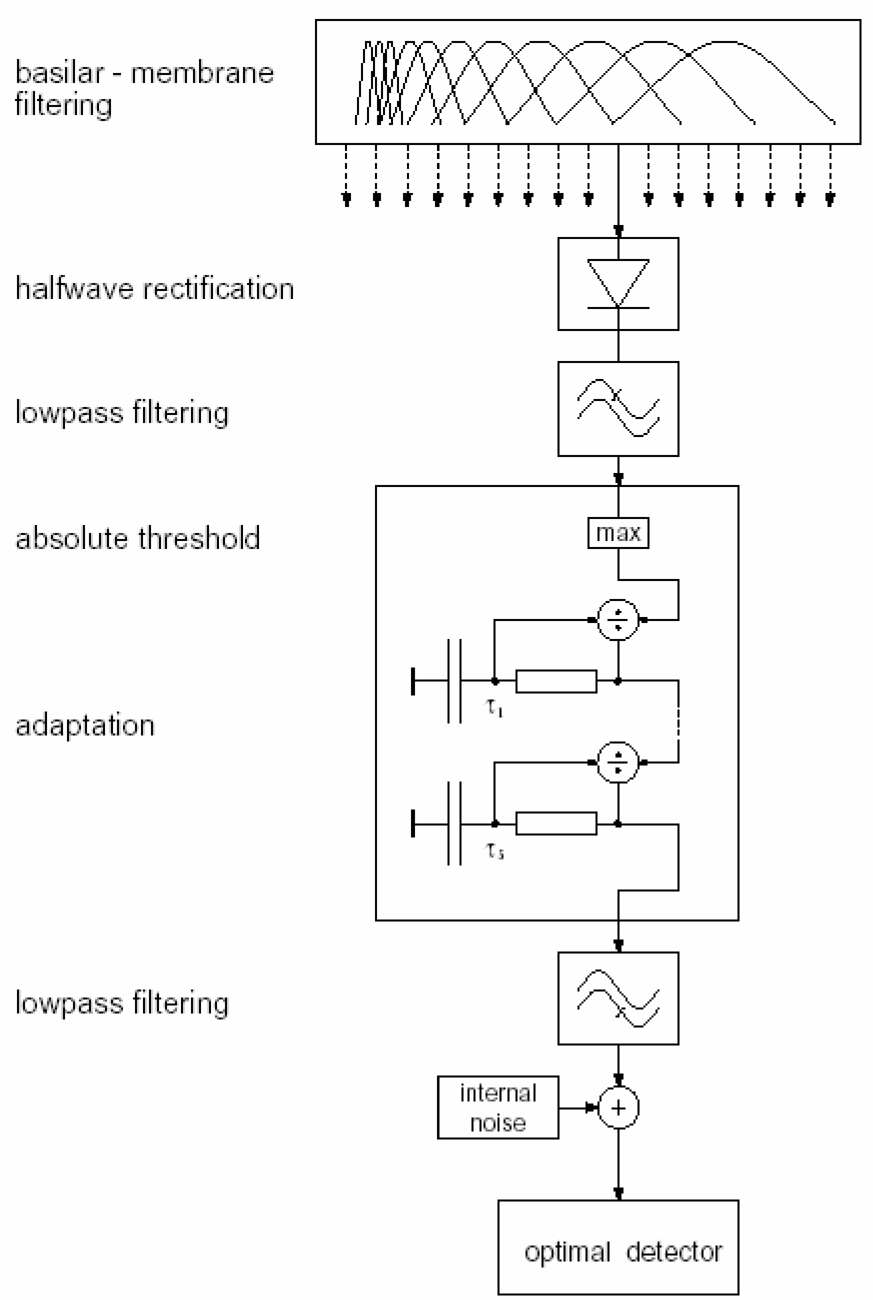


Figure 3.2: Block diagram of the psychoacoustical model for describing simultaneous and nonsimultaneous masking data with an optimal detector as decision device (Dau, 1992; Dau et al., 1996). The signals are preprocessed, fed through nonlinear adaptation circuits, lowpass filtered and finally added to internal noise; this processing transforms the signals into their internal representations. (Dau, 1996)

3.2.1 Description of the model

The main idea of the optimal detector is that a change in a test stimulus is just detectable if the corresponding change in the internal representation of that test stimulus - compared with an internally stored reference - is large enough to emerge significantly from the internal noise. In the decision process, a stored temporal representation of the signal to be detected (the template) is compared with the actual activity pattern evoked on a given trial. The comparison amounts to calculating the cross correlation between the two temporal patterns and is comparable to a „matched filtering“ process. The detector itself derives the template at the beginning of each simulated threshold measurement from a suprathreshold value of the stimulus. If signals are presented using the same type of adaptive procedure as in corresponding psychoacoustical measurements, the model could be considered as „imitating“ a human observer. The optimality of the detection process refers to the best possible theoretical performance in detecting signals under specific conditions. The calibration of the model is based on the 1 dB criterion in intensity discrimination tasks. In the first step of adjusting the model parameters, this value of a just-noticeable change in level of 1 dB was used to determine the variance of the internal noise.

In the model described above, the stimulus - in its representation after the adaptation stage - is filtered with a time constant of 20 ms. This stage represents the „hard-wired“ integrative properties of the model and leads - in combination with preprocessing and the decision device - to very good agreement between experimental and simulated masked-threshold data. However, for describing modulation detection data it is not reasonable to limit the availability of information about fast temporal actuations of the envelope in that way. In addition, as pointed out in the Introduction, results from several studies concerning modulation masking indicate that there is some degree of frequency selectivity for modulation frequency. It is assumed here that the auditory system realizes some kind of spectral decomposition of the temporal envelope of the signals. For this reason, the following model structure is proposed to describe data on modulation perception. [22]

4. Experiment: Motivation and hypotheses

„Spatial unmasking“ is an improvement in signal detection threshold when signal and noise are spatially separated. It is well known from previous studies that spatial unmasking of pure-tone stimuli depends on: energetic and binaural factors. The energetic factors refer to the fact that changes in spatial location of the stimuli cause a change in the signal-to-noise energy ratio with which the stimuli are received at the ears. The binaural factors refer to the improvement in signal detectability caused by differences in the interaural cues (ITD and ILD) of the signal relative to the noise. Gilkey and Good [7] showed that spatial unmasking of broadband stimuli depends on: energetic factors for all stimuli and additional binaural factors for lowfrequency stimuli. So when we manipulate with BRD stimuli we have two possibilities of showing the how the auditory system integrates information across multiple channels and how the auditory system chooses single best channel with most favorable SNR.

Kopco et al. [10, 11] found three characteristics that are typical for spatial unmasking of broadband stimuli in virtual anechoic environment. Here, we would like to re-evaluate these three characteristics by testing them in a more direct way. The characteristics are:
1) spatial unmasking of broadband stimuli is determined by the spectral content of the received stimuli at high frequencies.
2) binaural contribution to spatial unmasking is very small, less than 3-4 dB.
3) performance can be explained completely by considering the SNR in the single best peripheral auditory filter and that the best filter is always centered at high frequencies.

Kopco et al. draw the above conclusions without testing them directly, because their subjects always listened to broadband stimuli. The goal of the present study is to evaluate the Single-Best-Filter model behaviorally, by comparing thresholds obtained with broadband stimuli to thresholds for stimuli that contain energy in the single best filter.

Specifically, in the present study we first analyze the stimuli as they reach the listener’s ears from various locations and as they get processed by the auditory periphery (modeled by a bank of GammaTone filters). We compute the signal-to-noise ratios

(SNR) in all peripheral channels for all configurations of the target signal position and of a masking noise position. The goal of this analysis is to determine spatial configurations of the target and masker that test the Single-Best Filter model in specific ways:

1A: one channel dominates (this channel is always HF) – this is a condition where SBF model is expected work fine

1B: a configuration, where one channel dominates but the channel is at low frequency (around 1 kHz frequency) – here the SBF model might not work because binaural processing might contribute

2. at least two equivalent HF channels exist (here we can include also situation, when S and N are in the same position) – if the system is able to integrate information, the broadband threshold should be better than either of the monaural thresholds

3. one HF channel dominates, but there is also one LF channel, that is a little worse. Because auditory perception at low frequencies is improved by binaural processing, it's hard to predict which channel will be better and thus, which channel will determine performance and whether there will be any across-channel interaction.

4. condition with many similarly good channels – this is the most direct test of the SBF model

First of all we would like to show like in their experiments how it is in unmasking on HF by using our experimental data. Second how it is about binaural effect –if it is very small like in Kopco et al. [10, 11] or the things could be another. And the last point of their hypothesis if all frequencies can be predicted by Single-Best-Filter.

To test these hypotheses, an experiment was prepared and data were collected on several listeners, as is described in the following sections.

5. Experimental Methods

5.1 HRTF analysis

Similar to Kopčo [10] we try to confirm the Single-best-filter model hypothesis in the human auditory system.

However, the present approach is a little bit changed because here the hypothesis is tested by a very different approach. While Kopčo et al. [10, 11] tested the hypothesis by measuring only broadband thresholds and comparing them to the predictions of the single-best-filter model, here the thresholds for stimuli were filtered by the model of single best filter and measured as well.

In this part of the work we want to filter the convolutions of signal through the GammaTone filters, and we should get 60 signals for each ear at intervals of broadband between 300 and 12000 Hz. Signal is presented by 200ms 40-Hz chirp train and Noises 250ms white noise bursts. We make it for all signal azimuth position of -90° to 90° by step of 5° (together 37 position) and for noise azimuth position (-90°,-60°,-45°,-30°, 0°, 30°, 45°, 60°, 90°).

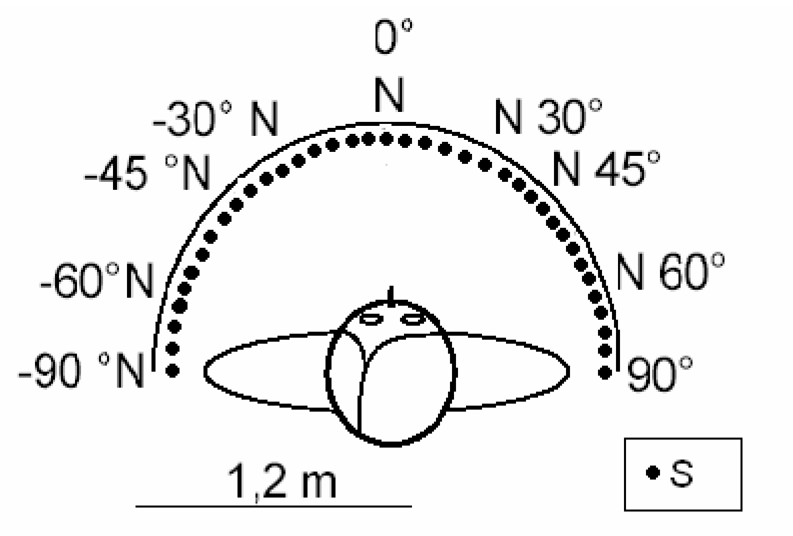


Figure 5.1 Simulated positions of S and N in making HRTFs analysis

When the listener is in a reverberant environment (e.g., in a room) the direct sound received at the ears is combined with multiple copies of the sound reflected off the walls before arriving at the ears. This reverberation acts like noise that deteriorates the spatial cues extracted by the auditory system.

Basically the reverberations arrive later to ours ears than straight signal, so generally we can get HRTFs without of wall reverberations by cutting off first impulse of these functions.

In that case if the simulation is measured in the anechoic room, the HRTFs will regulate this way: we separate reverberant signals from room walls and another barriers, and then we will work only with pure chirps, which will reach to pinna.

These separations should be the same for L and R ear.

To eliminate room effects, time-domain impulse responses were multiplied by 450-ms-long cos-squared time window (rise/fall time of 20 see fig. 5.2) to exclude all of the reverberant energy while retaining all of the direct-sound energy.

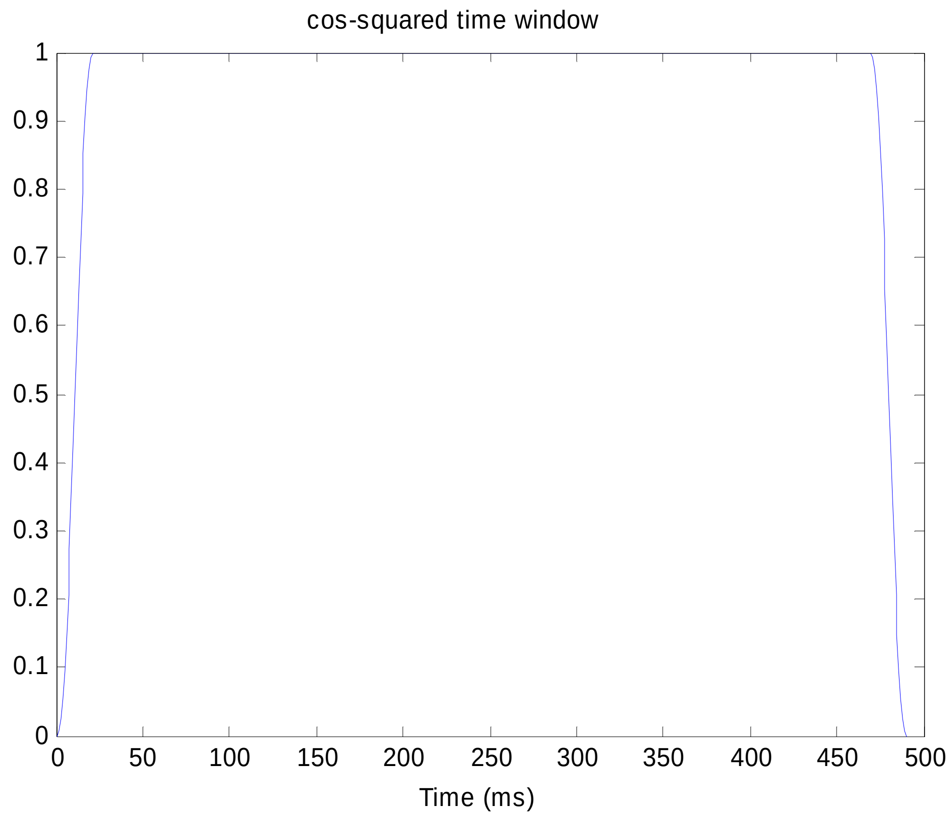


Figure 5.2 Shows cos-squared time window to eliminate room effects.

The resulting pseudo-anechoic HRTFs were used to simulate sources (and in all subsequent experiments and analyses).

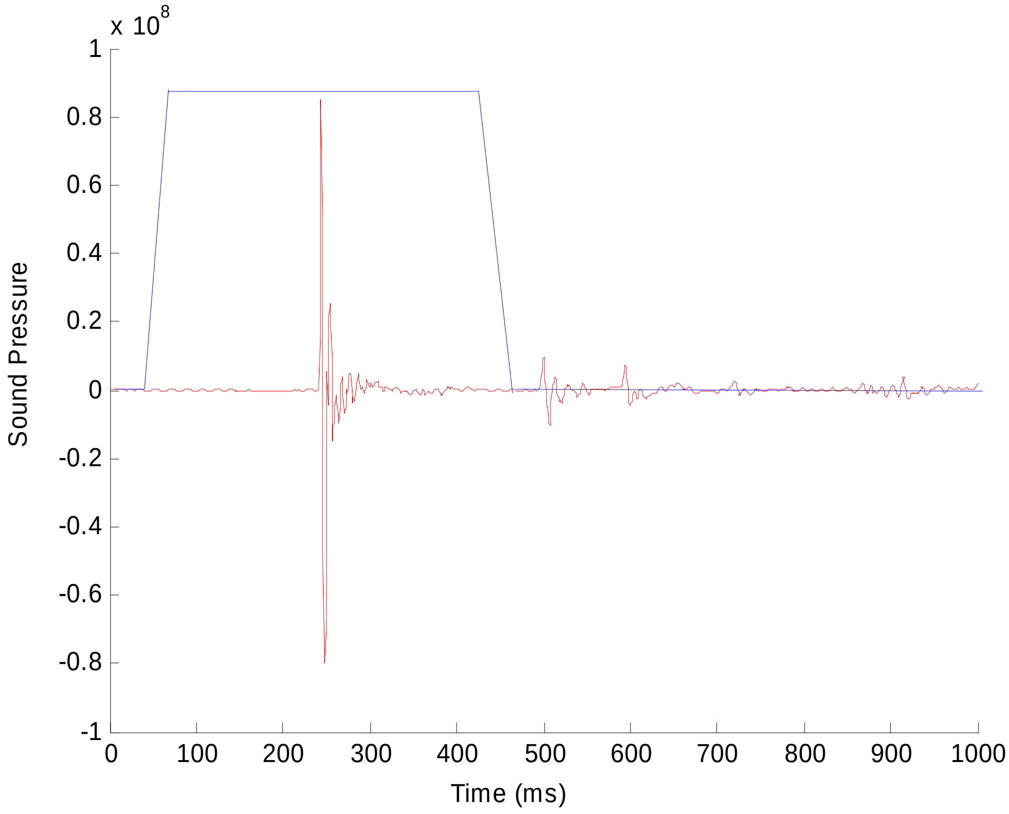


Figure 5.3: Time Window for Pseudo-Anechoic Head Related Transfer Function (HRTF) Creation

This figure shows a 490 ms-long portion at the start of a reverberant Head Related Impulse Response (HRIR; seen in red) plotted as a function of time (x-axis) versus sound pressure (y-axis). An example of a time window used to separate the direct sound from the reverberant information in the HRIR is shown in blue. The direct sound alone creates a pseudo-anechoic version of the reverberant HRTF.

Because the simulations is in progress in anechoic room, we had to convolute HRTFs (chirps) with broadband signals and broadband white bursts noises at all azimuth positions before.

Because the procedure was reverse how it used in previous study of Kopco at al. (2003), we want to confirm the method of using the single-best-filter. All used broadband stimuli (signals and noises) were convoluted with HRTFs and filtered by filterbank of 60 log-spaced GammaTone filters. So we got 60 short signals for each ear by using 60 frequency bands at intervals of 300 and 12000 Hz.

The best channel, of which threshold energy is the best, actives in auditory organ according to single-best-filter. For finding out which channel is the best, it is necessary to count its RMS energy for S and N according to:

$$RMS(dB) = 20 * \log_{10}\left(\sqrt{\sum\left(filtered\_signal\right)^2}\right)$$

When we want exactly predict azimuth combination from HRTFs analyze for experiment. So we must count from RMS the activated energy called SNR.

$$SNR(dB) = (chirp_{RMS}(dB) - noise_{RMS}(dB))$$

On base of these SNRs we chose such combinations by describing all factors of experiment, which are needed on spatial unmasking. By helping of HRTFs analyzes the experimental procedure can confirm the use of single-best-filter. Except that, we will look in detail for binaural benefit in unmasking broadband stimuli to the 2000 Hz.

When we have suspicion of binaurally benefit, we have to measure thresholds monaurally and also binaurally. We always have to measure moreover broadband signals.

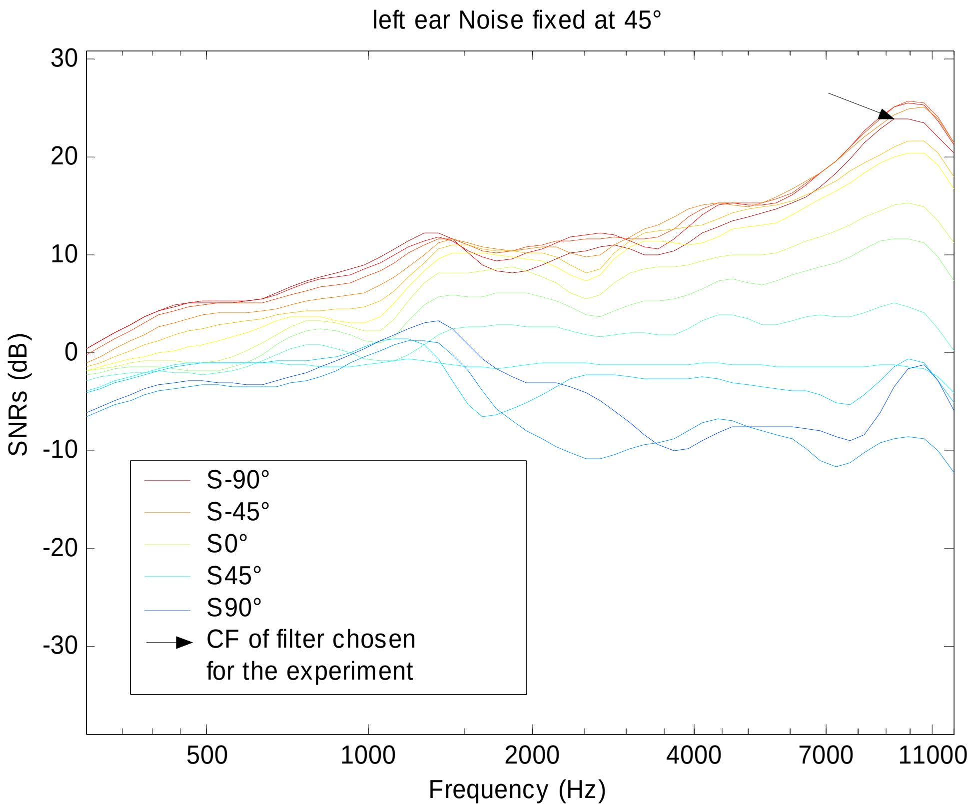


Figure 5.4 Spatial unmasking of BRD stimuli, SNRs differences between Signals azimuths and Noise fixed at 45° in L ear

As we can see from figure 5.4. We analyze left ear with N fixed at 45° and we are looking on S at -90°. We choose only one channel which dominates (as we can see this channel is HF). In our case L channel of 9947,7 Hz. We will measure: bin. Broadband channels, one L channel of 9947,7 Hz.

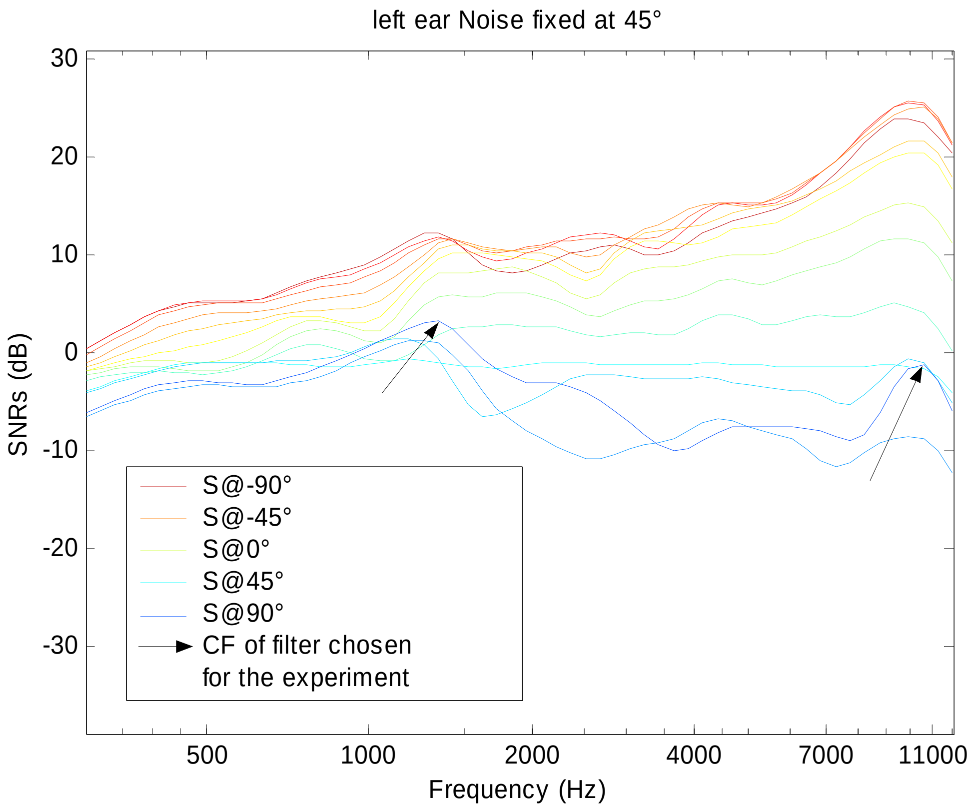


Figure 5.5 Spatial unmasking BRD stimuli, SNRs differences between Signals azimuths and Noise fixed at 45° in L ear

Figure 5.5 shows us also combinations where one channel dominates about 1345 Hz frequency. As we can see N is fixed at 45° and S is fixed at 90, we are looking at left ear. But as it is shown at this picture also fairly good is L channel of 10581 Hz frequency. So we choose these two channels for our experimental procedure. We will measure bin broadband channels, one channel of 1345 Hz, one channel of 10581 Hz, bin. 1345 Hz channels and monaurally broadband L channel.

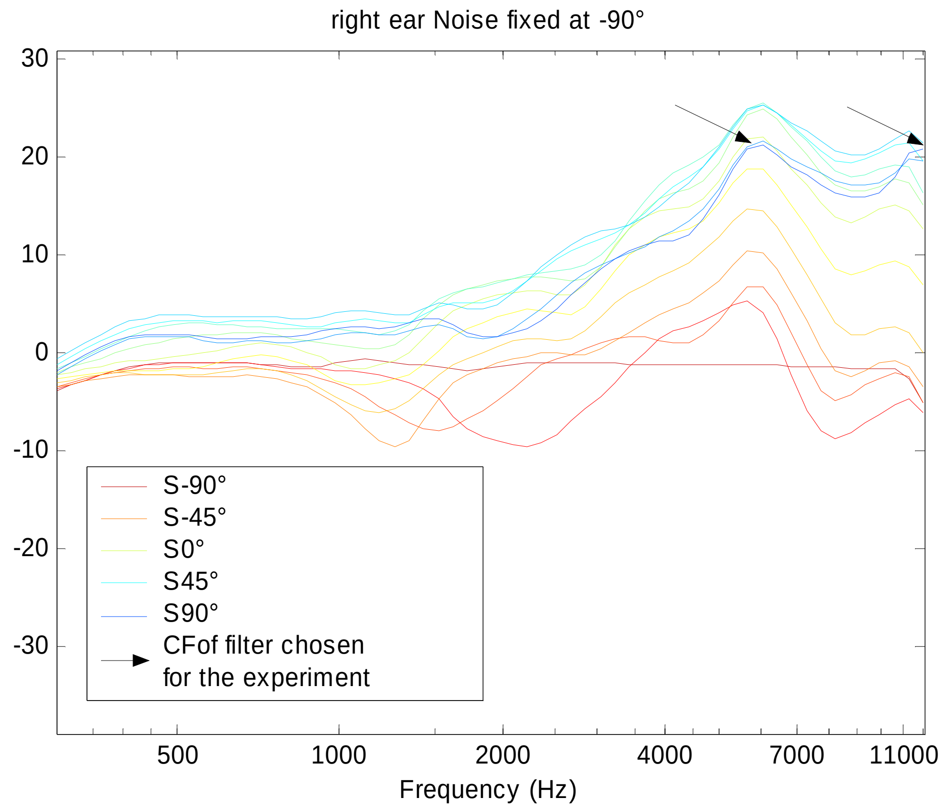


Figure 5.6 Spatial unmasking BRD stimuli, SNRs differences between Signals azimuths and Noise fixed at -90° in L ear

When we look at Figure 5.6 we see more in detail right ear with N fixed at -90° and S at 90°. We can see at least two equivalent HF channels. (here we can include also situation, when S and N are in the same position). These two channels are on following frequencies: 6032 Hz and 12 kHz.

We will measure binaurally broadband channels, one R channel of 6032 Hz, and one channel of 12 kHz frequency.

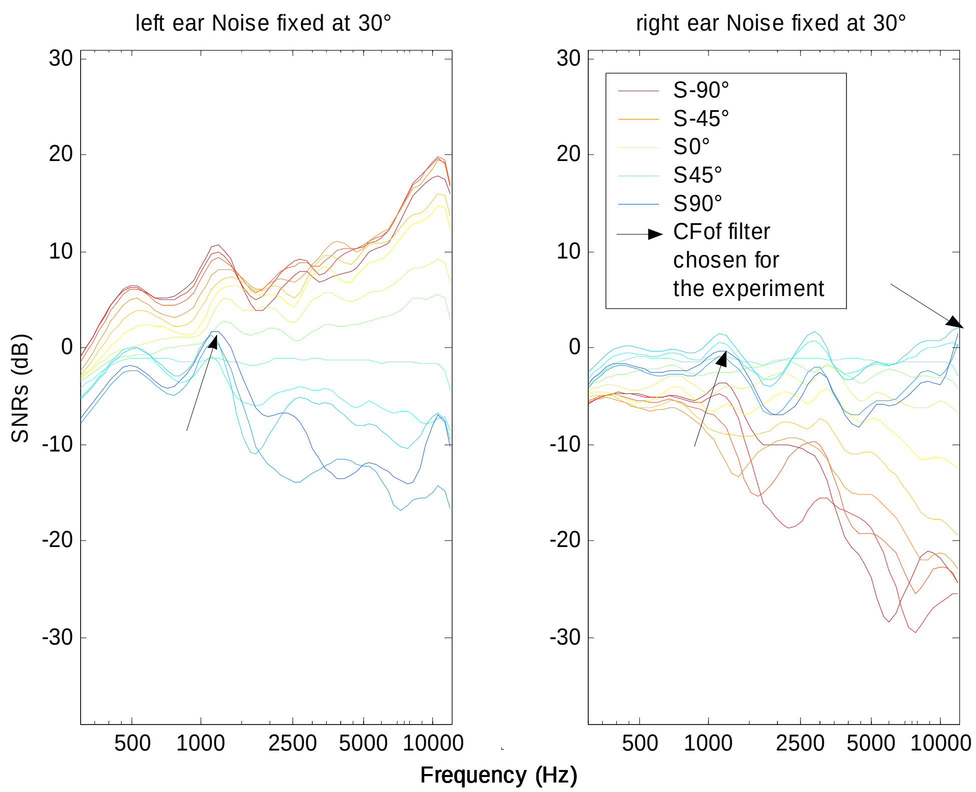


Figure 5.7 Spatial unmasking BRD stimuli, SNRs differences between Signals azimuths and Noise fixed at 30° in L and R ear

This figure 5.7 interprets one HF channel which dominates, but there is also one LF channel, that is a little worse. Because auditory perception is better by helping of binaural effect on LF, so beforehand isn't possible to say which channel will be better. So we choose 2 channels of frequencies: 1187,1 Hz; 12 kHz. Their maximal threshold is about 0 dB.

We will measure bin broadband channels, monaurally one L broadband channel, one L channel of 1187,1 Hz frequency, bin two channels of 1187,1 Hz frequency and one monaurally R channel of 12 kHz frequency.

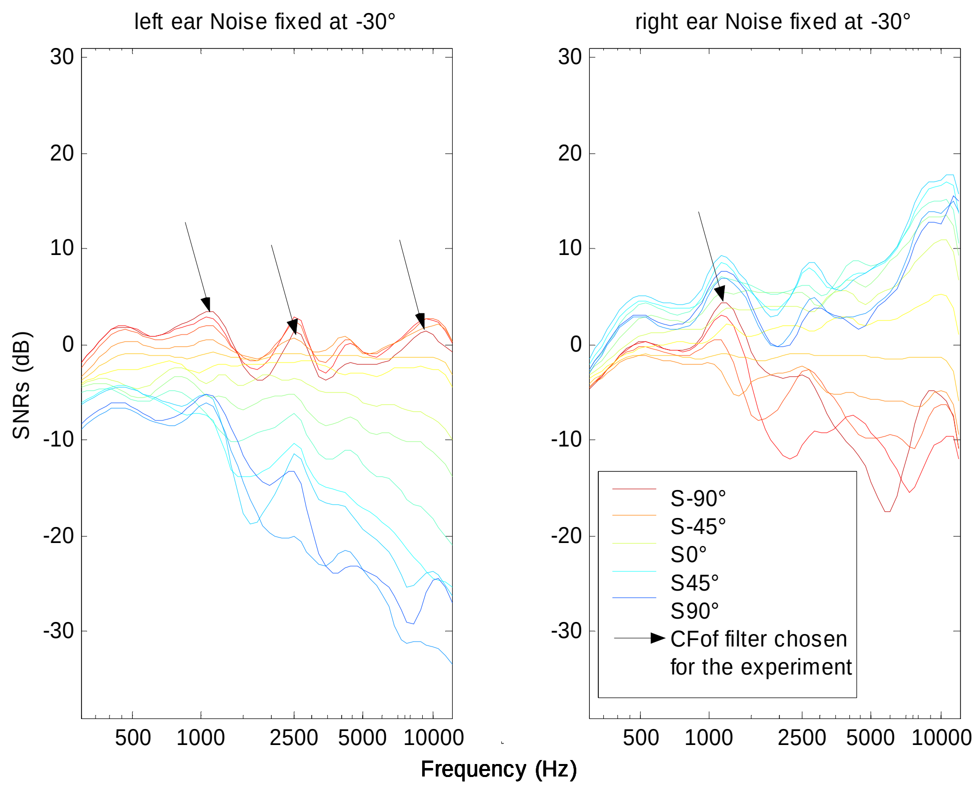


Figure 5.8 Spatial unmasking BRD stimuli, SNRs differences between Signals azimuths and Noise fixed at -30° in L and R ear.

In figure 5.8 is showed situation when we don't know what will dominate. We can ask a question: „What will happen when we will have two equally good channels in different ears? "

So this figure shows this situation: N is fixed at -30°, and we are looking at S of -90° in L and R ear. As it is shown on this picture we have 4 equally good channels of these frequencies: L 1047,6 Hz; L 2513,9 Hz; L 9344,7 Hz; R 1187,1 Hz).

We will measure bin broadband channels, L and R broadband channel, monaurally one L channel of: 1047,6 Hz; 2513,9 Hz; 9344,7 Hz frequency, one monaurally R channel of 1187,1 Hz frequency and two bin channels (in L ear at 1047,6 Hz frequency and in R ear at 1187,1 Hz frequency).

Through HRTFs analyses we choose these spatial combinations of S and N, and frequency signals to our experimental procedure.

AZIMUTH	NOISE	SIGNAL
M 45° T -90°	BRD	BRD
M 45° T -90°	R channel	R channel 9,9 kHz
M 45° T 90°	BRD	BRD
M 45° T -90°	R channel	1,35 kHz BRD
M 45° T -90°	R channel	R channel 10,6 kHz
M 45° T -90°	BRD	bin channel of 1,35 kHz
M 45° T -90°	R channel	R channel
M -90° T 90°	BRD	BRD
M -90° T 90°	L channel	L channel of 6 kHz
M -90° T 90°	L channel	L channel of 12 kHz
M 30° T 90°	BRD	BRD
M 30° T 90°	R channel	R channel
M 30° T 90°	R channel	R channel of 1,19 kHz
M 30° T 90°	BRD	bin channel of 1,19 kHz
M 30° T 90°	L channel	L channel of 12 kHz
M -30° T 90°	BRD	BRD
M -30° T 90°	R channel	R channel
M -30° T 90°	L channel	L channel
M -30° T 90°	R channel	R channel of 1 kHz
M -30° T 90°	R channel	R channel of 2,5 kHz
M -30° T 90°	R channel	R channel of 9,4 kHz
M -30° T 90°	L channel	L channel of 1 kHz
M -30° T 90°	BRD	bin channel of 1 kHz

Table 5.1: Final combinations of filtered Noises and Signals chosen for the experimental procedure. First column (right) represents azimuthal combination M/T, second (in the middle) shows to which ears Noisess come, and left column presents Signals frequencies.

5.2 Experimental procedure

Before experimental procedure we have to make pilot measurement. This measurement was needed because of fine-tuning and testing of functionality in our experimental procedure.

Measurement devices:

- Standard via computer with 15“ display monitor
- professional sound card: **Echo Darla 20**
- professional ear-phones: **Etymotic Research ER-4B**

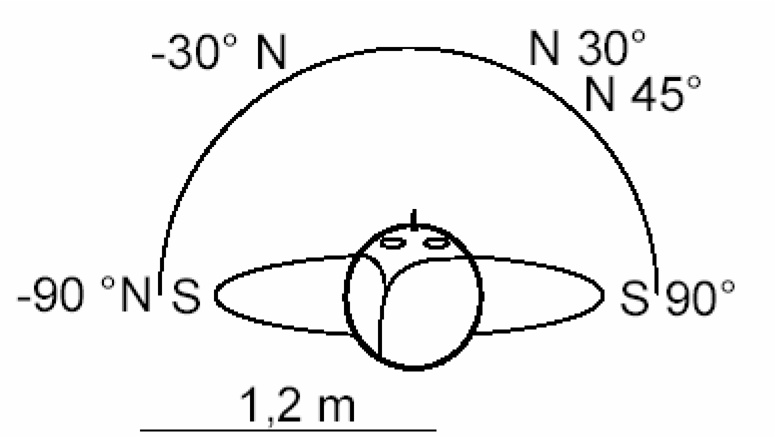
In our experiments were 4 subjects and all were men between 22-24 years. All subjects were before experimental procedure tested by audiogram. That was a little sound exam; it was because of comparing experimental results. By some subjects could dominate the ability of better hearing by R ear opposite L ear. This could have some unfriendly effects for the experimental procedure. Other subjects could have big threshold's deviations on similar frequencies. Or their threshold could be too high over 15 dB over of average listener.

All participants had thresholds at or below 15 dB HL. The audiograms were measured by frequencies of 125 Hz, 250 Hz, 500 Hz, 1 kHz, 2 kHz, 4 kHz and 8 kHz.

On the basis of the above-mentioned signals generated by HRTFs we make experimental procedure for broadband signal measuring. We select for the experimental procedure 23 combinations of S and N, where each of these combinations had been played for each subject 3 times in random sequence. That is 69 measurements for each subject. All sounds were at 1,2m distance. In this experimental procedure subjects made 4 sessions (3 sessions by 20 measurements and 1 session by 9 measurements). The total time of experiment durations for one subject was 4 hours.

The simulations were made in anechoic auditory space.

- 5 simulated position : Noise at 45^o Signal at -90^o
  Noise at 45^o Signal at 90^o
  Noise at -90^o Signal at 90^o
  Noise at 30^o Signal at 90^o
  Noise at -30^o Signal at -90^o



Figures 5.9 Simulated position of used signals (S) and noises (N) in experimental procedure.

The subjects' task was to identify in which of 3 consecutive stimuli (noises) was played chirp. A 3-down-1-up, three-interval, two-alternative, forced-choice procedure was used to estimate detection thresholds (Levitt, 1971), defined as the 79.4% correct point on the psychometric function.

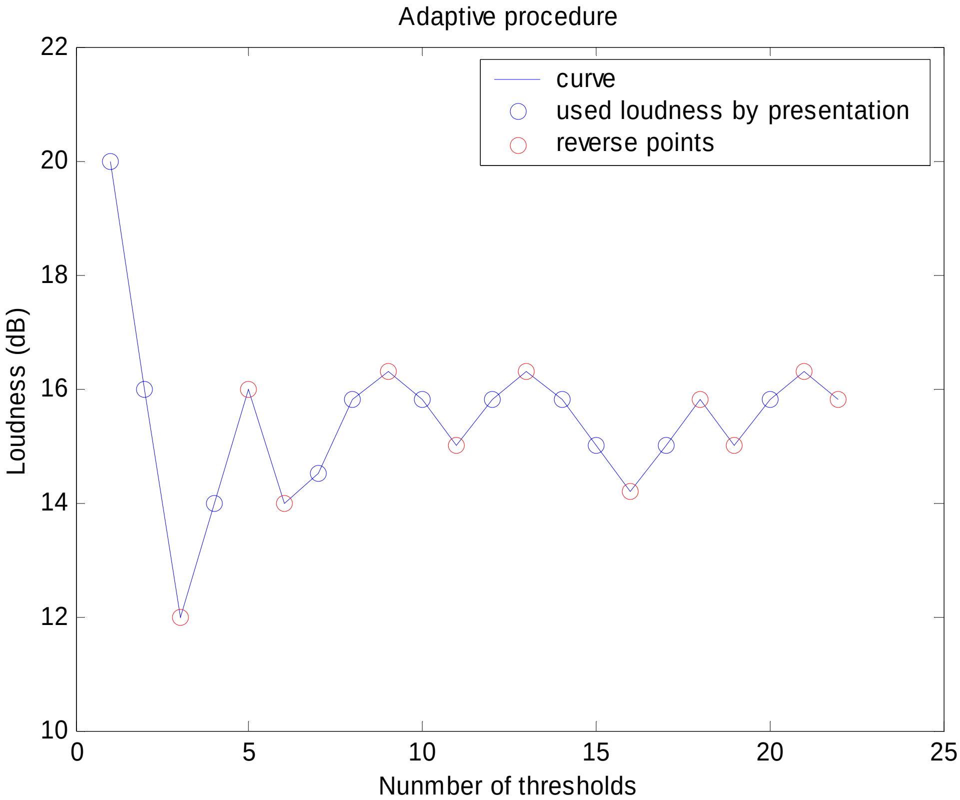


Figure 5.10 Marking reverse points on psychometrics curve in adaptive/staircase method.

This method is similar to the method of limits in hoc that by one measurement it measure only one point on psychometrics curve. The experimenter can choose which point will be measured. His sequence is to define under what conditions stimuli increase or decrease. For example, we say that the loudness decreases when the subject hears sound and vice-versa that loudness increases always when the subject not hears the sound. The course of this measurement can be following: The subject heard by the first three presentations chirp, so the loudness gradually decreases. Then he don't hear two times chirp so the loudness increases twice. Then he hear, etc. etc. The loudness by which will change movement of volume (in case of previous measurement the loudness decreases and consequently increases) is called reverse point (see fig. 5.10)

Each run started with the T at a clearly detectable level and continued until 11 "reversals" occurred. The T level was changed by 4 dB on the first reversal, 2 dB on the second reversal, and 1 dB on all subsequent reversals. For each adaptive run, detection threshold was estimated by taking the average T presentation level over the last six reversals. At least three separate runs were performed for each subject in each condition. Final threshold estimates were computed by taking the average threshold across the repeated adaptive threshold estimates. Additional adaptive runs were performed as needed for every subject and condition to ensure that the standard error in this final threshold estimate was less than or equal to 1 dB.

In this experiment the signal (200ms 40 Hz chirp train) was presented by BRD tones masked by 250ms white noise bursts and we measure threshold loudness of tone detection. White noise was always present by uniform loudness and identically for both channels (L and R earphone). We change only the loudness of signal to destine its threshold value.

In this experimental procedure we try to measure experimental thresholds for chosen combination of S and N location. So these thresholds were measured for broadband signals, for signals filtered by the best expected channels, also for binaural signals and also for monaural broadband signals. The results of this experimental study are showed in chapter 6.

The experimental results were averaged by arithmetic mean according to $\overline{x}_j = \frac{1}{n}\sum_{i=1}^{n} X_{ij}$ and drew in to the figures included with the standard error of the mean – calculated by $\sigma_M = \frac{std}{\sqrt{n}}$. Where STD is standard deviation rated by $std = \left(\frac{1}{n-1} \sum_{i}^{n} \left(x_i - \overline{x} \right)^2 \right)^{\frac{1}{2}}$.

6. Results

This chapter first summarizes the overall data by presenting the across-subject average of the thresholds measured in all conditions (Figure 6.1). Then, in Figures 6.2 through 6.6, individual subject data are presented and discussed.

6.1 Overall results

Figure 6.1 shows the across-subject means in the measured threshold SNRs for all conditions as a function of the target/masker configuration. In each configuration, filled symbols show the broadband thresholds while open symbols show the thresholds measured with filtered stimuli, as described in the labels next to the symbols.

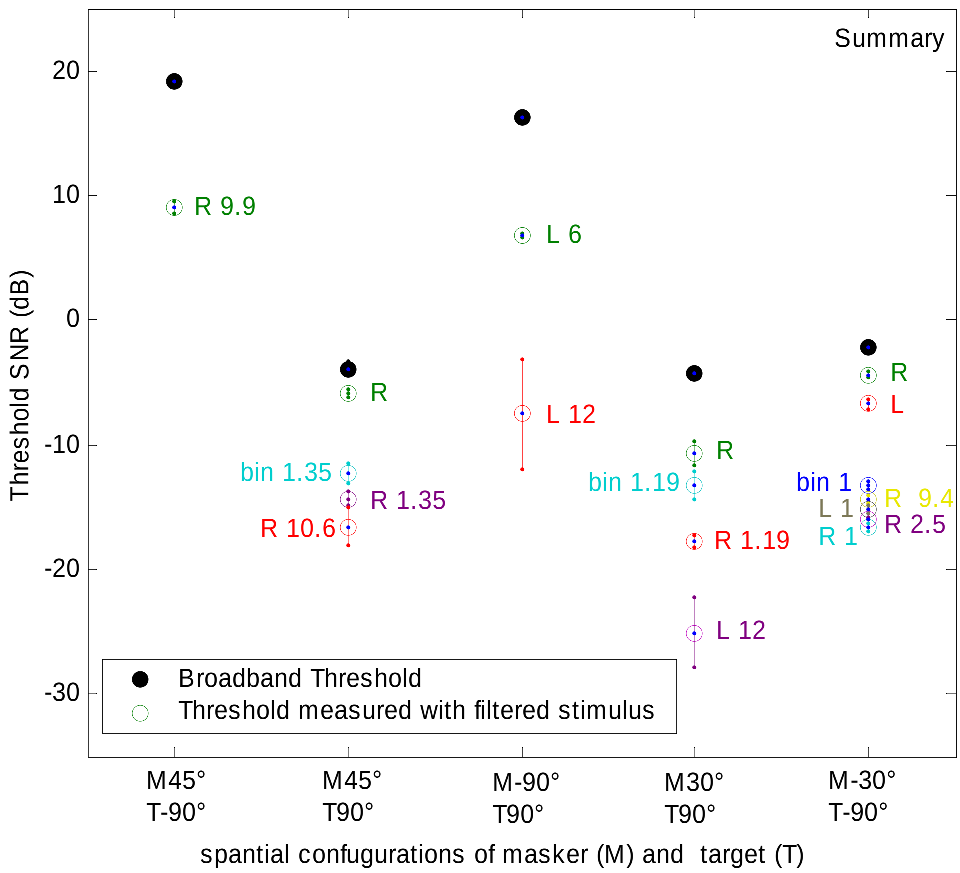


Figure 6.1 Overall Results: Across-Subject Means and standard errors in threshold SNRs for all conditions measured in the experiment. Filled symbols represent broadband thresholds, open symbols represent thresholds obtained with filtered stimuli. Every filtered threshold has a legend that describes the filter: R –

right ear only, L – left ear only, bin – binaural (hearing with two ears). The number near some filtered thresholds means frequency of the filter in kHz.

Overall, there are very small across-subject differences (small standard errors) with exception of several thresholds for single-filter measurements at high frequencies). This increased difference might have been caused by increased high-frequency thresholds in some of the listeners.

When the masker (M) is at 45° and the target (T) is at -90° (left-most column in Fig. 6.1, condition described in Fig 5.4) the difference between the broadband threshold (filled symbol) and the threshold obtained with the best single GammaTone-filtered stimulus (right ear, center frequency 9.9 dB) is approximately 10 dB. That means that the single-best filter does not determine the broadband threshold in this condition.

In the configuration of M at 45° and T at 90° (second column from left in Fig. 6.1, condition described in Fig 5..5) we can see that the binaural contribution is only 2-3dB as reported in the previous study. However, the difference between the broadband threshold and the SBF threshold is still more than 7dB.

In the condition of M at -90° and T at 90° (center column in Fig. 6.1, condition described in Fig 5.6) a large 10-dB difference between the broadband and the SBF thresholds is again observed.

The 4th measured configuration was M at 30° and T at 90° (second column from right in Fig. 6.1, condition described in Fig 5.7). Here we can see that binaural contribution is fairly large (more than 5 dB), both in a single channel and for broadband stimuli. And the difference between the SBF thresholds and the broadband thresholds is again 10dB.

The last simulated configuration was M at -30° T at -90° (the right-most column in Fig. 6.1, condition described in fig 5.8). There is a small binaural contribution both for broadband and for single-channel thresholds. However, the SBF threshold is again more than 7 dB below the broadband one.

In contrast with the hypotheses stated in Chapter 4, these results suggest that:

1. Thresholds measured with broadband stimuli are always better than the single-best filter ones. This does not support the hypothesis that

spatial unmasking of broadband stimuli masked by noise is determined by the single peripheral auditory channel with the most favorable signal to noise ratio.

2. Low frequencies as well as the high frequencies can be important for spatial unmasking of binaural stimuli.
3. Binaural processing can contribute as much as 5 dB to spatial unmasking in certain target/masker configurations.

6.2 Individual subject results

Figures 6.2 through 6.6 show the individual subject results. In general, the results have the same trends as the overall results discussed in Chapter 6.1. The only significant difference is in that some of the subjects (Subject BB in Figure 6.3 and Subject MB in Figure 6.4) have their thresholds with high-frequency stimuli much lower than the other thresholds in the M-90° T90° and M30° T90° conditions. This difference can be a consequence of a mild hearing loss of the two subjects at high frequencies. As can be see in the subject's audiograms (see Appendix), subject BB has a high-frequency hearing loss. No hearing loss is shown for subject MB, however the audiograms are measured only for frequencies of up to 8 kHz, and therefore they don't test the listener's sensitivity for frequencies of 12 kHz at which the differences were observed.

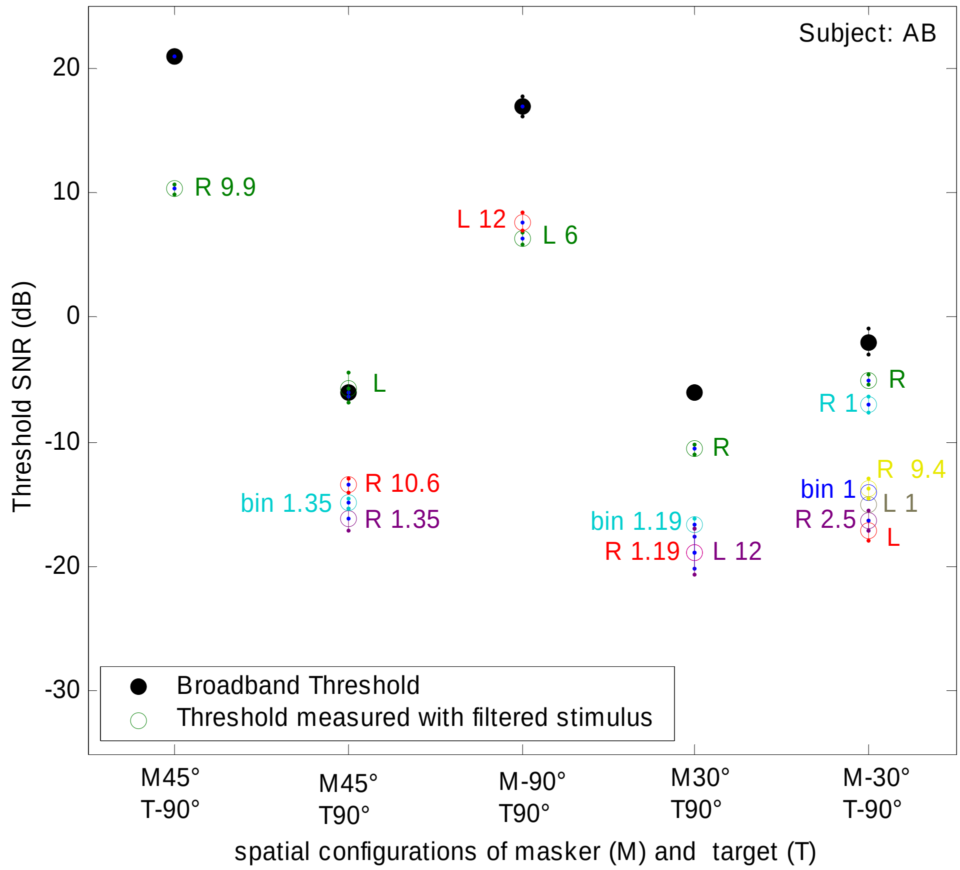


Figure 6.2 Means and standard errors in Threshold SNRs for subject AB. For detailed description see the caption of Figure 6.1.

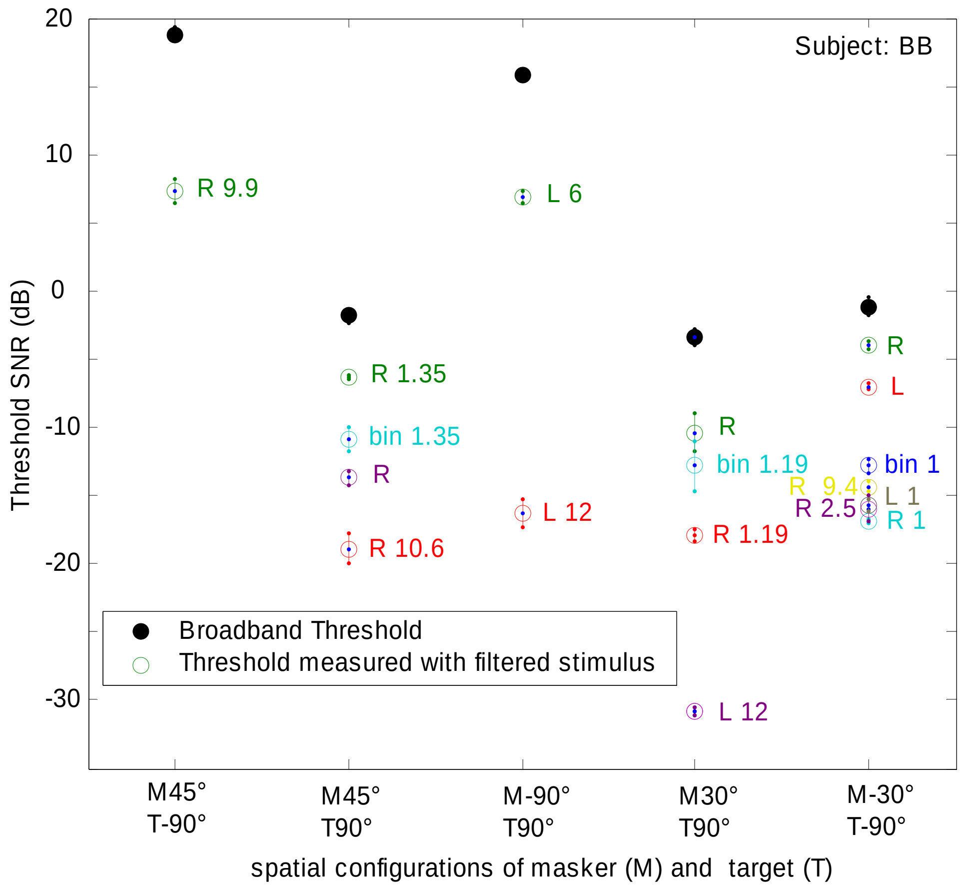


Figure 6.3 Means and standard errors in Threshold SNRs for subject BB. For detailed description see the caption of Figure 6.1

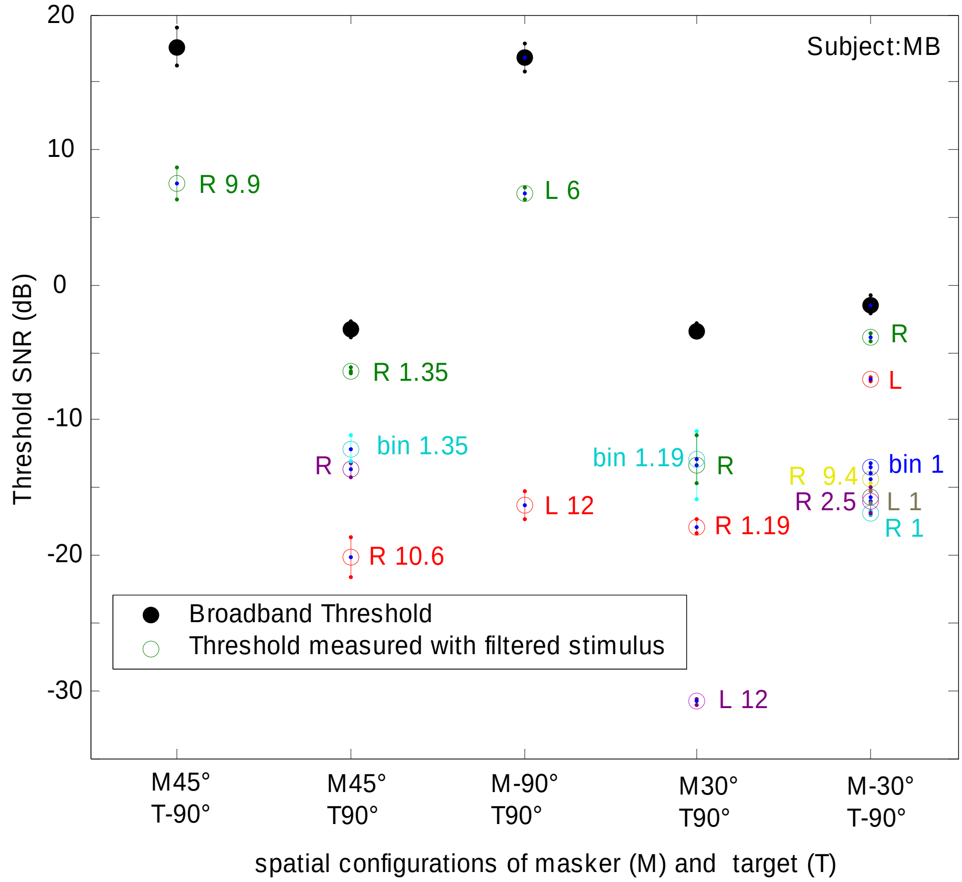


Figure 6.4 Means and standard errors in Threshold SNRs for subject MB. For detailed description see the caption of Figure 6.1

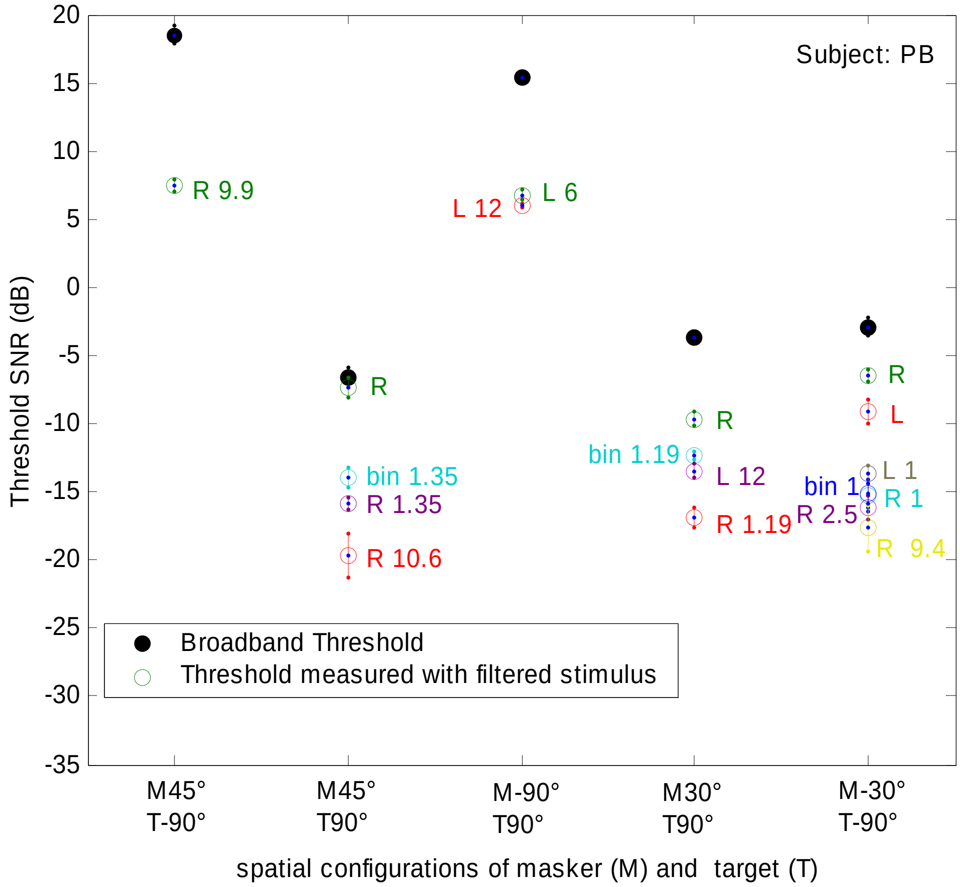


Figure 6.5 Means and standard errors in Threshold SNRs for subject AB. For detailed description see the caption of Figure 6.1

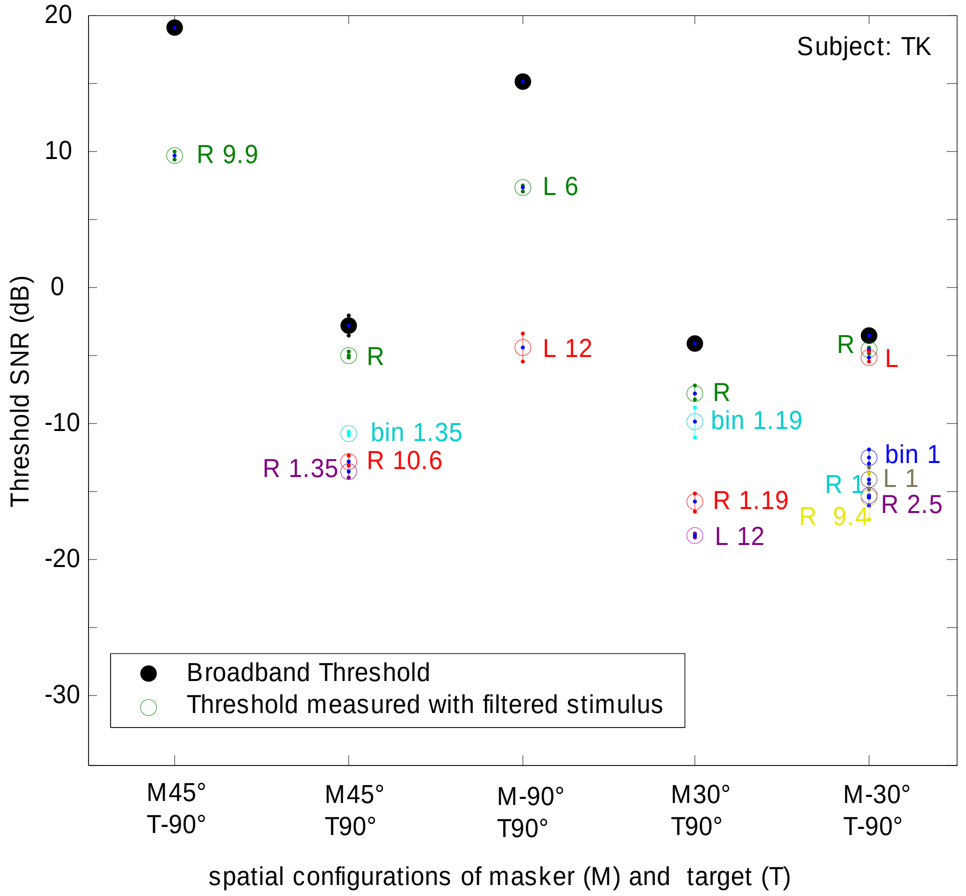


Figure 6.6 Means and standard errors in Threshold SNRs for subject AB. For detailed description see the caption of Figure 6.1

6.3 Analysis of other parameters

Several other parameters, like standard deviations, were also analyzed for the current data. The results of this analysis can be found in the appendix.

7. Conclusions & Summary

The current study differs from the previous studies of spatial unmasking in that it test the single-best-filter hypothesis directly by comparing behaviorally measured thresholds obtained with broadband stimuli to thresholds for stimuli that contain energy only in the single best filter.

The current results differ significantly from the ones obtained by Kopco et al (2003) in three ways:

1. In contrary to Kopco et al (2003) we do find spatial configurations where low-frequency spectral content is important for spatial unmasking.
2. Binaural contribution to spatial unmasking can be as much as 5 dB.
3. The single best filtered thresholds are always lower than the broadband thresholds, i.e., information is always integrated across multiple channels when subject listens to broadband stimuli.

Based on these results it can be concluded that a model that is more complex than the Single-best filter model has to be considered. The model has to be extended in at least two aspects:

1. to consider across-channel integration
2. to include improvement in detectability due to binaural processing

Dau (1996) proposed a model that implements across-channel interaction as well as another feature that might be important for modeling of the current data – amplitude modulation. This model, in combination with available models of binaural processing, can be more successful in predicting the current data.

One important question is why the current results are so different from the results of Kopco et al. (2003). One hypothesis is that the experimental procedures weren't completely comparable. However, a detailed analysis of all stimuli and experimental code will be necessary to determine the real cause of the differences.

Evaluation of the tasks defined for this thesis:

The main goal of this thesis was to study spatial unmasking of broadband stimuli in a virtual auditory environment. This main goal was divided into several subtasks, the fulfillment of which can be evaluated as follows:

1st task. In chapters 2 and 3 give basic information about spatial auditory perception, spatial unmasking, the cocktail party problem and on Dau's model of processing of sound signals in peripheral auditory system.

2nd task. In chapter 5 the procedure for experimental measurement of detection thresholds of broadband stimuli is described.

3rd task. Realized in the Perception and Cognition Laboratory at TU Košice.

4th task. Analyses and evaluation of the experimental data is in chapters 6 and 7.

8. References

[1] GOOD MD, GILKEY RH, BALL JM (1997): The relation between detection in noise and localization in noise in the free field. In RH Gilkey and TR Anderson (Eds, Binaural and Spatial Hearing in Real and Virtual Environments. Lawrence Erlbaum, Mahawah, NJ, pp. 349-376.

[2] JOHANNESMA PIM (1972): The pre-response stimulus ensemble of neurons in the cochkear nucleus. in IPO Symposium on Hearing Theory, edited by B. L. Cardozo, E. de Boer, & R. Plomp (IPO, Eindhoven, The Netherlands), pp. 58-69.

[3] KOPČO N, SHIN-CUNNINGHAM BG (2000): Spatial unmasking of nearby pure-tone sources in a simulated anechoic environment. J. Acoust. Soc. Am. 107(5 Pt. 2), 2849 (Presented at the 139th meeting of the Acoustical Society of America).

[4] KOPČO N, SHIN-CUNNINGHAM BG (2003): Spatial unmasking of chirp trains in a simulated anechoic environment: Behavioral results and model predictions. Association for Research in Otolaryngology Abstract #935.

[5] YOST WA (2000): Fundamentals of Hearing: An Introduction, Parmly Hearing Institute, Loyola University Chicago, ACADEMIC PRESS

[6] MOORE BCJ (1997): An Introduction to the Psychology of Hearing (4e). San Diego, CA: Academic Press.

[7] GILKEY RH and GOOD MD (1995): Effects of frequency on free-field masking, Human Factors 37(4): 835-843.

[8] LANE CC, DELGUTTE B and COLBURM HS (2003): A Population of ITD-Sensitive Units in the Cat Inferior Colliculus Shows Correlates of Spatial Release from Masking, ARO Abstract #706, Session Q10.

[9] MRAZ LISA (1999) "Understanding Speech in a Noisy Room: The Effect of Distance" Unpublished Senior Thesis, Boston University Dept of Biomedical Engineering.

[10] KOPCO N (2003): Auditory spatial perception and its neural code) In Sinčák et al. (Eds.) Slovensko-české rozpravy o umelej inteligencii (Collection of papers presented at the 3rd Slovak-Czech Seminar on Cognition, Artificial life, and

Computational intelligence CALCI 2003, May 15-17, 2003, Stará Lesná, Slovakia), pp. 213-218

[11] LANE CC, KOPCO N, DELGUTTE B, SHINN-CUNNINGHAM BG and COLDBURN HS (2004): A cat's cocktail party: Psychophysical, neurophysiological and computational studies of spatial release from masking. In: Auditory signal processing: Physiology, psychoacoustics, and models. (Pressnitzer, D., de Cheveigné, A, McAdams, S., and Collet, L., Eds), pp 327-333, Springer, New York. (Proc. International Symposium on Hearing, Dourdan, France, Aug. 24-29, 2003)

[12] CHERRY EC (1953): Some experiments on the recognition of speech with one and with two ears, J. Acoust. Soc. Am. 25, 975-703

[13] Ebata M, Sone T, Nimura T (1968) Improvement of hearing ability by directional information. Journal of the Acoustical Society of America 43:289-297.

[14] Saberi K, Dostal L, Sadralodabai T, Bull V, Perrott DR (1991) Free-field release from masking. Journal of the Acoustical Society of America 90:1355-1370.

[15] Kidd G, Mason CR, Rohtla TL, Deliwala PS (1998) Release from masking due to spatial separation of sources in the identification of nonspeech auditory patterns. Journal of the Acoustical Society of America 104:422-431.

[16] Freyman RL, Helfer KS, McCall DD, Clifton RK (1999) The role of perceived spatial separation in the unmasking of speech. Journal of the Acoustical Society of America 106:3578-3588.

[17] Hawley ML, Litovsky RY, Colburn HS (1999) Speech intelligibility and localization in a multi-source environment. Journal of the Acoustical Society of America 105:3436-3448.

[18] Saberi K, Dostal L, Sadralodabai T, Bull V, Perrott DR (1991) Free-field release from masking. Journal of the Acoustical Society of America 90:1355-1370.

[19] http://xenia.media.mit.edu/~barons/cocktail.html

[20] DAU T, PÜSCHEL and KOHLRAUSCH A (1996): "A quantitative model of the "effective" signal processing in the auditory system. I. Model structure", *Journal of the Acoustical Society of America* 99, 3615-3622

[21] Dau, T., Püschel, D. and Kohlrausch, A. (1996): "A quantitative model of the “effective“ signal processing in the auditory system: II. Simulations and measurements", *Journal of the Acoustical Society of America 99,* 3623-3631

[22] DAU T (1996): "Modeling auditory processing of amplitude modulation", *Bibliotheks - und Informationssystem der Universität Oldenburg, ISBN 3-8142-0570-7.*

[23] KOPCO **N** (2003): Spatial hearing, auditory sensitivity, and pattern recognition in noisy environments, Graduate School of Arts and Sciences, Boston University. Published by UMI.

[24] http://hincapie.psych.purdue.edu/Research/Cocktail.html

9. Appendix

1. CD medium
2. Users and Systems manual

- annex a: all figures and tables
- annex b: user’s and system’s manual

10. List of Figures and Tables

List of figures:

List of tables:

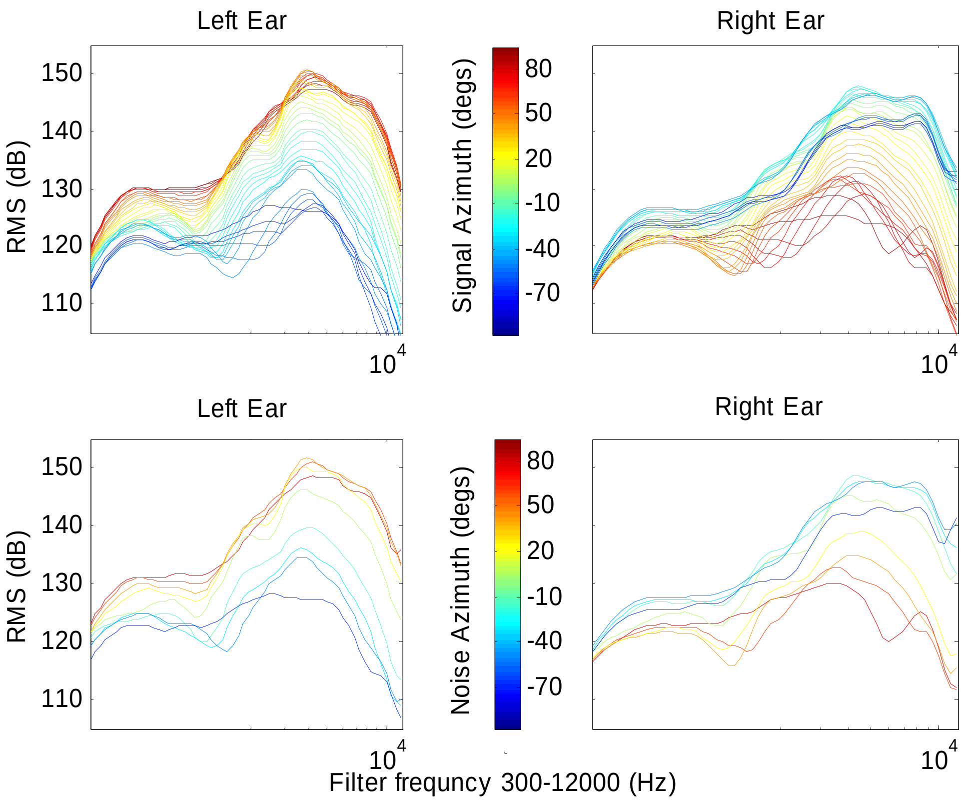


Figure 10.1 Spatial unmasking of BRD stimuli, SNRs differences between all simulated Signals (up) azimuths and all Noises (down) azimuths

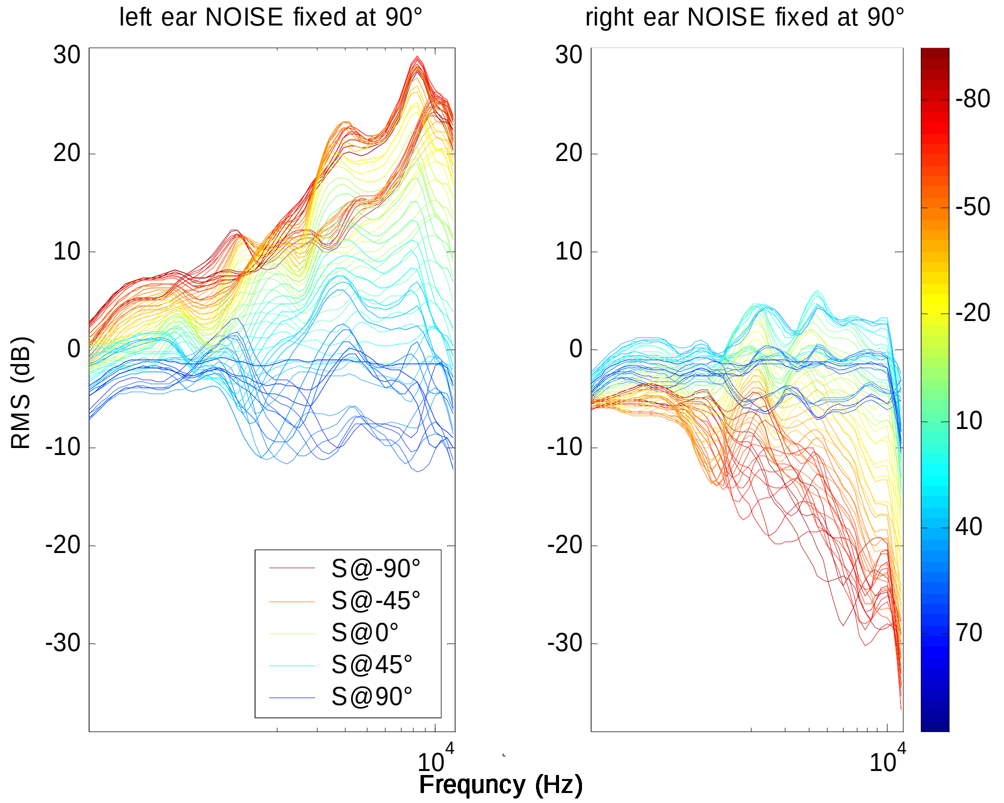


Figure 10.2 Spatial unmasking of BRD stimuli, SNRs differences between all Signals azimuths and Noise fixed at 90° in L and R ear.

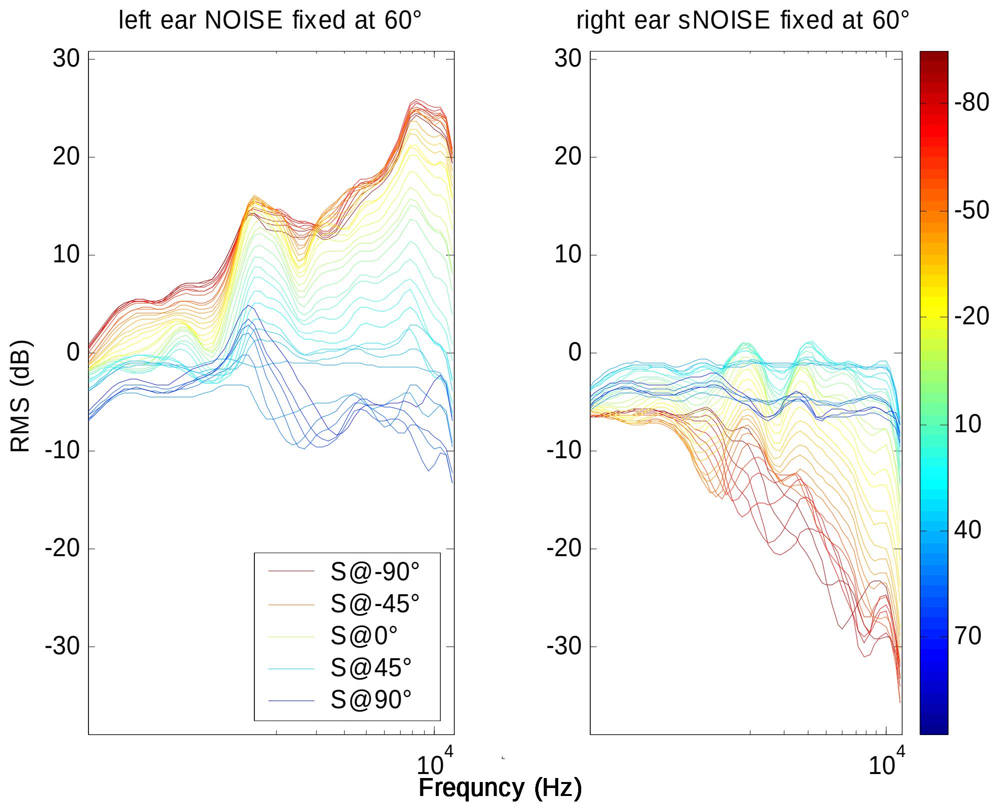


Figure 10.3 Spatial unmasking of BRD stimuli, SNRs differences between all Signals azimuths and Noise fixed at 60° in L and R ear.

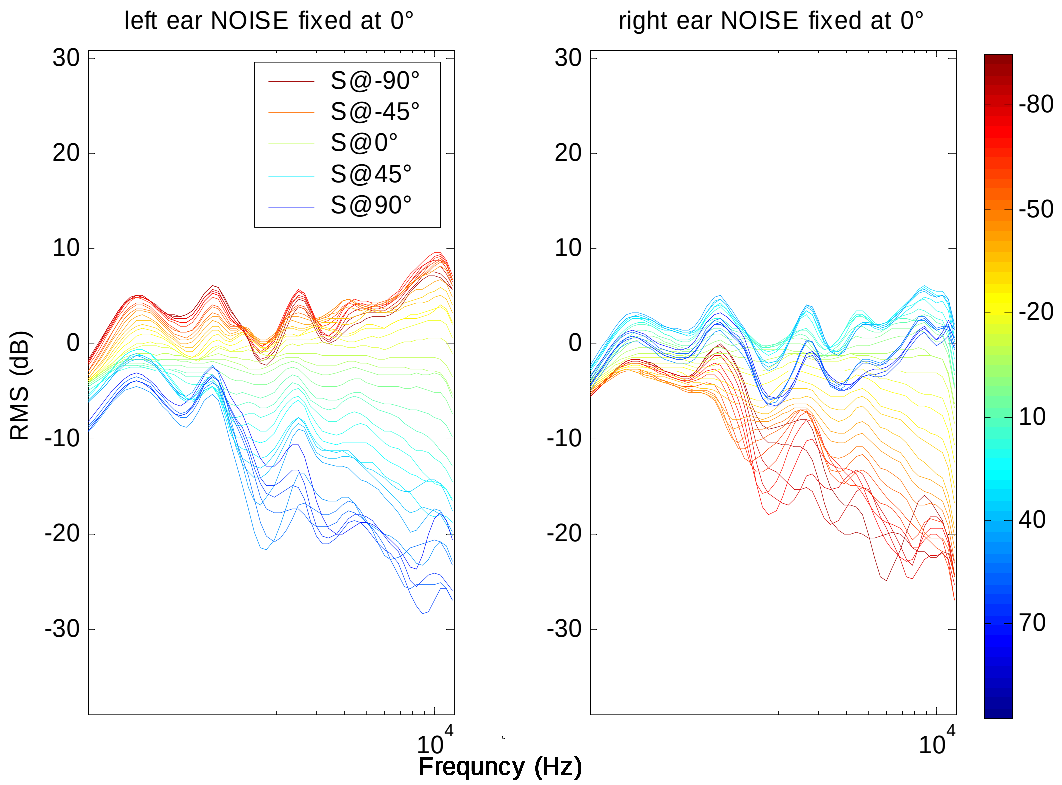


Figure 10.4 Spatial unmasking of BRD stimuli, SNRs differences between all Signals azimuths and Noise fixed at 0° in L and R ear

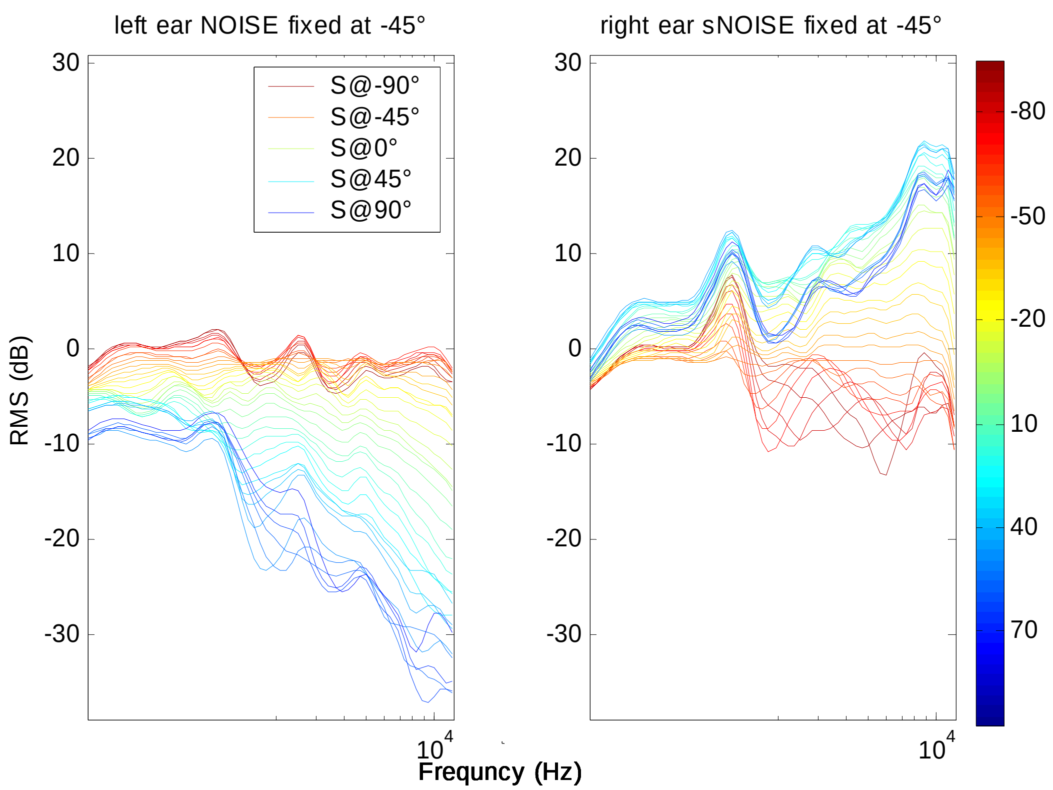


Figure 10.5 Spatial unmasking of BRD stimuli, SNRs differences between all Signals azimuths and Noise fixed at -45° in L and R ear

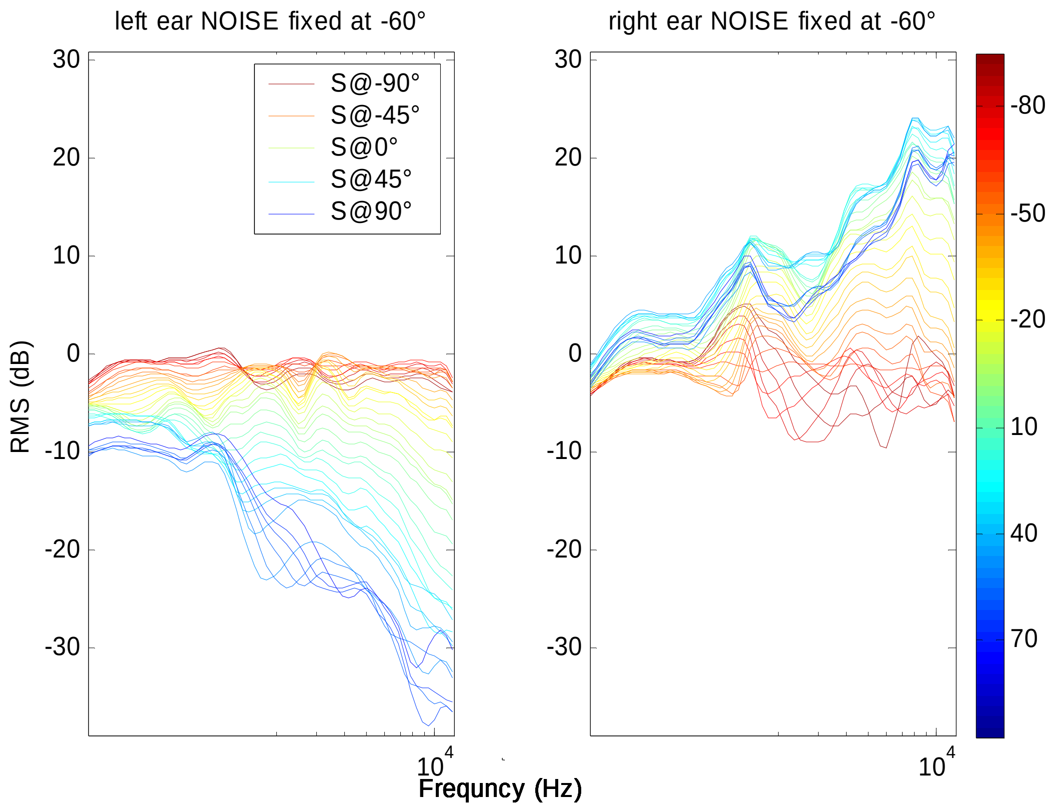


Figure 10.6 Spatial unmasking of BRD stimuli, SNRs differences between all Signals azimuths and Noise fixed at -60° in L and R ear

Subject AB:

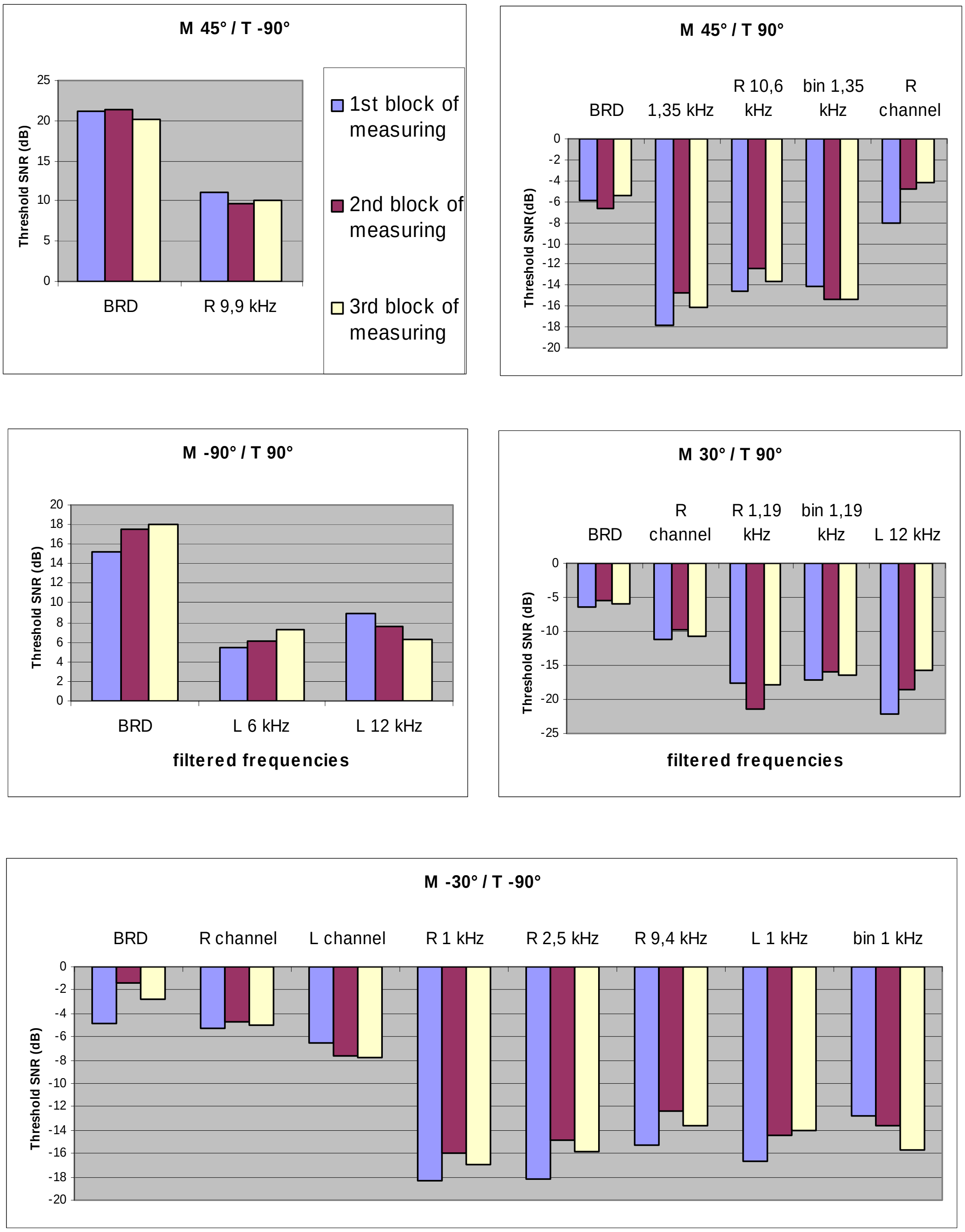


Figure 10.7 Graphs of measured SNRs thresholds for showed azimuthal combination (tab. 5.1) of M/T for subject: AB. Three columns per each M/T combination represent 3 blocks of measuring.

Subject BB:

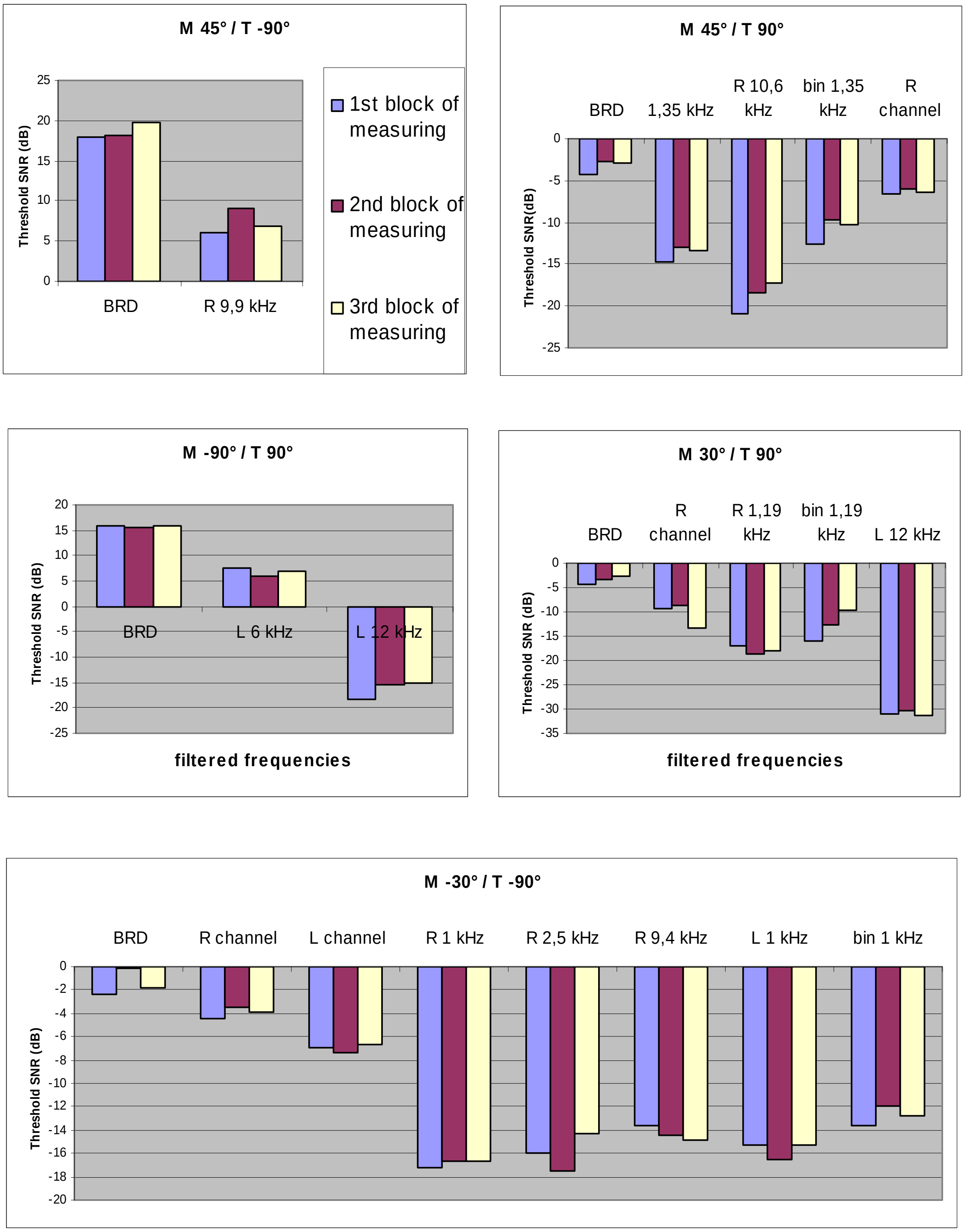


Figure 10.8 Graphs of measured SNRs thresholds for showed azimuthal combination (tab. 5.1) of M/T for subject: BB. Three columns per each M/T combination represent 3 blocks of measuring.

Subject MB:

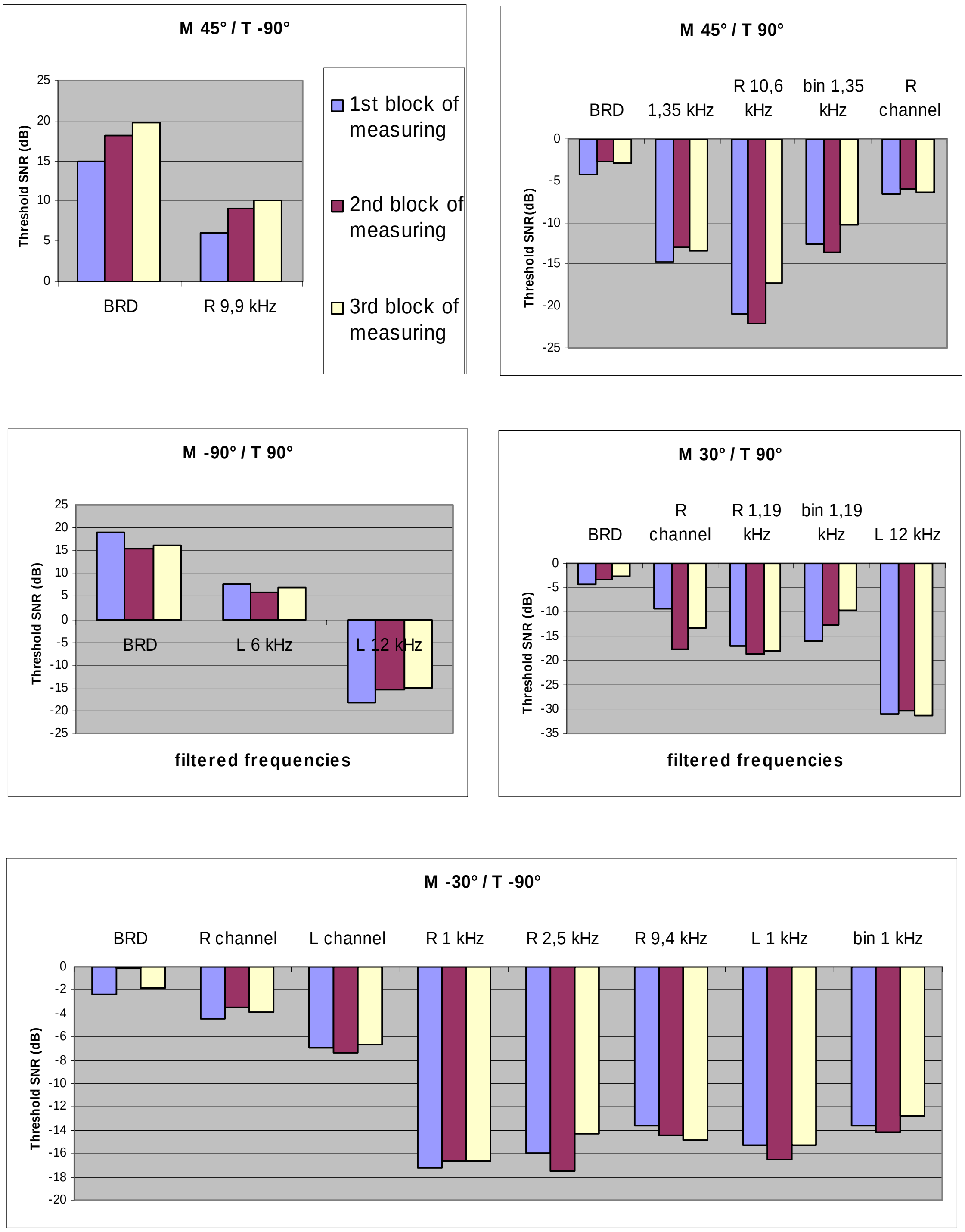
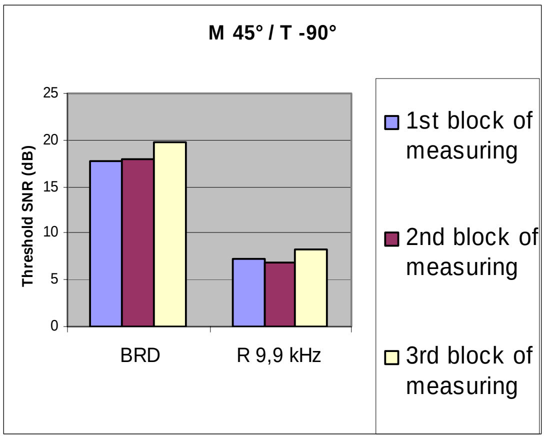
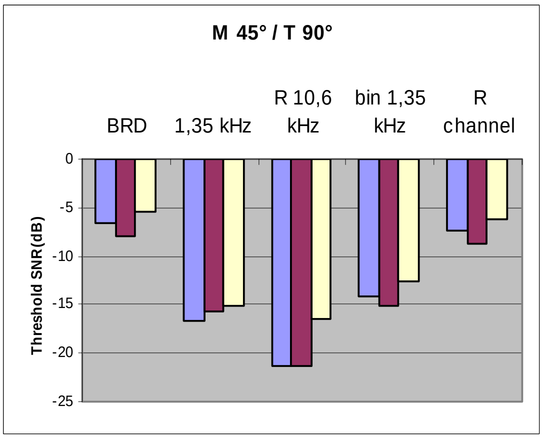
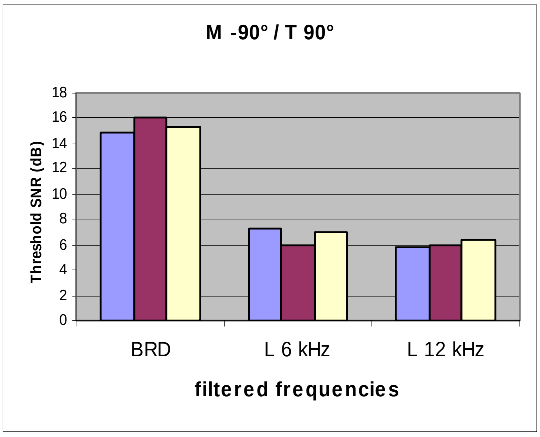


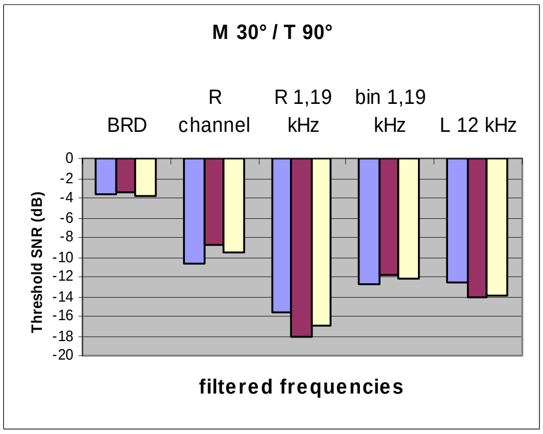
Figure 10.9 Graphs of measured SNRs thresholds for showed azimuthal combination (tab. 5.1) of M/T for subject: MB. Three columns per each M/T combination represent 3 blocks of measuring.

Subject PB:









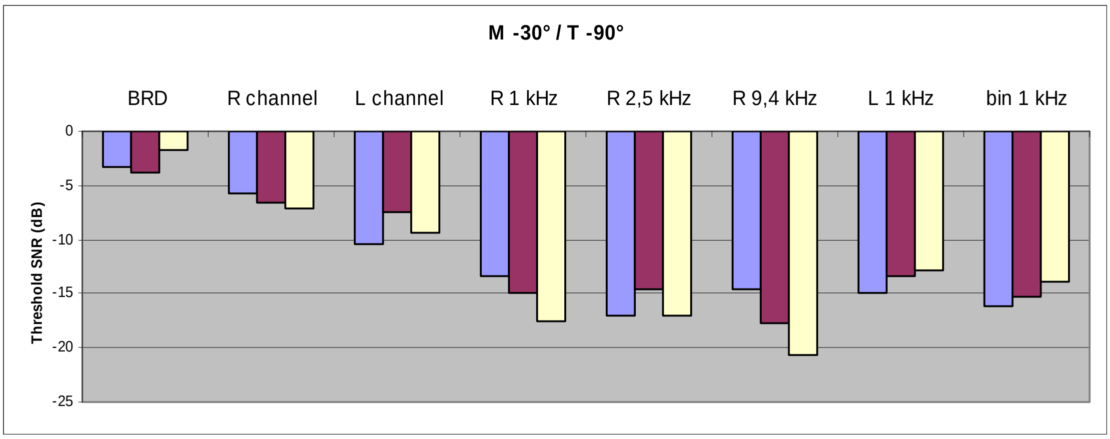


Figure 10.10 Graphs of measured SNRs thresholds for showed azimuthal combination (tab. 5.1) of M/T for subject: PB. Three columns per each M/T combination represent 3 blocks of measuring.

Subject TK:

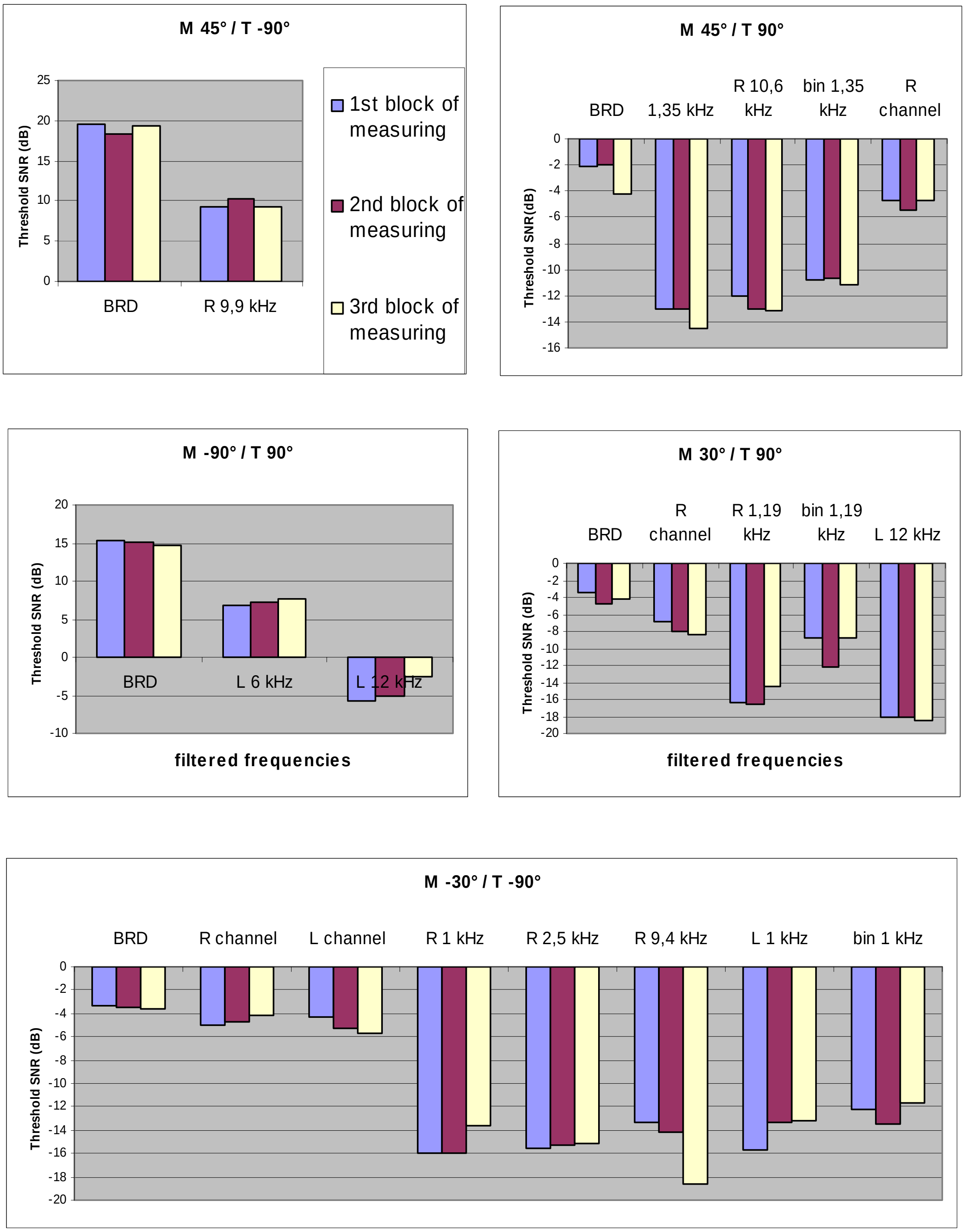


Figure 10.11 Graphs of measured SNRs thresholds for showed azimuthal combination (tab. 5.1) of M/T for subject: TK. Three columns per each M/T combination represent 3 blocks of measuring.

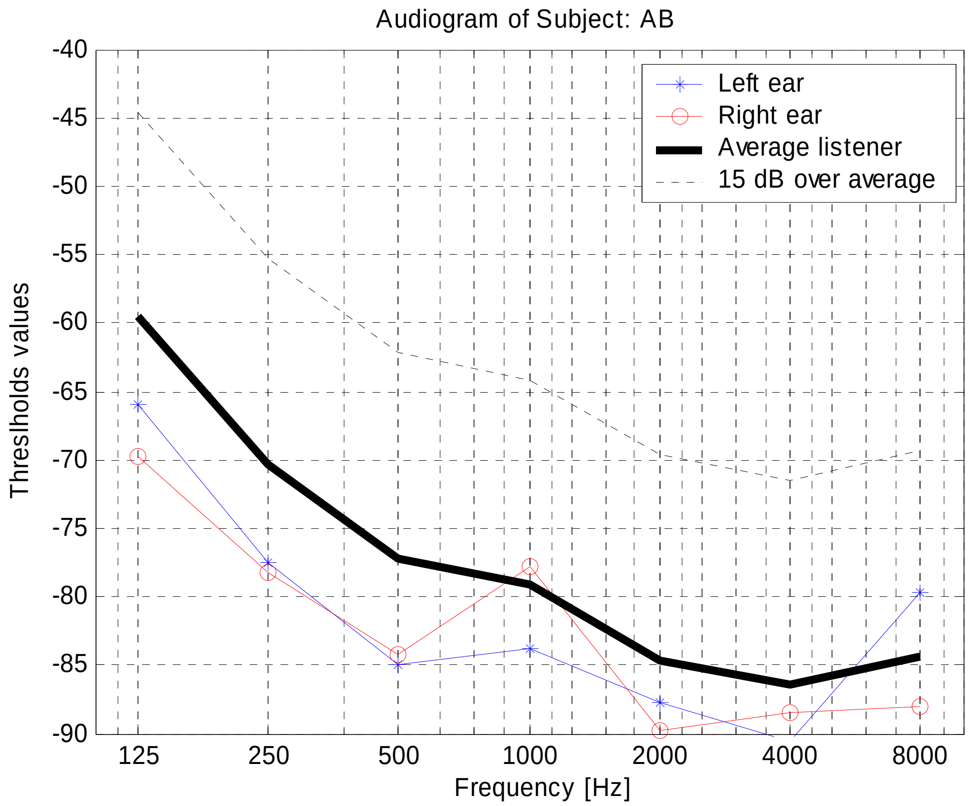


Figure 10.12 Thresholds values of the subject AB, blue line means L - ear, red line means R – ear, black line means average listener and dashed line means15 dB over average. Circles and stars mean measured thresholds on showed frequencies.

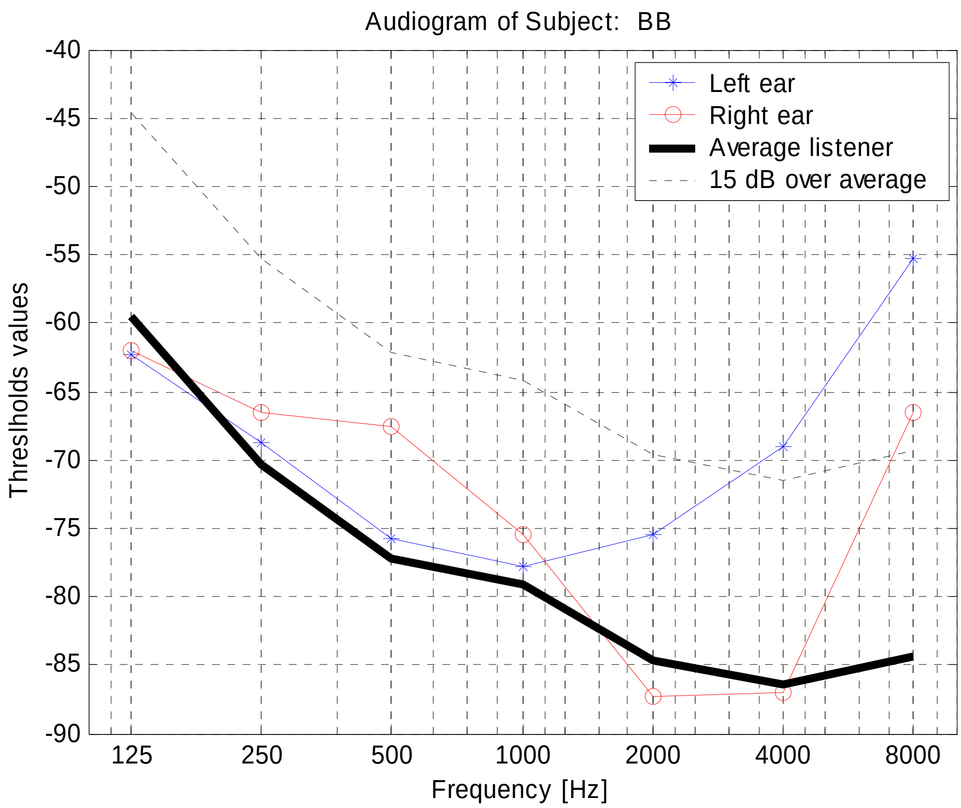


Figure 10.13 Thresholds values of the subject BB, blue line means L - ear, red line means R – ear, black line means average listener and dashed line means15 dB over average. Circles and stars mean measured thresholds on showed frequencies.

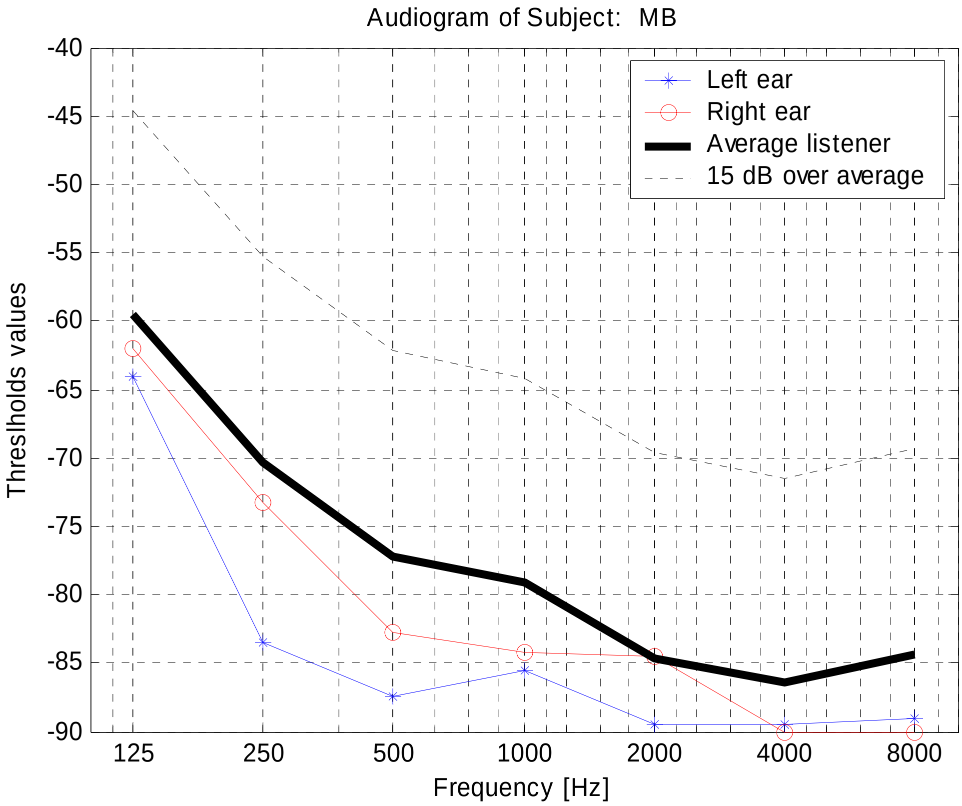


Figure 10.14 Thresholds values of the subject MB, blue line means L - ear, red line means R – ear, black line means average listener and dashed line means15 dB over average. Circles and stars mean measured thresholds on showed frequencies.

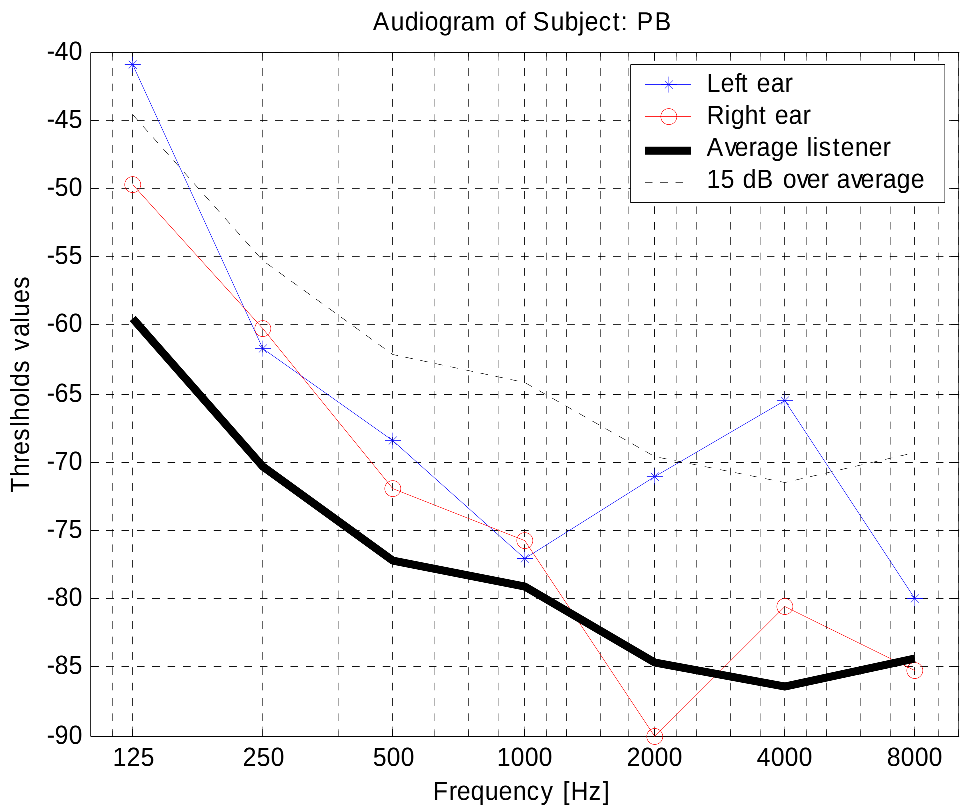


Figure 10.15 Thresholds values of the subject MB, blue line means L - ear, red line means R – ear, black line means average listener and dashed line means15 dB over average. Circles and stars mean measured thresholds on showed frequencies.

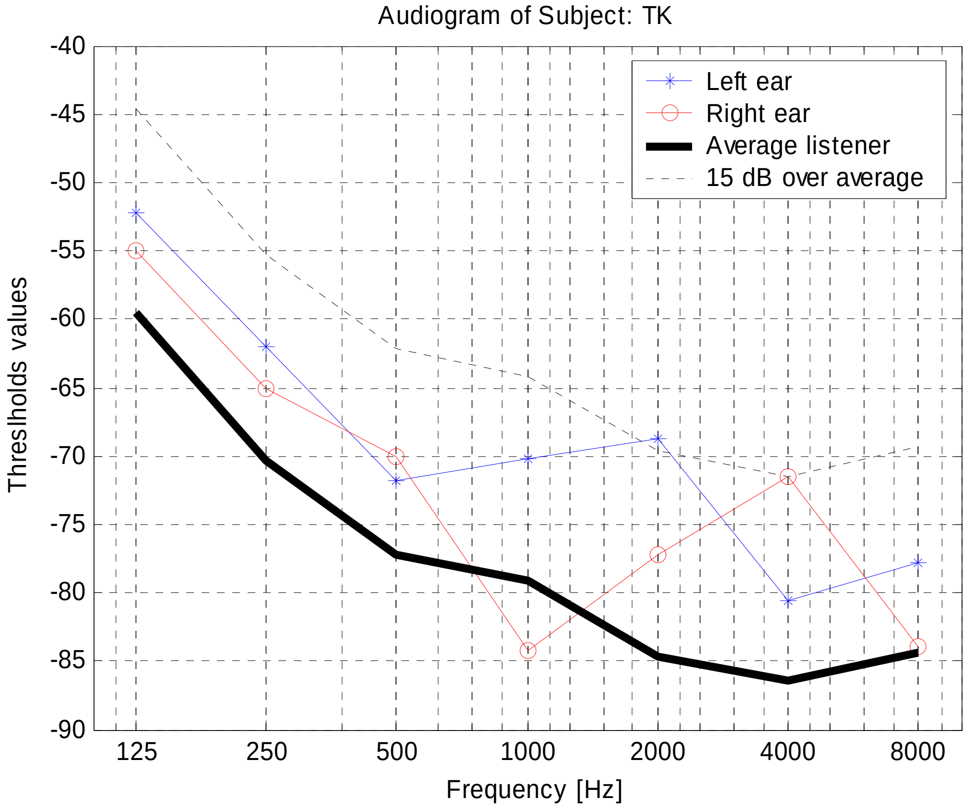


Figure 10.16 Thresholds values of the subject AB, blue line means L - ear, red line means R – ear, black line means average listener and dashed line means15 dB over average. Circles and stars mean measured thresholds on showed frequencies.

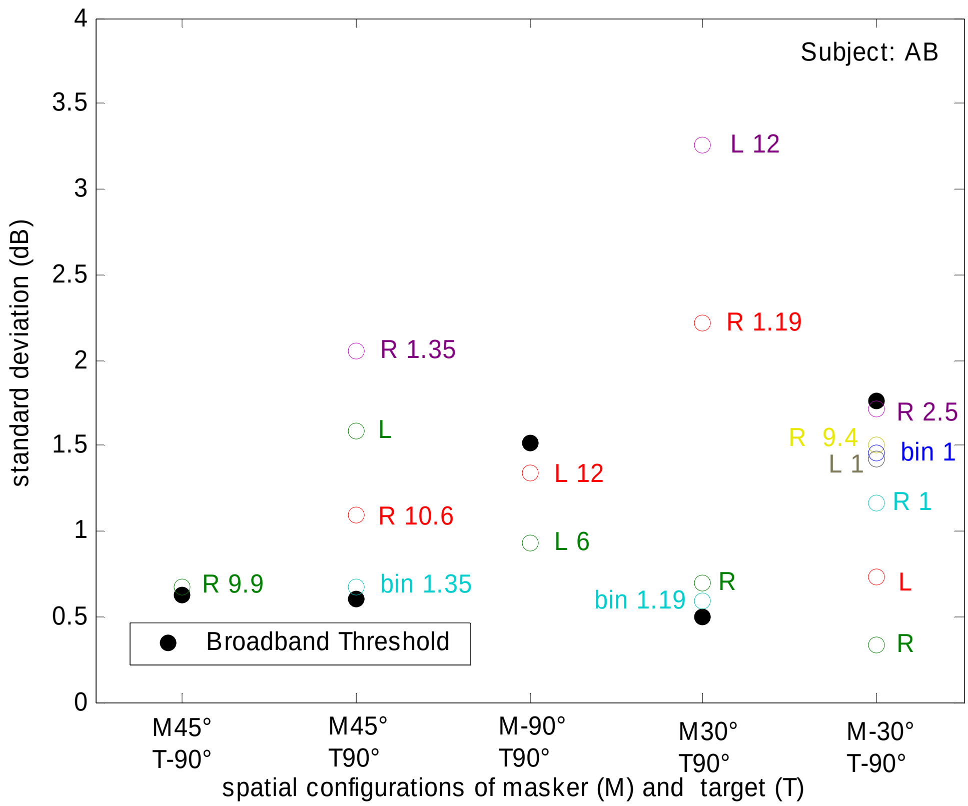


Figure 10.17 Measured standard deviation filtered stimuli for subject AB

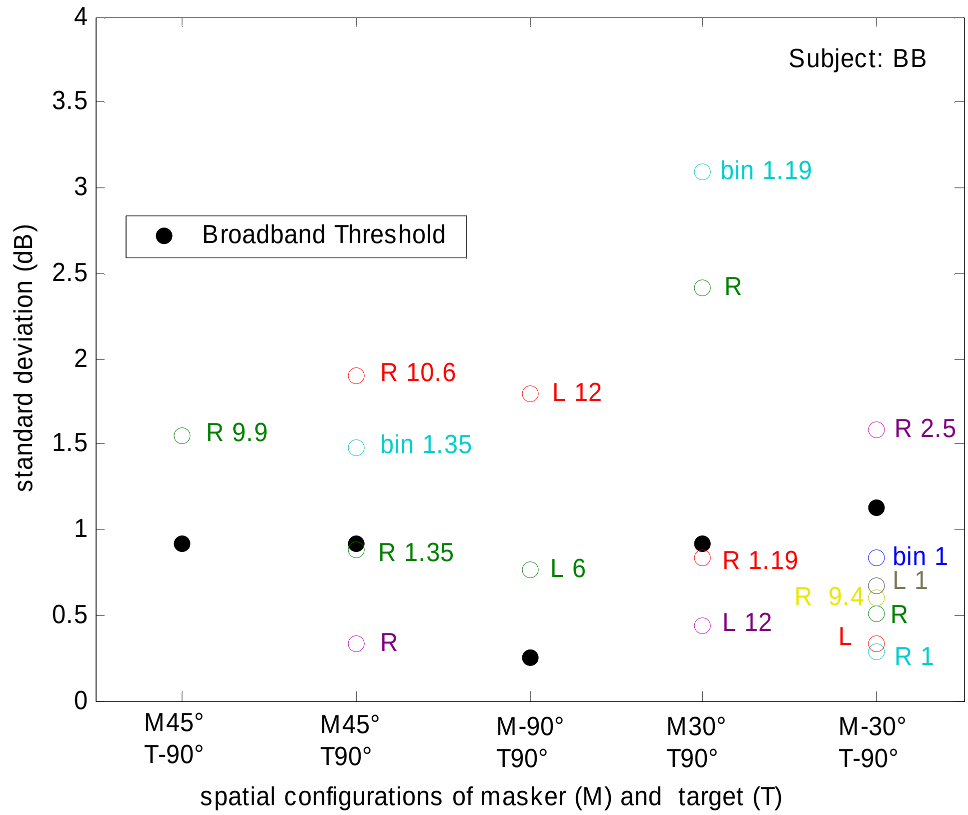


Figure 10.18 Measured standard deviation filtered stimuli for subject BB

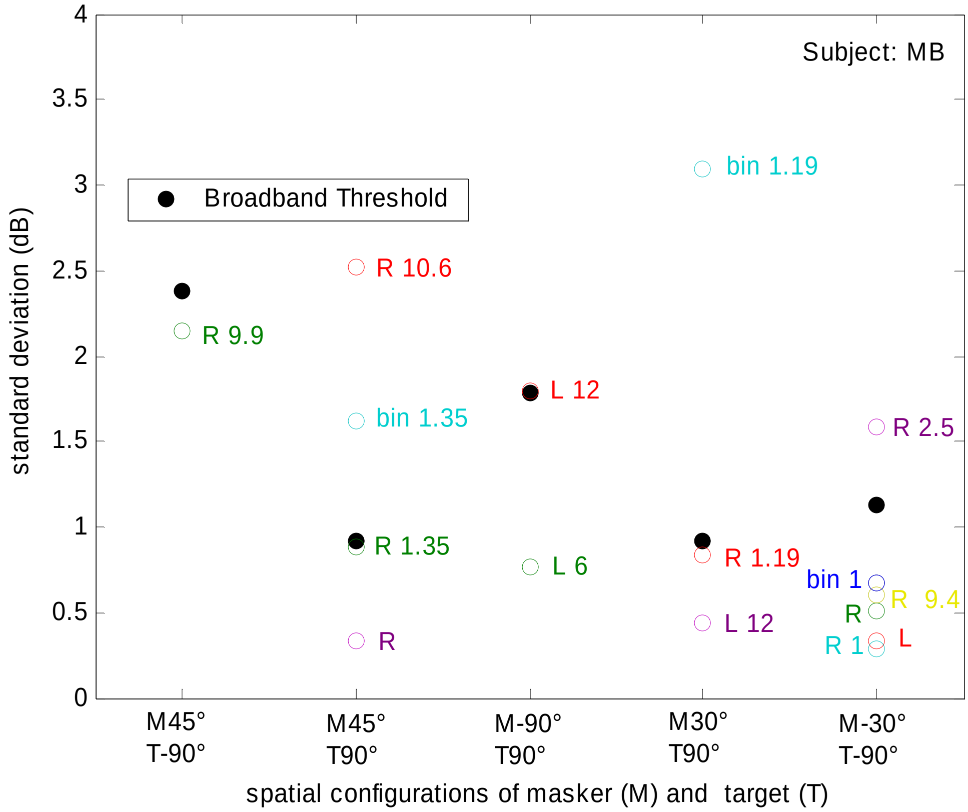


Figure 10.19 Measured standard deviation filtered stimuli for subject MB

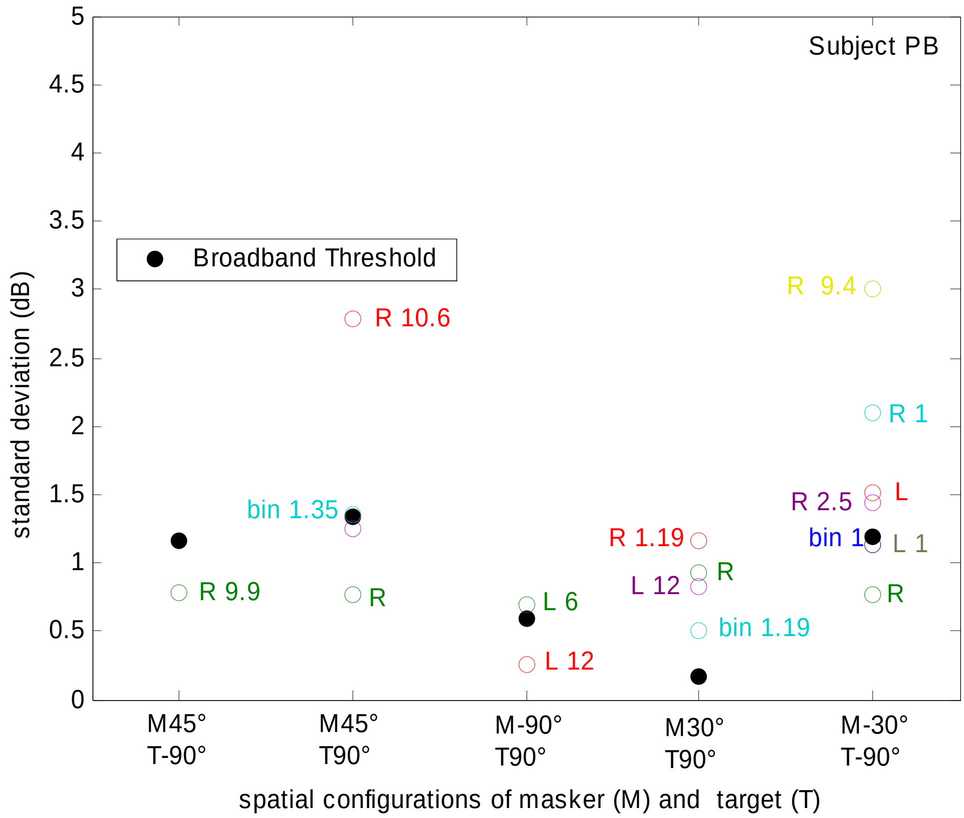


Figure 10.20 Measured standard deviation filtered stimuli for subject PB

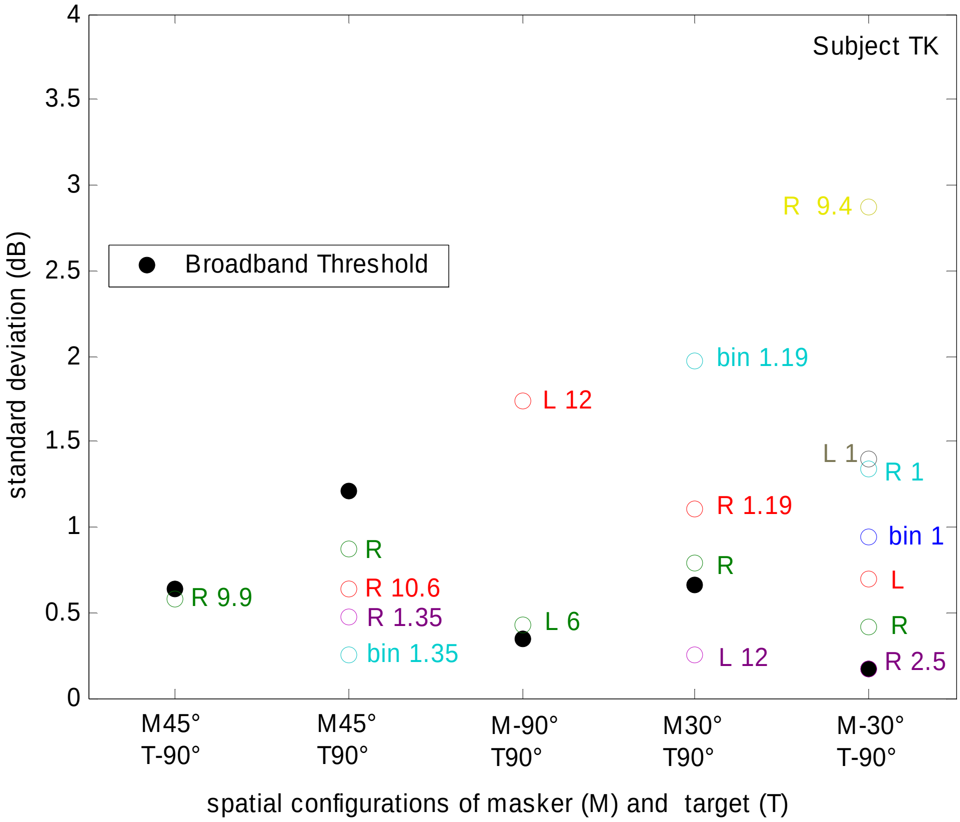


Figure 10.21 Measured standard deviation filtered stimuli for subject TK

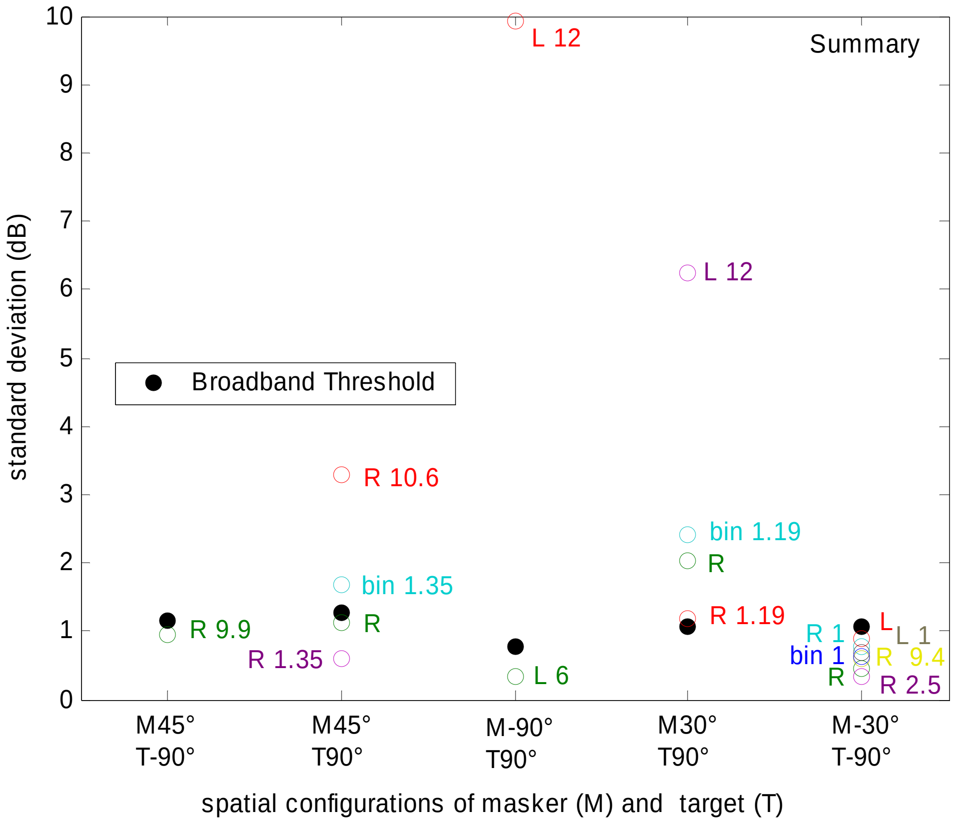


Figure 10.22 Overall standard deviations from all 5 subjects