Usability Evaluation of a Transparent Hearing System

TI-Project Report

Rūdolfs Liepiņš

Supervisor: Dr. Georgios Marentakis

Graz, May 24, 2013

kunst uni graz

institut für elektronische musik und akustik

iem

Abstract

The increasing use of portable audio appliances in public spaces poses limitations in the way people perceive their auditory environment. These can range from inconvenience in cases when someone needs to interact with other individuals, to risks concerning traffic safety. Both auditory masking of outside stimuli and distraction can be considered as possible causes. A solution to this problem could be obtained when environmental sound recorded via microphones attached to the headphones is mixed with the sound the individual is hearing. In the course of this work an already existing such headset was evaluated with respect to localization, and sound quality. A localization experiment was carried out to examine whether such a system affects the ability to localize discrete sound sources. In addition, an evaluation study was performed to see how the system affects the perception of human voice, and if sound quality can be improved by using equalization filters.

Contents

1 Introduction

While headphone listening can provide good sound quality and do not affect or disturb others, the resulting acoustic isolation has certain disadvantages when used in public: The lack of acoustic feedback from the surrounding environment can be potentially dangerous for pedestrians or cyclists in terms of traffic safety. A recent study by Lichenstein et al. [LSAM12] shows that between 2004 and 2011 the number of deaths or injuries of pedestrians wearing headphones has almost tripled. Furthermore, headphone listening is not convenient for interaction with other individuals, e.g., when buying a bus-ticket, getting asked for directions, and so on. A potential solution might be the superposition of sound from the acoustic environment, recorded with binaural microphones attached to the headphones, to the sound people listen through their headphones.This concept has been called *transparent hearing*, but is also found in audio mixed reality (AMR) design proposals, where similar headset configurations are used.
The task of the present work was to objectively and subjectively characterize an already existing headset. This was done by measuring the distortion induced to HRTFs due to the presence of such a headset, and evaluating its usability in terms of sound-source localization and sound quality in subjective experiments.

1.1 Related Work

Augmented Reality (AR), where information is artificially superimposed on the stimulation we receive from the environment, is gaining increased attention in recent time. The advent of costly affordable portable electronic devices with real-time signal-processing capabilities has facilitated the realization of applications, especially in the visual domain. However, the audio aspect of augmented reality remains more or less experimental to date. Terms like *augmented reality audio (ARA)* or *audio mixed reality (AMR)* refer to a design approach where the natural acoustic environment around a user is combined with a spatial auditory display in real time. The basic setup of such a system is similar to the *transparent hearing* idea and consists of microphones mounted on a pair of headphones, providing a binaural recording of the surrounding soundscape. In AMR designs, binaurally encoded acoustic events are used to augment the auditory scenery of the user. Proposed applications for ARA-systems include among others automated tour guides in exhibition halls [Ben95], social networking and navigation for cyclists [MRK^{+}10], or a audio-menu interface for music-playback devices [VAB10]. Another interesting idea is the design of spatial auditory displays for air traffic controllers or other human operators, who have to deal with a huge amount of information - distributing the perceptual load over both visual and auditory (modal) channels can be a benefit [SC98].
Mueller and Karau have designed an augmented audio system with a pair of binaural

microphones mounted on closed (noise-canceling) headphones [MK02]. They call their concept "Transparent Hearing", which is akin to the idea of pseudo acoustic environment. However, instead of adding binaural signals to create a mixed reality scenario, the focus lies on the processed playback of the real environment. One proposed application is to improve conversation quality when listening to music. Playback is stopped, if some prospective dialog partner is detected using infrared distance sensors. The question arises, if a similar setup can be successful without interrupting the music listening.
Fundamental research in the field of ARA has been done at Helsinki University of Technology. The framework for mobile augmented reality audio systems as described above was introduced by Härmä et al. in [HJT$^+$04]. Besides a theoretical description of the acoustic properties of such a system, listening experiments with two different prototype configurations were carried out in order to examine the systems behavior in terms of the users ability to distinguish between a real outside sound source and a virtual one. Furthermore, the grade of externalization, and the effect of front-back confusion for the two source types were evaluated. Amongst others it was shown that under the right circumstances the user cannot distinguish between a virtual acoustic event and an event that comes from the real acoustic environment. This is essential in order to realize an authentic embedding of virtual sounds into a real world situation, but not of big relevance for the scope of this particular work.
Tikander [Tik09a] deals more in-depth with ARA-related problems like acoustic requirements, effects of equalization, head tracking and positioning crucial that are crucial for the real-time alignment of virtual objects with the real surroundings, as well as usability issues. In [Tik09b] a field-study concerning usability in real-life situations was performed. The subjects had to wear an ARA-headset for long period in their daily routines and write down observations. Apart from that, factors like spatial impression, timbre and location accuracy were evaluated under lab conditions. The approach was to judge those factors by the experimenter shouting out loud and making finger snaps at different locations in the room. Since the aspect of event localization is of crucial importance for the *transparent hearing* concept, it should be investigated it in more detail. It is of particular interest if such a system has a significant influence on the ability to localize surrounding events, compared to conventional headphone listening. Localization with earmuffs and military hearing protectors has been evaluated in [BKES03]. It was shown that bandwidth might be a limiting factor of localization performance using electronic hearing protection. Algazi [VRAT99] suggests, that the spectral cues related to localization are more or less preserved in HRTFs measured slightly outside a blocked ear canal. However, it is unknown how they perform in practical realizations. In this particular work a simple prototype for *transparent hearing* was built and evaluated against it's localization performance in a listening experiment. In particular, it was tried to treat localization aspect isolated from other sound quality aspects. Subsequently, a subjective evaluation of sound quality was carried out to see if music listeners can benefit from such a system

in situations where they interact with others.

1.2 Theoretical Background

1.2.1 Sound Localization

The interaction of several factors enables the human hearing-system to localize sound events. When building a transparent hearing device, we have to make sure, that the cues relevant to the localization capability are preserved in the best way possible.
A big part of the localization process can be explained by the so called duplex theory [Car96]. Two factors are responsible for localization in the horizontal plane, namely inter-aural time differences (ITDs) and inter-aural level differences(ILDs). A sound-wave will reach the left and right eardrum at different time instances, when the source location lies outside the median plane. This offset is described by the ITDs and takes effect in the frequency band between 80 Hz and 1500 Hz approximately. Above 800 Hz, the wavelength of the incoming wave is smaller than twice the distance between the ears making the phase-information ambiguous. Starting from around 300 Hz , the ILDs take effect, and dominate at high frequencies above 1600 Hz where ITDs become ambiguous as the wavelength is much smaller than the size of the head; at this frequency range, the shadowing-effect yields different intensity levels at both ears, that are used to determine sound source azimuth. For lower frequencies the ILDs are irrelevant due to diffraction of the sound-wave around the head.
However, evaluation of both ITD and ILD alone are not sufficient for proper localization of arbitrary directions. Considering the horizontal plane only, there is an ambiguity between front and back: The distance between a source and both ears will be the same whether it is located behind or in front of the head. In three-dimensional space, the so called cone-of-confusion determines the region of ambiguous directions around the head. Head-movements are capable of resolving this ambiguity to a certain degree. Head related transfer functions (HRTF) facilitate localization of elevated sources and contributes to eliminating the mentioned ambiguities. The frequency dependent energy profile of a sound reaching the eardrum is altered by the human's head, torso and concha(outer ear) for every source direction, yielding a certain transfer function. The shape such a HRTF is different for different directions and is used by the brain to determine elevation and perform front-back discrimination. It is of interest, to what degree the role of HRTFs on enabling localization can be preserved when using a *transparent hearing* setup.

1.2.2 Pseudo Acoustic Environment

The goal of the pseudo acoustic environment idea is to achieve a representation of the real acoustic environment as accurate as possible. Figure 1 shows the signal path both for unblocked ears and a normal earphone-listening situation. The spectral HRTF cues are distorted by the presence of the earplug, since the outer-ear geometry is altered and only low frequency components leak around the earplug to the ear canal. When an earphone is placed on the ear, the ear canal resonance behavior is different from the open ear case. For the open ear case, the ear canal can be seen as a quarter wavelength resonator, where the first resonance is occurring between 2kHz and 4kHz. When the ear is blocked the ear canal behaves as a half wavelength resonator and the first resonance is shifted up to around 5kHz-10kHz. [Tik09a] As a consequence, the transfer function between headphone speaker and ear drum should be equalized in order to allow a more natural perception. Unfortunately, it has been shown that the ETF curves are highly individual. Thus, it would be necessary to measure the ETFs for every individual in order to achieve perfect equalization. In section 4.1 the equalization was implemented using generic ETFs measured on a dummy head setup and was used to evaluate to what extent a simple ETF equalization can improve the sound quality of the system. According to

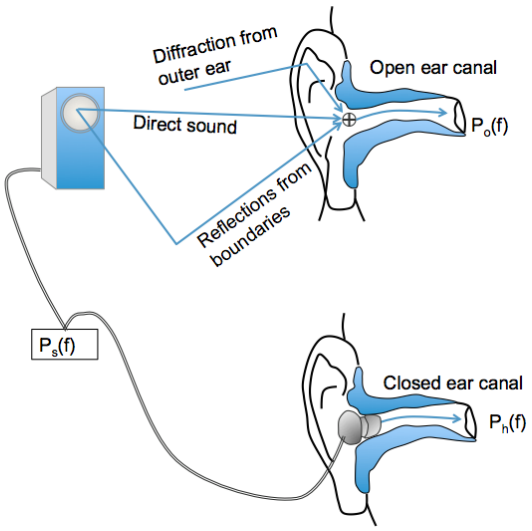


Figure 1: Signal transmission with unblocked ears and when using headphones. [Tik09a]

Härmä, in the pseudo acoustic case the signal path for the given headset configuration can be characterized as

$$y_{ear}(z) = [H_t(z) \cdot H_m(z) + E(z)] \cdot x(z). \quad (1)$$

The transfer function describing the signal path consists of the earphone transfer function (ETF) $H_t(z)$, a microphone transfer function $H_m(z)$ and additional direct sound leakage $E(z)$. Some leakage from the direct path is inevitable for headphone listening, even for insert type headphones. While higher frequencies are attenuated quite well, low frequency components tend to leak around the earphone. Since the leakage has no delay, it is difficult to design a digital filter for compensation due to latency. Apart from keeping the latency as low as possible, coloration caused by direct sound leakage can be masked by choosing an appropriate ratio between the physical attenuation of the earplug and the signal level [HJT+04]. The transfer function $H_m(z)$ describes the signal path between a sound source and the microphone in the headset. It is basically similar to an HRTF when the headset is mounted on the ear. However, since the microphone is located outside the ear canal, certain discrepancies have to be expected. Apart from that, the frequency characteristics and distortion effects of the microphone itself contributes to $H_m(z)$, since no ideal transducer can be assumed.

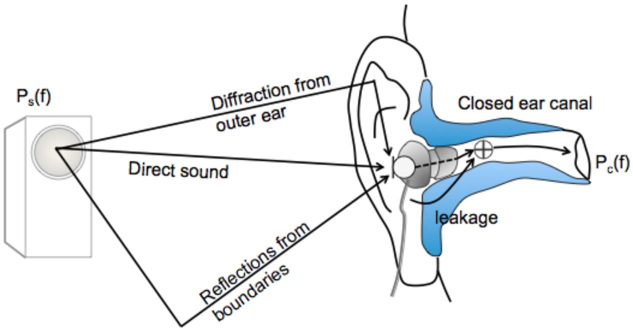


Figure 2: Signal transmission using a *transparent hearing* setup. [Tik09a]

1.3 Evaluated Setup

A simple *transparent hearing* setup was implemented in order to provide the basic functionality of a transparent hearing environment. The setup consists of two main elements: Headphones combined with binaural microphones, and a mixing device. In [HJT+04] the ARA-mixer has been implemented in analog hardware in order to allow mobile usage and keep the system latency to a minimum. For this project however, a more straight forward implementation was used, since acquiring the necessary knowledge in analog circuit design would lie beyond the scope of the work. Since it was not possible to spot a portable appliance on the market that would meet the desired requirements without blowing the budget at the time, the decision was made to carry out the evaluation tasks in a fixed environment. Furthermore, there was no previous knowledge of the headset's acoustic properties, as we used a product available on the market rather than developing an own prototype.

1.3.1 Headset

The headset of choice was the CS-10em system by Roland(see 4). It consists of dynamic in-ear headphones with attached omnidirectional binaural electret-condenser mi-

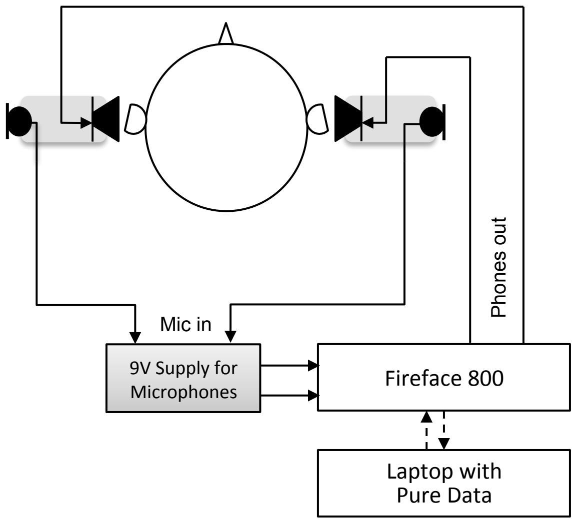


Figure 3: Evaluated transparent hearing setup

crophones. The microphones require a supply voltage of 2V to 10 V, which is provided

Figure 4: Roland CS-10em

by a portable audio recorder by default. Since the headset was run in combination with an audio-interface and a notebook computer without supply voltage capability, an external circuitry had to be implemented in order to provide the necessary voltage using a 9V block-type battery. Audio processing was carried out on a notebook computer. Connecting the headphones to the computer's output jack directly led to significant background noise on the headphone speakers. It turned out to be a reported issue with the particular notebook model (MacBookPro late 2007), when using headphones with low impedance. When using the sound-cards pre-amplifiers the background noise could be avoided successfully.

1.3.2 Mixing Device

The mixer was implemented in Pure Data Vanilla running on a MacBook Pro (late 2007 model). The overall latency (between direct sound and playback) of 10.5 ms turned out to be small enough to provide an acceptable listening experience without noticeable comb filter effects. In order to suppress the above mentioned background noise, a RME Fireface 800 audio interface was used for audio I/O.

2 Acoustic Measurements

As a starting point of the evaluation study it was necessary to gain some knowledge about the acoustic properties of the configuration under test. It was decided to carry out preliminary impulse response measurements using a dummy head at several source positions under different test conditions. In a next step, it was evaluated how well localization cues like ILD, ITD and HRTF-curves are preserved when wearing the headset.

2.1 Impulse Response Measurements

The impulse responses were obtained in the azimuth plane at steps of 10° and at 16 elevation positions, yielding a total amount of 576 source positions. They were measured for both left and right channels of the dummy head system. The three measurement conditions were chosen to be the same as in the subsequent listening test:

– **Earphones mounted**, *transparent hearing* **off** .
– **Earphones mounted** , *transparent hearing* **on**.
– **Dummy-Head only**, as a control condition.

2.1.1 Setup

Figure 5: Measurement Setup for Impulse Response Measurements

The main elements of the measurement setup were a Brüel & Kjær Head and Torso Simulator (HATS) system and a semi circular loudspeaker array built by Boris Müller at IEM. The array depicted in 5 consisted of 16 drop-shaped loudspeakers placed spanning the elevation form $-90°$ to $90°$ in steps of $11.25°$. The radius of the structure is considered to be 1 m. The correct position of the dummy head had to be determined with the aid of a plumb-bob and a laser rangefinder, however, absolute precision could not be attained due to the bulkiness of the setup. The loudspeakers were driven by two Behringer Ultragain Pro-8 preamplifiers, connected to a Debian Linux workstation via ADAT. The HATS dummy-head was placed on a ethernet-controlled turntable; its microphone outputs were connected with the computer workstation via a Brüel & Kjær 2629 Nexus Conditioning Amplifier for recording. All control tasks like excitation signal playback, recording, turntable movement were executed by a single Pure Data patch developed by Franz Zotter.

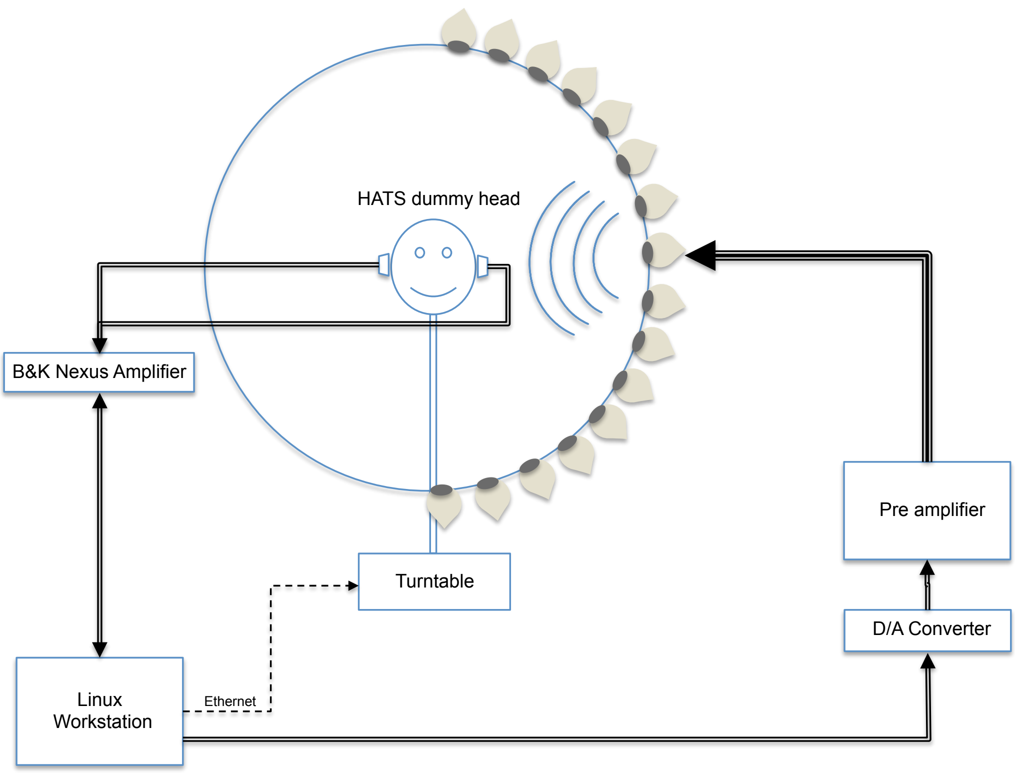


Figure 6: simplified signal flow of the measurement setup

2.1.2 Measurements

The impulse response measurements were carried out using the multiple exponential sweep method, as described in [MBL07]. It has the advantage of being less time-consuming than other methods, which is an important criterion when a large amount of data is to be acquired. Since the procedure is based on the standard exponential sweep method by Farina [Far00], its robustness concerning weak system non-linearities is maintained.

2.2 Analysis

The acquired binaural impulse responses were examined with regard to localization cues, namely interaural level differences, interaural time differences and spectral cues. When the system is turned on, an overall amplification level of approximately 25 dB relative to the off-case was observed for all directions. While very low frequencies show more or less the same level for both cases, higher frequencies of the direct path are attenuated dramatically.

2.2.1 ILD and ITD

Based on the impulse response data, estimates of both ILDs and ITDs were computed in a straightforward manner. Level differences for all azimuth angles were determined by dividing the particular magnitude spectra (Eq.2). The time lag between left and right impulse responses, i.e. the lag between the first peaks of the broadband time-domain responses, was used as an estimate for the ITD of a certain direction. Figure 7 shows the resulting curves for all azimuth directions. The ILD cues are well preserved when the system is on, whereas for the blocked ears case the cues show notable impairment.

$$\Delta L(f) = 20 \log_{10} \left| \frac{X_{right}(f)}{X_{left}(f)} \right| \tag{2}$$

The ITDs on the other hand show the same behavior, apart from a small offset because the microphones are located away from the eardrum, for all configurations, which is to be expected since the direct sound is not delayed by the presence of earphones. Furthermore, any latency of the *transparent hearing* system can be assumed the same for both channels.

2.2.2 Spectral Cues and Quantification Error

It was of particular interest to see how the spectral localization cues are affected when the ear canal is blocked by earphones and when *transparent hearing* is turned on. Inspection of the HRTF curves shows that both system on and system off heavily change the shape of the transfer functions. As an example, figure 8 depicts the HRTF spectra for both ears for a single direction. For the blocked ears only case the magnitude drops massively at higher frequencies, where the important cues are located. Apart from that, the shape

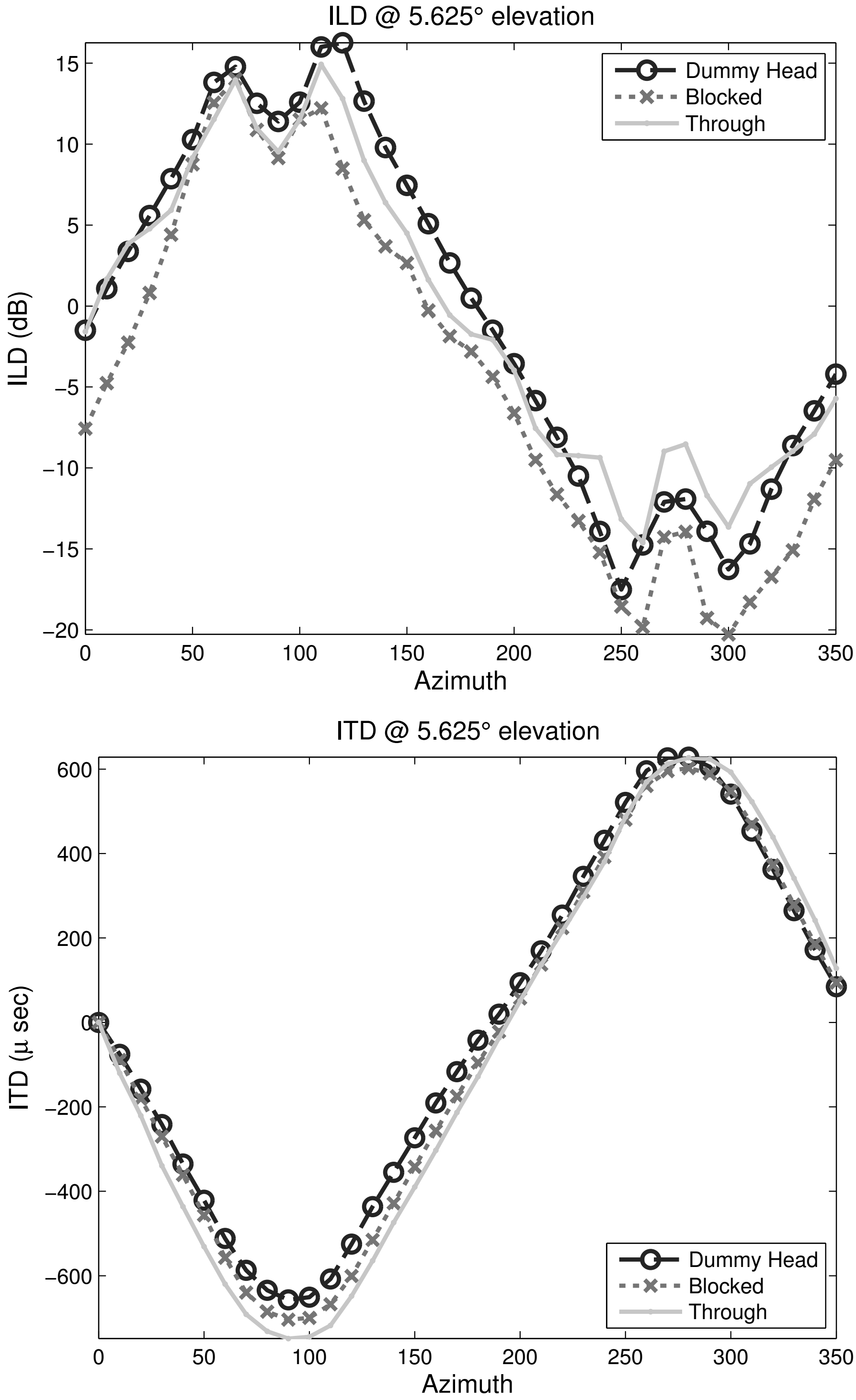


Figure 7: Upper plot: ILD for all azimuth values and different configurations, lower plot: ITD for all azimuth values and different configurations

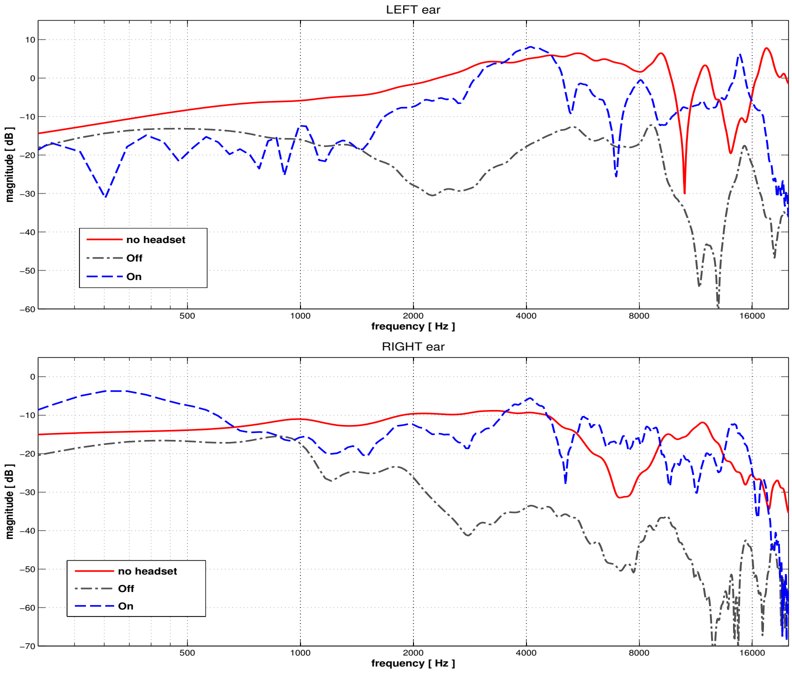


Figure 8: (Head Related) Transfer Functions of all conditions. azimuth: 60°, elevation: 5°

of the curve is different from the open ears case. When the system is turned on, higher frequencies are attenuated less, but the transfer function is, again, altered heavily. This suggests that the microphone/headphone system's own frequency response is not flat enough to preserve localization cues in an appropriate manner.

Since 576 impulse responses in total for each ear were measured, it was desired to represent the data in a more compact way for further investigation. In particular, a way had to be found to characterize the amount of mismatch between the HRTF spectra of individual directions. Lemaire et al. [LCB+05] have proposed the "quantification error" as a metric for spectral distortion of HRTFs, as described in Eq. 3.

$$E_{\phi,\theta} = \sum_{j=1}^{N} \left| 20 \log_{10} \hat{h}_{\phi,\theta}(f_j) - 20 \log_{10} h_{\phi,\theta}(f_j) \right| \text{ [dB]} \tag{3}$$

Figure 9 visualizes the quantification error for all azimuth and elevation angles, and both ears. HRTF spectra were normalized to each other before error computation. Certain differences between left ear and right ear are apparent for both configurations, the general structure of the plots, however,is similar (but mirror-inverted). Most likely the disparity of the channels is due to inaccurate placement of the earplugs on the dummy head, or, some biased error in the measurement chain. For the blocked ears case the difference between individual directions in terms of error-level are less pronounced compared to "system on". In general, "system on" has a larger quantification error for a multitude of directions, in particular for contralateral directions and below the horizontal plane.

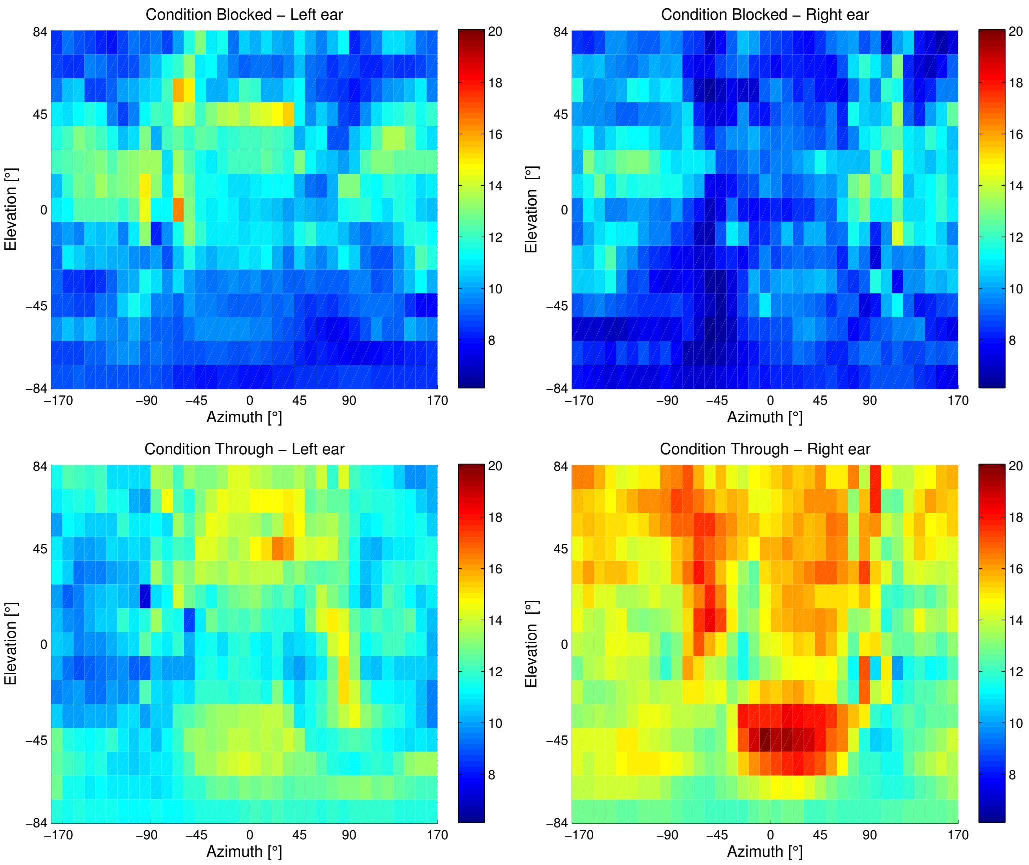


Figure 9: Quantification error in dB for all directions.

3 Localization Experiment

Based on the foregoing observations a listening test was designed and carried out in order to see how localization is affected by the system when utilized by human test subjects. As stated above, the idea of the transparent hearing system is to improve the ability to locate sonic events more effectively, while listening to music in a real world scenario. However, in order to examine the system's performance in a strict localization context it was decided not to include a music signal to keep things more simple. Addressing the interaction of localization cues of two different sound sources states a psychoacoustic problem (for example, see [GG96]) that was outside the scope of this work. Instead, the influence of additional music was investigated in the course of the subsequent sound quality evaluation (section 4). The decision was made to introduce another test condition using a pair of open (acoustically transparent) headphones to see how they perform in comparison with the system and if the transparent hearing idea brings any benefit compared to such headphones.

3.1 Test Setup

The test environment was set up in the Experimental Studio of IEM, a small room with the possibility to rig loudspeakers on the ceiling. Since the room had to be shared with other projects, it was not possible to arrange the setup in the center but had to be placed in one corner of the studio.

3.1.0.1 Excitation Signal Playback From the analysis of the measured impulse responses it was apparent, that the localization properties are more or less the same for both ears. Thus the decision was made to design the experiment setup only for one hemisphere around the test position. This brought the benefit to be able to evaluate more different position angles, since the number of channels was restricted to 16 in order to keep the duration of the experiment in a tolerable manner. The loudspeakers used for the experiment were designed and built by Sebastian Blamberger [Bla12]. Because of their flexible mounting possibilities and their compact size they proved to be well suited for stimulus playback. They were arranged around a central spot as depicted in Figure 10. The exact positions were determined once again using a plumb-bob and a laser rangefinder. Speakers corresponding to -45° elevation and 0° elevation respectively were attached to ordinary microphone stands, the upper speakers were mounted on a pre-assembled construction that was affixed to the ceiling of the room. In order to prevent visual feedback regarding the actual loudspeaker positions acoustically transparent screens were placed between the listening position and the loudspeakers.

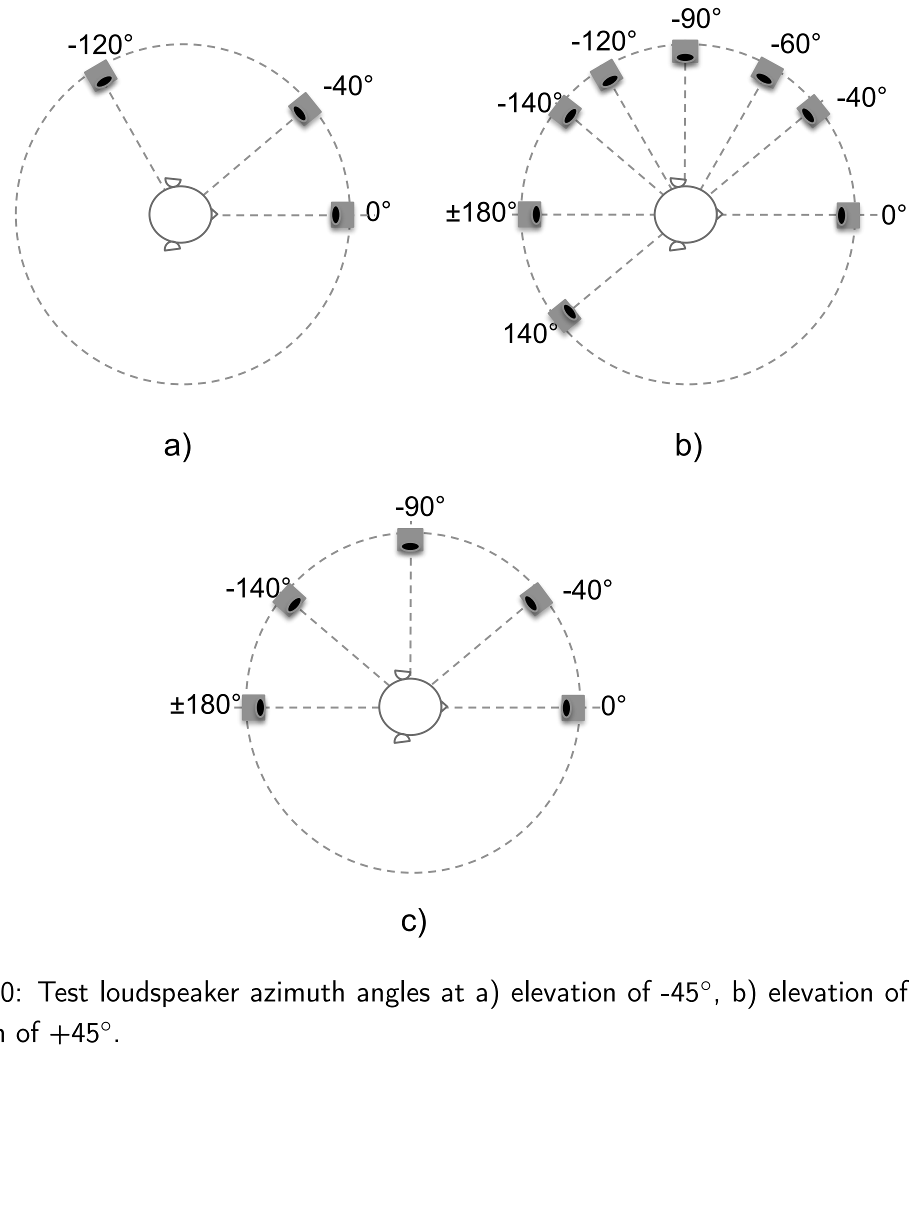


Figure 10: Test loudspeaker azimuth angles at a) elevation of -45°, b) elevation of 0° and c) elevation of +45°.

The excitation signal playback was controlled via a Pure Data patch on a notebook computer. The speakers were driven by a custom made, 48 channel amplifier connected to a set of RME ADI 8 AD/DA converters, which in turn were connected to a RME Fireface 800 audio interface. The individual loudspeaker output levels was calibrated to assure a constant excitation signal level from all directions. A level meter was placed in head position, then the level was adjusted manually within the Fireface Mixer.

3.1.0.2 Control Software The control scheme for the experiment was implemented in a single Pure Data patch, as depicted in figure 11. It has three main purposes: excitation signal playback control, communication with the GUI and recording the test subjects response data. The playback order of the excitation signal has been generated using a randomization algorithm in MATLAB and saved in a text-file. It is ensured that for each repetition every speaker is triggered only once. The structure of such a random file is exemplified in figure 12. An array of zeros with a "one" corresponding to the speaker ID is used to control the gating behavior of a `[multiline~]` object. Since the GUI was implemented in Processing (see next section), it communicates with the control patch via Open Sound Control protocol. Once the test person enters a response, the next trial is triggered after 1.5 seconds. At the same time, the response data is written to a text file in order to store the result in terms of azimuth and elevation.

3.1.0.3 User Interface Designing a useful response input interface for the particular task is quite a challenge, since we have to deal with three-dimensional space. It is very important to provide an intuitive way of mapping the perceived directions, in order to keep the bias caused by the user interface as low as possible. First of all, the interface has to provide sufficient accuracy concerning the input. This implies a proper technical realization as well as some kind of a (typically visual) feedback mechanism to ensure the test subject is aware of what is actually recorded as an answer. Furthermore, it is important to keep the input task as simple as possible in order not to overstrain the participant.

The input data consists of elevation and azimuth values, thus an obvious idea was to use a three-dimensional representation for input. One solution could have been a pointing device, where the test person would point to the perceived direction. An experimental setup was implemented using a Nintendo Wiimote controller in combination with Pure Data.

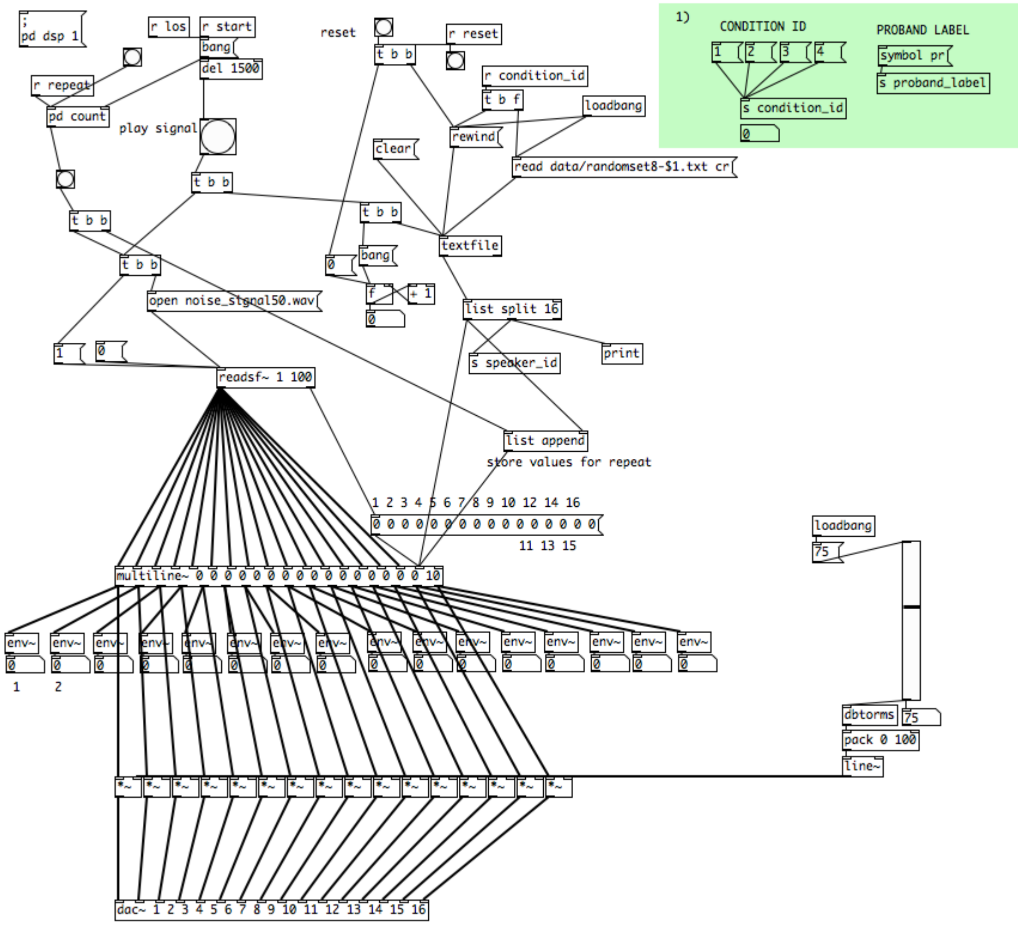


Figure 11: Pure Data patch for stimuli control.

The Wiimote features a pitch, roll and yaw sensor and can be connected to a computer via bluetooth. Unfortunately the yaw-angle (corresponding to azimuth) detection turned out to be unreliable and imprecise. One possible solution could have been the use of an optical tracking system like VICON, but since it was not possible to utilize the IEM Cube this was not an option. Experimenting with a mobile IR camera system revealed that the camera fans generate massive background noise, making it useless for the particular purpose. A further drawback emerged in conjunction with the pointing procedure itself: While it does the job quite well when pointing to a position in front of you, the precision dramatically decreases for targets on the back. Another 3D-concept could be a spherical touch-sensitive surface, as proposed for example in [BWB08]. An input system like this

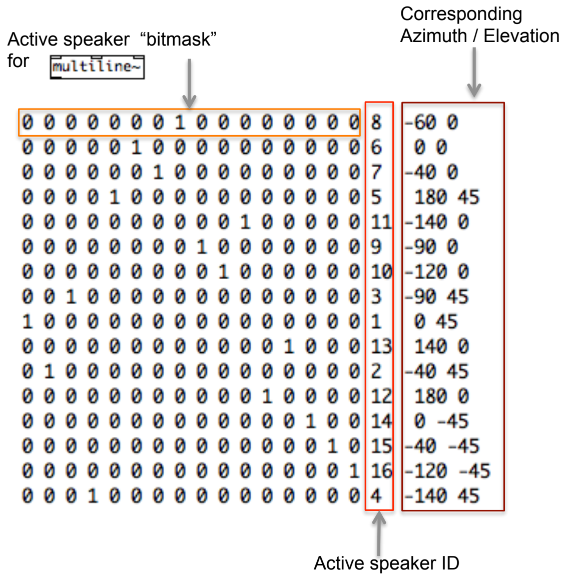


Figure 12: dataformat of speaker control messages

would have the ability to map spherical coordinates directly on a physical metaphor, representing a fairly intuitive interaction method. Nevertheless, due to the complexity of such an implementation, it was decided to settle for an ordinary GUI with mouse input. The GUI was implemented in Processing, a java-based programming language convenient for fast and straightforward graphics programming.

Since the input data format consists of two variables (i.e angles), it turned out to be easier to deal with them separately rather than creating some spherical GUI element. Not only it would be quite demanding to implement that in a reasonable way, but also a three-dimensional representation on a two-dimensional screen does not provide the desired input accuracy. Therefore, the final GUI solution consisted of two circular elements, representing the azimuth and the elevation plane, respectively. The user had to place a marker on each circle to indicate the perceived angle (see Figure 13).

3.2 Participants

A total of 12 test subjects participated in the localization experiment. They were aged between 19 and 45, five of them female, and seven male. Except from one, none of them had previous experience with localization experiments.

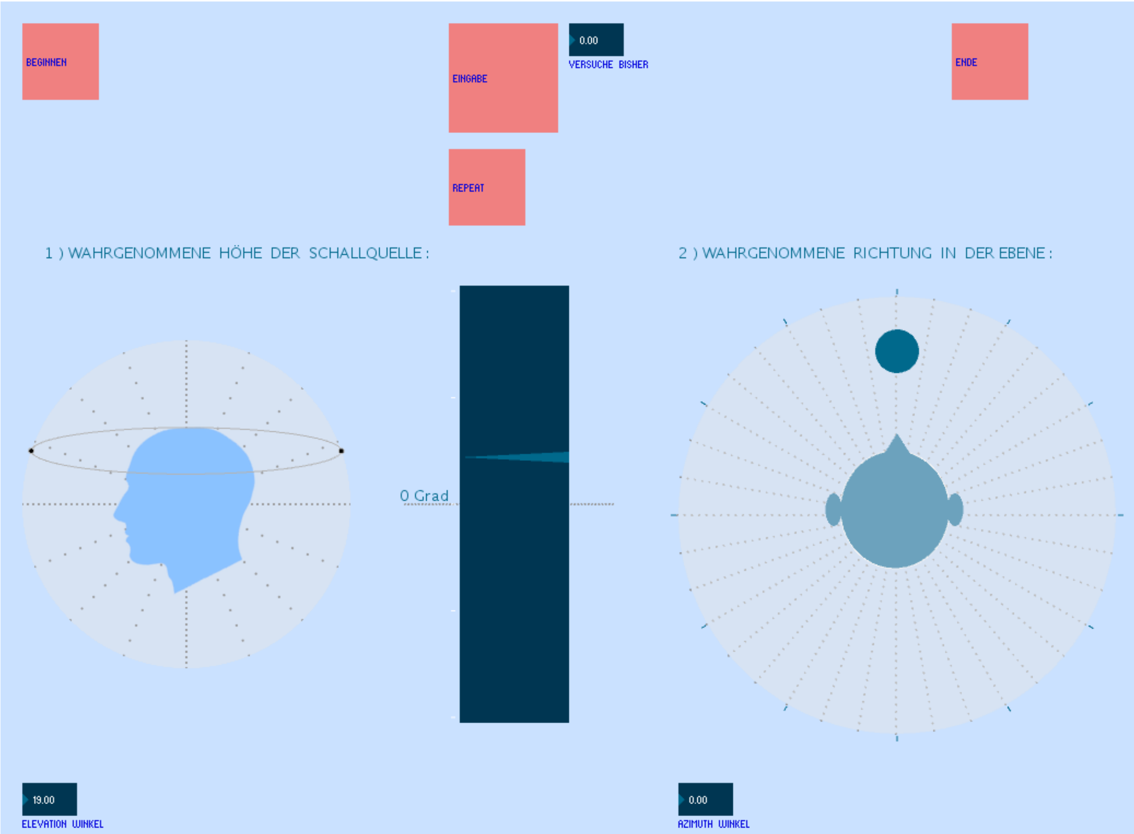


Figure 13: final version of the user input interface. The test person has to enter the perceived stimulus location by placing an indicator on planes representing elevation and azimuth angles respectively.

3.3 Test Procedure

Test subjects were seated within the loudspeaker array. Since the participants were of unequal body height, they were placed on a vertically adjustable chair to be able to set the correct ear level. Furthermore, they were instructed to keep their head straight by focusing a spot in front of them.
The task had to be accomplished for four different conditions:
Condition 1 - without earplugs
Condition 2 - wearing earplugs, *transparent hearing* off
Condition 3 - wearing earplugs, *transparent hearing* on
Condition 4 - wearing open headphones
The time of each session was restricted to roughly one hour, since the test subjects participated voluntarily. Furthermore, due to the monotonous nature of the task it could not be guaranteed that the subjects would be able to keep focused appropriately for a longer period of time. As a consequence, it was decided to carry out 5 repetitions for each condition. Each repetition consisted of 16 stimuli played back from random source positions. Since, the number of participants was 12, in total 60 trials were completed for each loudspeaker position and each test condition. The running-order of conditions was shuffled by means of a latin square of 4th order to prevent bias.
Each participant was briefed about the task and had to perform 16 trials in advance in order to get familiar with the GUI and gain some routine concerning the task. As the test procedure a broadband white noise stimulus of 0.5 s length was played back through a randomly chosen loudspeaker. The participant then had to adjust the guessed azimuth and elevation respectively on the GUI using a computer mouse. It was possible to repeat each stimulus once, but the test subjects were instructed to make use of this possibility only in cases when they lost attention or forgot their guess while operating the input interface. After a block of 96 trials (i.e. after one test condition) participants were advised to take a break of five minutes. Figure 14 shows the listening test setup.

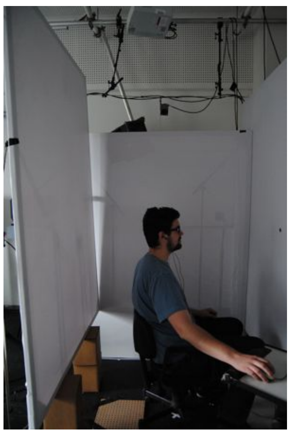


Figure 14: Localization test setup.

3.4 Results

3.4.1 Spherical Data Analysis

Since the localization data was acquired in polar coordinates(azimuth θ, elevation ϕ), it can be beneficial to use spherical data analysis and visualization methods for further examination. Instead of dealing with azimuth and elevation separately, the data is mapped on unit-sphere, where it's origin represents the listener's head position. For further localization error analysis it was assumed that the data follows a Kent distribution which models asymmetric data on a sphere. Unlike the Fisher distribution which assumes rotational symmetry and unimodality of the data, the Kent distribution gives information about the directions where a dataset has its widest and smallest variance respectively [LC98]. Both distribution types were observed on the dataset depending on the source position, but since the Kent distribution is a generalization of the Fisher distribution it effectively models symmetrical data as well.

3.4.1.1 Judgement Centroid The direction of the judgement centroid vector describes the average direction of all judgements in a dataset from the origin, or head position. It is computed by first transforming the polar coordinates to Cartesian coordinates:

$x_i = \sin\theta_i \cos\phi_i, \quad y_i = \sin\theta_i \sin\phi_i, \quad z_i = \cos\phi_i$ where $i = 1...n$th datapoint
The obtained directional cosines then are summed up:

$$S_x = \sum_{i=1}^{n} x_i, \quad S_y = \sum_{i=1}^{n} y_i, S_z = \sum_{i=1}^{n} z_i \tag{4}$$

The resulting length of the centroid is calculated as:

$$R = \sqrt{S_x^2 + S_y^2 + S_z^2} \tag{5}$$

Since the unit sphere only allows vectors of length 1, the resulting directional cosines are computed as

$$\bar{x} = \frac{S_x}{R}, \quad \bar{y} = \frac{S_y}{R}, \quad \bar{z} = \frac{S_z}{R} \tag{6}$$

and finally, transformed back to polar coordinates:

$$\bar{\theta} = \arccos(\bar{z}), \bar{\phi} = \arctan(\frac{\bar{y}}{\bar{x}}) \tag{7}$$

3.4.1.2 Judgement Spread The length of the judgement centroid R (eq. 5) is ranged between 0 and n and can be interpreted as a measure of dispersion. Large values correspond to low dispersion, and values near 0 correspond to a more or less uniform distribution of the data on the unit sphere. Figure 15 visualizes the idea of data dispersion around the centroid.

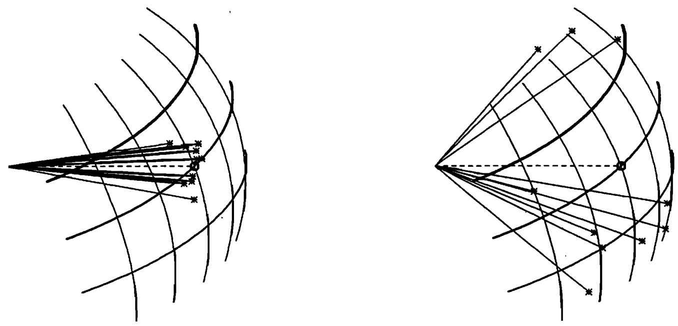


Figure 15: Left: low dispersion of data, Right: high dispersion of data.

3.4.1.3 Spherical Correlation Coefficient The spherical correlation coefficient of the perceived and the actual (target) locations is a useful metric for the overall localization performance [CLH97]. It is computed as follows:

$$S_{XX^*} = \det(\sum_{i=1}^{n} X_i X_i^{*T}) \tag{8}$$

,

$$S_{XX} = \det(\sum_{i=1}^{n} X_i X_i^{T}) \tag{9}$$

$$S_{X^*X^*} = \det(\sum_{i=1}^{n} X_i^* X_i^{*T}) \tag{10}$$

$$\rho = \frac{S_{XX^*}}{\sqrt{S_{XX} S_{X^*X^*}}} \tag{11}$$

where
X_i ... $n \times 3$ matrix of the direction cosines of n *perceived* locations, and
X_i^* ... $n \times 3$ matrix of the direction cosines of n *actual* locations

The value of ρ lies in the range between -1 and 1. If the actual and perceived location-sets can be transformed to the other by a rotation, the value is 1, whereas a value of -1 indicates that one set can be seen as a reflection of the other.

So called front-back confusion errors occur where a subject correctly identifies an azimuth angle with respect to the median plane, but detects the target in the wrong hemisphere. The same applies to up-down confusions with respect to the horizontal plane. The typical way to approach [LC98] the problem is to examine the confusion errors first, then remove them from the data set to examine the localization error related to individual target locations. It was chosen to analyze individual locations in terms of error centroid and judgement spread and the overall localization performance using the spherical correlation coefficient.

3.4.2 Source Confusion Error

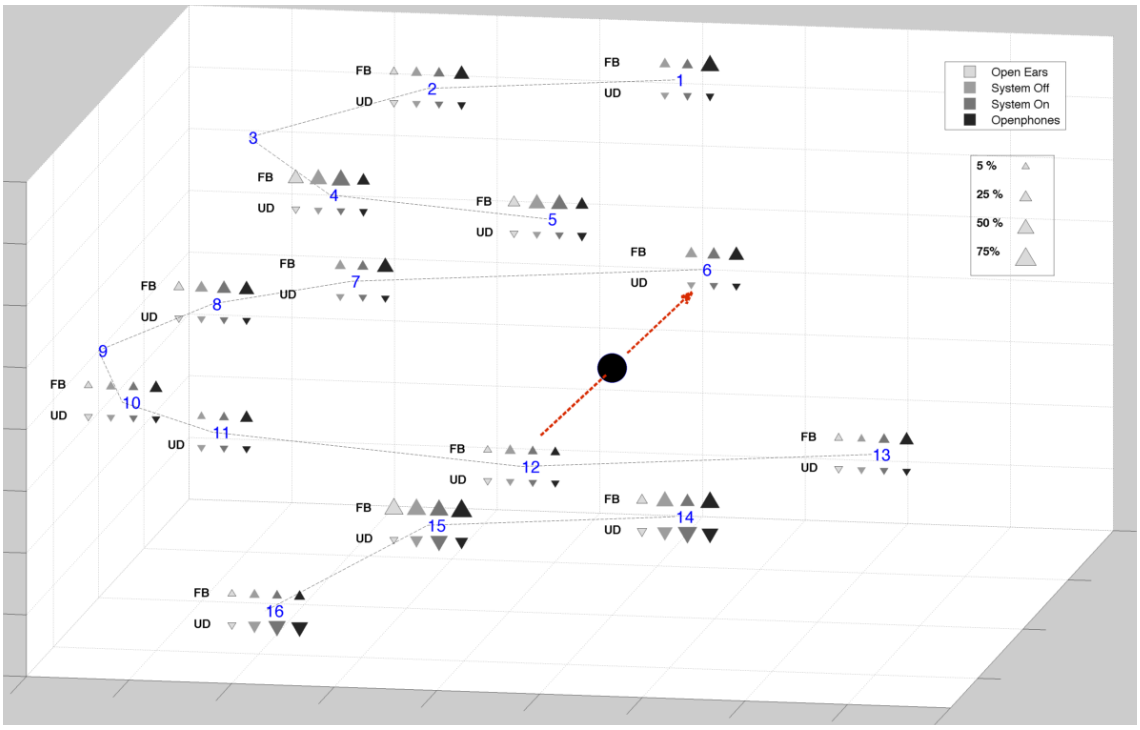


Figure 16: Front/back- and up/down-confusions for all directions and systems.

Overall, front-back and up-down confusions increased when comparing the unblocked condition to the others and was highest for open headphone condition. Even with un-blocked ear-canals front-back confusion rate was high. The problem was localized in five out of the 16 directions, namely directions 4,5,8,14 and 15 where a rate of 45%,21%,13%,18% and 65% respectively was observed. In the other directions, this was never more than 3%, a rate that is considered normal [CLH97]. Figure 16 illustrates the amount of confusions for single directions. The rate of up-down confusions in the unblocked condition was 0% all directions, save 13 (1%) and 14 (10%). When considering front-back confusions, performance worsened with blocked ear-canal for all locations (see Table 1). At first inspection, there is no difference between the blocked and the transparent conditions, however, the apparent similarity of the two systems is largely due to individual differences. In particular, 5 out of the 12 participants performed better with the *transparent hearing* system than the blocked system, while 7 out of 12 performed worse. For the first group, there is an improvement of 4% when using the

	No Headset	System Off	System On	Open HD
$F \rightarrow B$	14.5%	22.8%	32.6%	56.1%
$B \rightarrow F$	10.5%	22.6%	24.5%	28.3%
Total	12.5%	28.1%	28.5%	42.2%
$U \rightarrow D$	0%	1%	2%	8.6%
$D \rightarrow U$	3.8%	36.1%	61.6%	45.5%
Total	1.4%	14.1%	24.3%	22.5%

Table 1: Front-Back and Up-Down Confusions observed in the experiment

transparent system, while a drop of 5% is observed for the second. The overall trend with respect to the participants' performance remained the same both in high and in low confusion rate directions.

3.4.2.1 Statistical Analysis Statistical hypothesis tests for main effects and interaction were carried out using SPSS statistics software. The data was tested using ANOVA for within-subjects effects, pairwise comparisons were performed with t-Tests. A two-way ANOVA yielded significant main effects of System, $F(3, 33) = 54.387, p < 0.001$, Location, $F(13, 143) = 11.061.p < 0.001$ and a significant interaction between System and Location $F(39, 429) = 2.279, p < 0.001$. All reproduction systems apart from *System On* and *System Off* were significantly different between them ($p < 0.004$), with Open Headphones yielding the worse performance. It is worth noting that there was a significant improvement when using the *transparent hearing* system, $t(4) = -3.51, p = 0.02$, for the group showing improved performance with it, and a significant drop, $t(6) = -7.4, p < 0.01$, for the group showing decreased performance. Direction 15 and 4 yielded the highest confusion rate ($p = 0.001$), followed by 4 and 14 and finally 8. There was no difference in front-back confusion rate for the rest of the locations. The interaction between System and Direction is mainly due to the fact that the confusion rate increased dramatically for all positions for the Open Headphone case, while the ordering for *System On* and *System Off* remained the same as for the unblocked ear-canal case.

When considering up-down confusions, again performance worsened in all systems compared to the unblocked condition. Of considerable interest is that the sounds that were below were significantly more likely to be confused as coming from above than the opposite in all systems, $t(11) = 4.5066, p < 0.001$. Here, *System On* is overall performing at par with *System Off* and better than the open headphone case for sounds coming from above the listener, but yields the worse performance for sounds coming from below. This trend was observed for the majority of participants, on average 8 of the participants performed worse when using the *transparent hearing* system, 3 better and one was at par.

A two-way ANOVA was performed for up-down and down-up confusions separately. For up-down confusions, only a main effect of System was observed, $F(3,33) = 5.659, p = 0.003$. Open Headphones were significantly worse that all other systems, but no difference was observed otherwise. For up-down confusions, again only the effect of System was significant, F(3,33)=13.385, $p < 0.001$ with unblocked ears performing significantly better than all other systems, $p < 0.01$, *System On* performing worse than all other systems ($p < 0.05$) and no difference between *System Off* and the open headphones.

3.4.3 Localization Error

In order to calculate the localization error front-back and up-down confusions were removed from the dataset. The localization performance is examined calculating the error centroid for individual directions. The table in Fig. 17 summarizes this mean error direction in terms of azimuth and elevation angles for all conditions and locations. Figures

ID	1		2		3		4		5		6		7		8	
	Az	El	Az	El	Az	El	Az	El	Az	El	Az	El	Az	El	Az	El
Location	0°	45°	-40°	45°	-90°	45°	-140°	45°	180°	45°	0°	0°	-40°	0°	-60°	0°
No HD	-2,2	56,0	-54,6	53,2	-80,6	51,0	-125,4	59,8	172,5	64,3	-1,2	4,8	-49,2	11,1	-76,8	5,9
Sys Off	-6,3	47,7	-57,6	47,3	-75,9	37,4	-129,4	49,0	170,3	49,0	-6,0	22,0	-61,2	26,9	-74,6	11,5
Sys On	-1,8	33,4	-52,7	43,9	-76,6	39,0	-133,1	45,3	172,6	51,2	-6,1	39,1	-62,8	30,5	-79,3	9,7
Open HD	-5,5	47,2	-49,9	37,1	-107,6	22,9	-142,4	43,3	174,7	28,0	-3,7	21,9	-63,0	13,5	-81,4	8,4
ID	9		10		11		12		13		14		15		16	
Location	-90°	0°	-120°	0°	-140°	0°	180°	0°	140°	0°	0°	-45°	-40°	-45°	-120°	-45°
No HD	-96,5	6,1	-124,1	8,5	-138,9	6,2	173,1	3,6	121,7	6,4	0,2	-43,4	-68,5	-47,6	-131,6	-45,5
Sys Off	-92,4	6,3	-120,2	7,2	-135,7	3,6	173,8	3,5	122,6	6,2	-1,9	-21,6	-55,4	-34,3	-127,4	-39,7
Sys On	-94,9	5,6	-116,5	-0,6	-130,9	1,8	171,1	10,6	123,8	6,3	-2,3	-36,2	-62,4	-31,9	-120,9	-24,5
Open HD	-93,0	4,8	-98,7	6,5	-129,6	13,3	174,8	-1,7	126,6	6,5	4,3	-23,1	-55,4	-33,4	-122,4	-25,1

Figure 17: Localization error centroid for all conditions and locations

18-20 show the resulting judgement centroids and judgment spread on the unit sphere for all conditions. As described above, using Kent-distribution, the data spread can be visualized by an ellipsis indicating directions of highest and lowest dispersion. As expected, the open ears case has the lowest spread for the single directions. Compared to the other conditions the ellipses seem less "stretched", the judgements tend towards a Fisher-distribution. Locations in the horizontal plane were located more precisely than elevated or lowered speakers. Locations between $40°$ and $140°$ azimuth have the tendency to be shifted towards the $90°$ direction.

All other conditions show much higher judgement spread for most of the directions. It is of particular interest that the elevation spread is extremely high in the median plane for all conditions compared to the control condition. Figures 21-22 show modified versions of the distributions. The judgement variances are scaled down and assumed symmetric around the centroid for easier inspection.

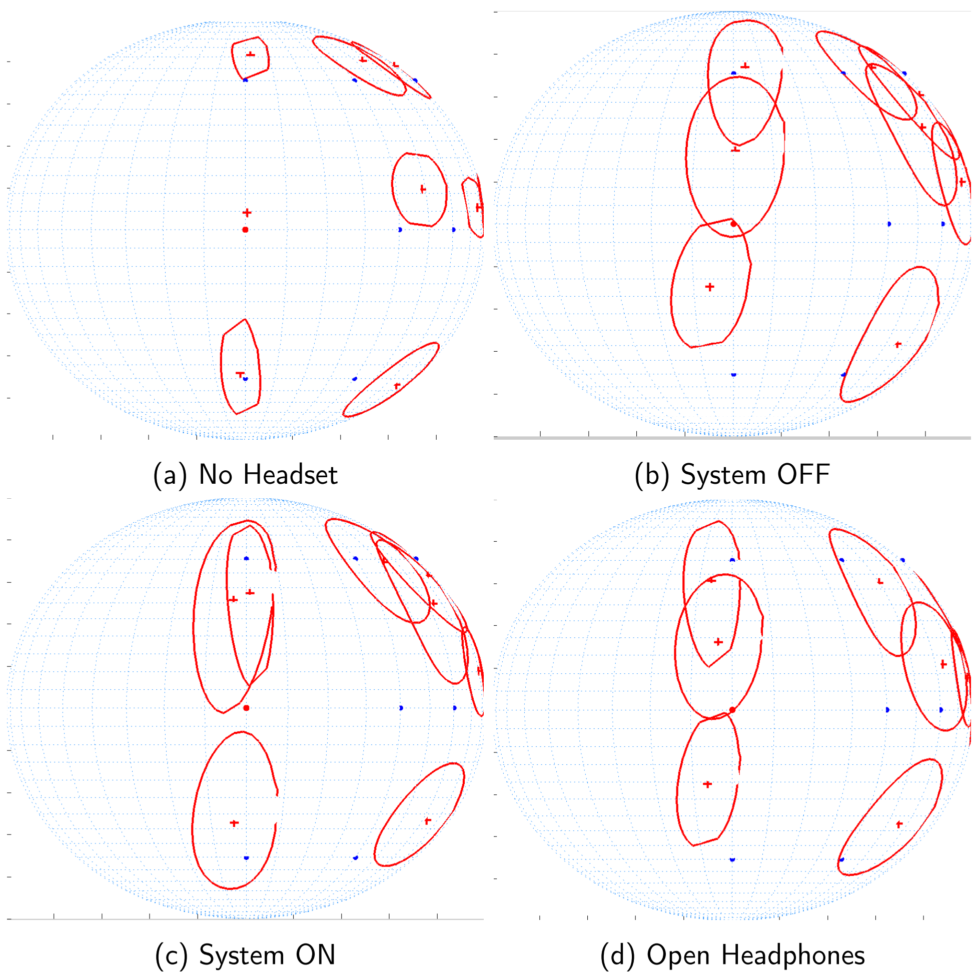
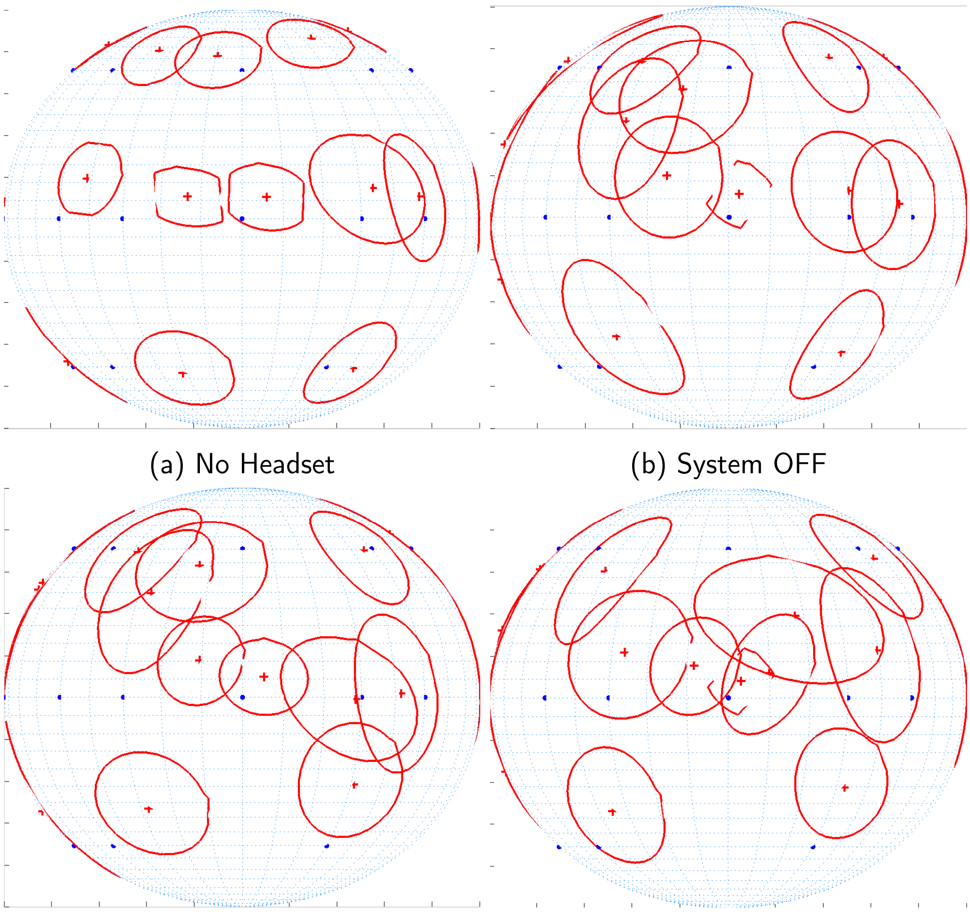


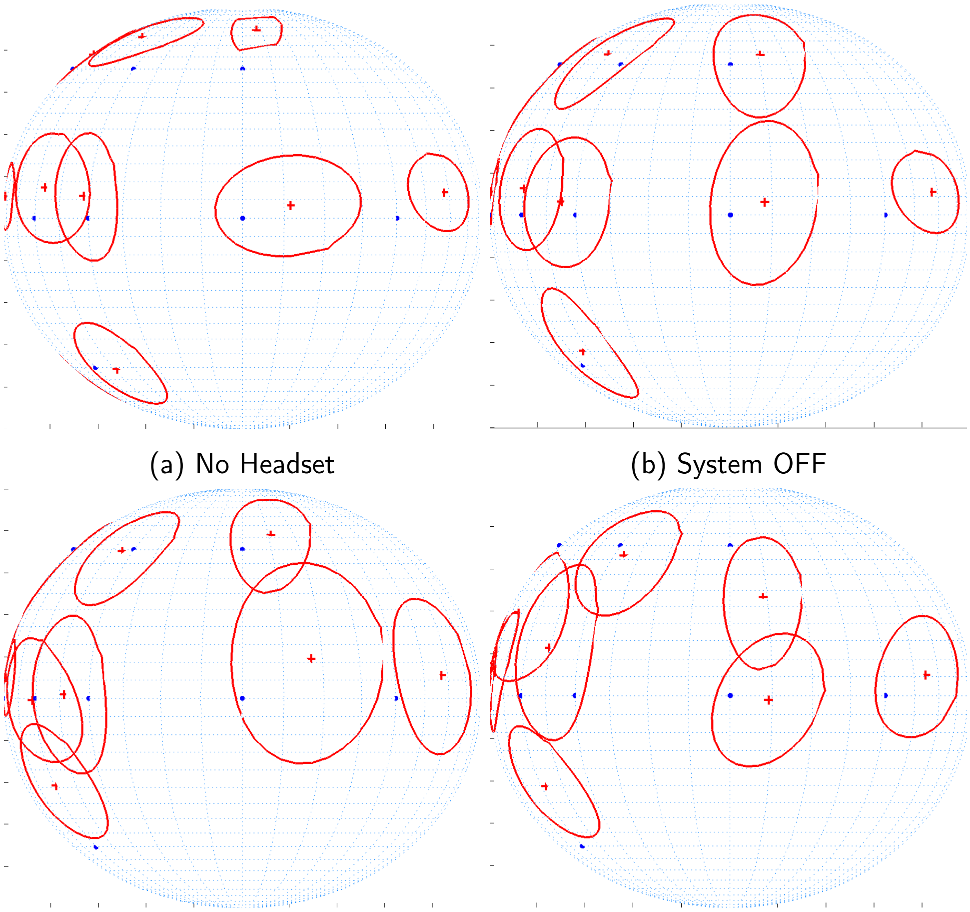
Figure 18: Judgment Centroid and Judgement Spread from viewpoint: $0°$/ $0°$. Both horizontal and vertical lines indicate an angular distance of $10°$



(a) No Headset (b) System OFF

(c) System ON (d) Open Headphones

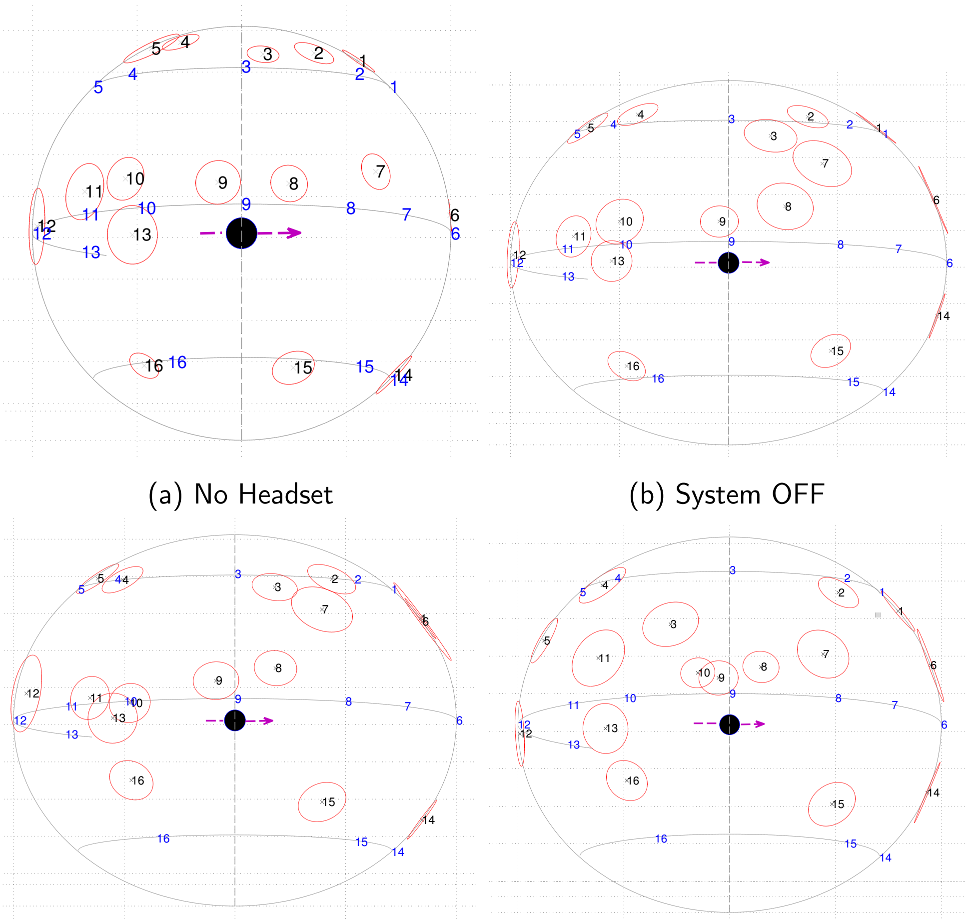
Figure 19: Judgment Centroid and Judgement Spread from viewpoint: $90°$/ $0°$. Both horizontal and vertical lines indicate an angular distance of $10°$



(a) No Headset (b) System OFF

(c) System ON (d) Open Headphones

Figure 20: Judgment Centroid and Judgement Spread from viewpoint: $180°$/ $0°$. Both horizontal and vertical lines indicate an angular distance of $10°$



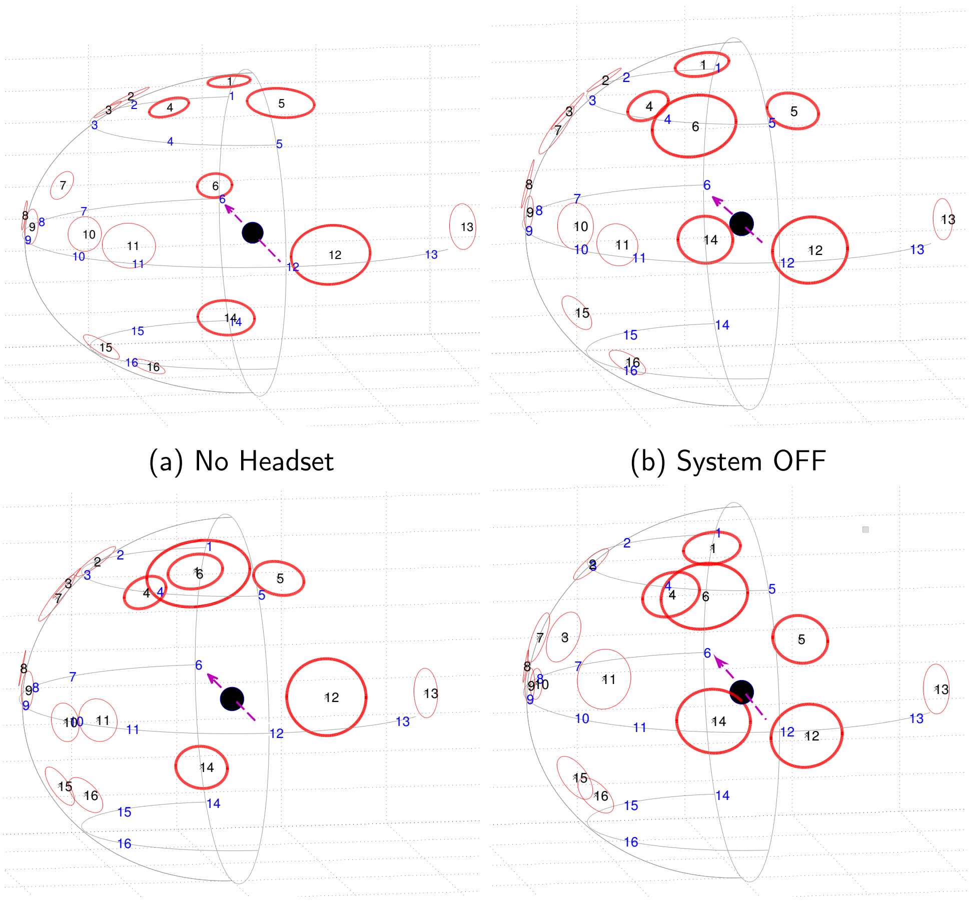
(a) No Headset

(b) System OFF

(c) System ON

(d) Open Headphones

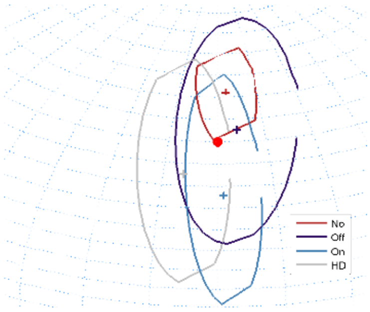
Figure 21: Centroid and Spread, rescaled for easier inspection.



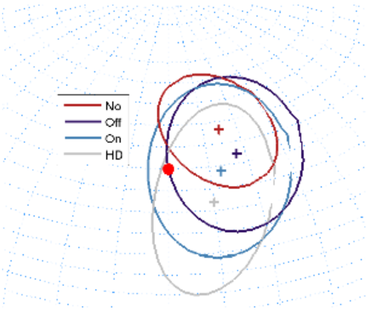
(a) No Headset (b) System OFF

(c) System ON (d) Open Headphones

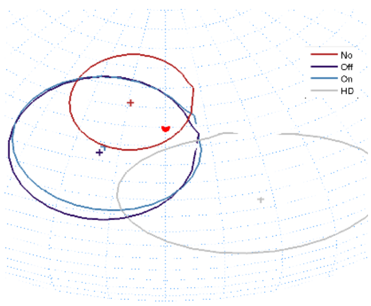
Figure 22: Centroid and Spread, rescaled for easier inspection.



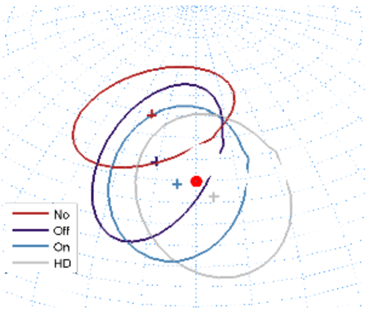
(a) $0°/\ 45°$



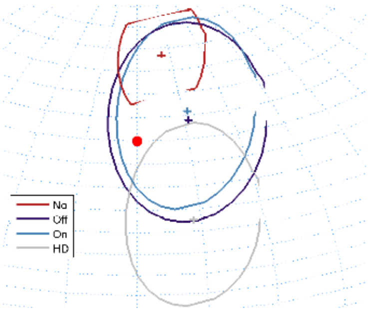
(b) $40°/\ 45°$



(c) $90°/\ 45°$



(d) $140°/\ 45°$



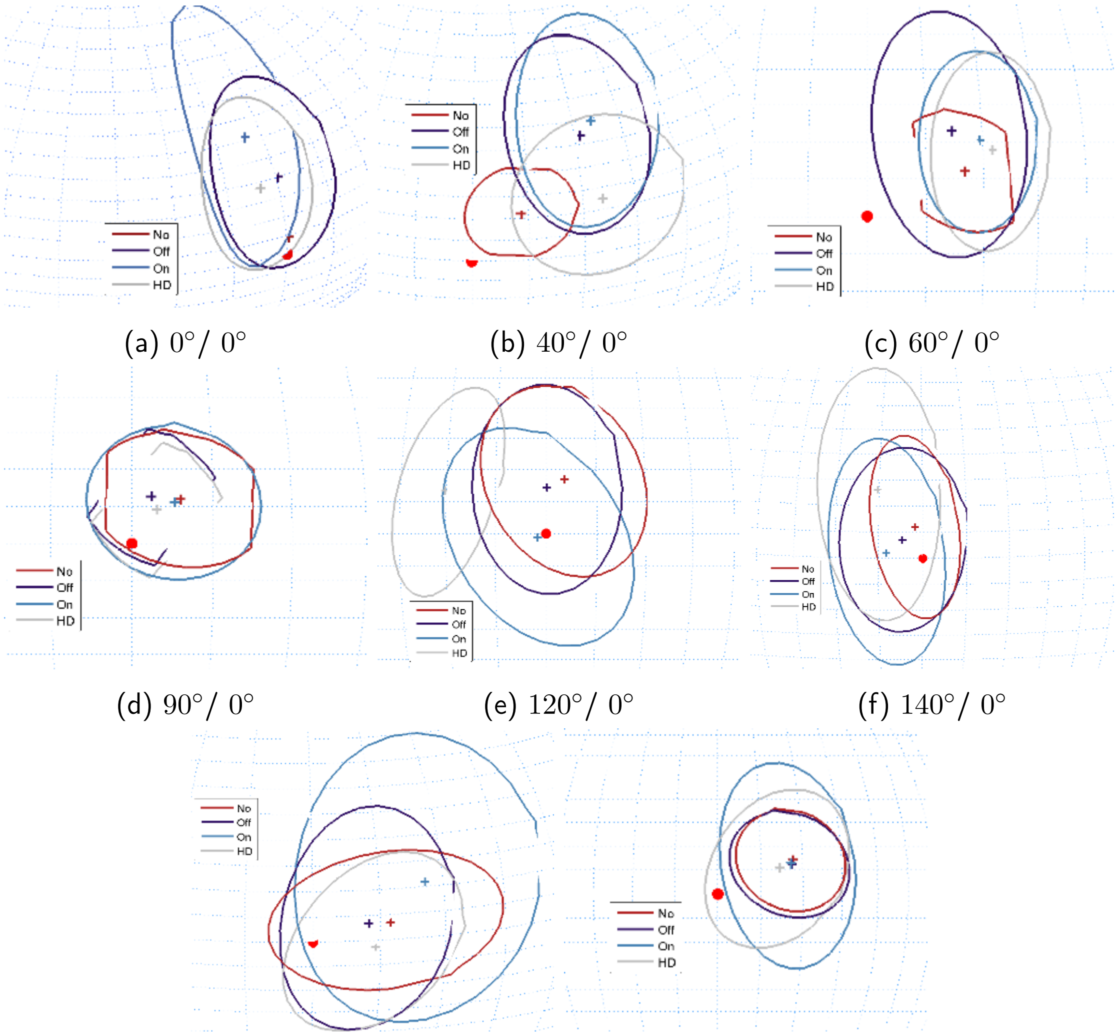
(e) $180°/\ 45°$

Figure 23: Judgment Centroid and Judgement Spread for upper elevation angles.

Figures 23-25 depict the judgment centroid and judgement spread for all locations separately. For elevated locations the spread for all systems is bigger than without headphones, except location 140°/ 45° (Fig. 23-d). The centroids for the open headphones tend to differ notably from the other conditions (see Fig. 23 a, c, and e, or Fig. 24 b, e, and f). The "System ON" case shows an significantly higher dispersion for certain directions especially for azimuth directions in the median plane (Fig. 24 a, g, and Fig. 25 a.)

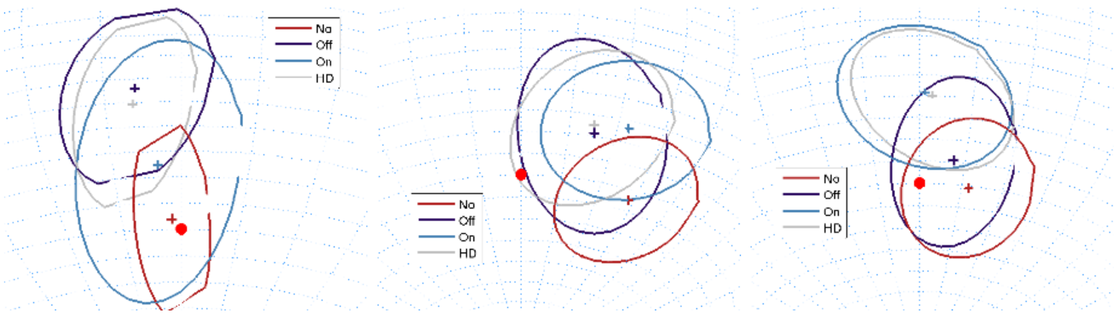
3.4.3.1 Statistical Analysis of judgement spread A two-way-ANOVA in SPSS was used to test the judgement spread R for main effects of "System" and "Location" as well as for interaction effects. It turned out that there were significant main effects of "System" ($F(3,33) = 4.792.p = 007$) as well as of "Location" ($F(15,165) = 2.823, p < 0.001$). No significant interaction effects were detected.

Pairwise comparisons were performed using t-Tests. The control condition (no headphones) is significantly different from system on ($p = 0.044$) and open headphones ($p = 0.021$) . There are no further significant effects between systems, however, no system and system off as well as system off and open headphones are show almost significant differences ($p = 0.066$ and $p = 0.06$ respectively). Maybe a higher number of test subjects would lead to significant effects for those cases.



(a) 0°/ 0° (b) 40°/ 0° (c) 60°/ 0°
(d) 90°/ 0° (e) 120°/ 0° (f) 140°/ 0°
(g) 180°/ 0° (h) 220°/ 0°

Figure 24: Judgment Centroid and Judgement Spread in horizontal plane.



(a) 0°/ −45° (b) 40°/ −45° (c) 120°/ −45°

Figure 25: Judgment Centroid and Judgement Spread for lower elevation angles.

Speaker ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1									sig		sig					
2								sig	sig		sig			sig		
3								sig	sig							
4								sig	sig							
5								sig	sig							
6									sig							
7								sig	sig							
8		sig	sig	sig	sig		sig			sig	sig	sig		sig		sig
9	sig	sig	sig	sig	sig	sig	sig			sig	sig	sig		sig	sig	sig
10								sig	sig			sig				
11	sig	sig						sig	sig				sig		sig	
12								sig	sig	sig					sig	
13											sig			sig		
14		sig						sig	sig				sig		sig	
15									sig		sig	sig		sig		
16								sig	sig							

Figure 26: Summary of significant effects between locations in terms of judgement spread R.

Figure 26 summarizes significant effects between different loudspeaker conditions.It should be noted, that locations 8 ($60°$/ $0°$) and especially 9 ($90°$/ $0°$) show significant differences to almost any other location except each other and location 13 ($220°$/ $0°$). Location 9 (Fig. 24 d) has a low dispersion through all conditions, since it is quite intuitively to detect. However, location 8 (Fig. 24 c) has a notably small spread as well, although the judgement centroids are very close to the ($90°$/ $0°$) direction, indicating that location 8 was mistaken for location 9 frequently.

The SCC was computed for all 12 participants separately and is summarized in Table 2. When considering the mean SCC over all participants, the unblocked ears case shows the best localization performance. It is of interest that in terms of the SCC the *transparent hearing* system shows a considerably worse localization accuracy than System Off.

3.4.3.2 Statistical analysis of spherical correlation coefficient The spherical correlation coefficients of the four conditions were tested for significant main effects using the non-parametric Friedman-Test, since no normal distribution of the data could be assumed. A significant main effect (Chi-Sq$(3, 33) = 19.8$, $p < 0.001$) was detected for the factor "system". For pairwise comparisons, the Wilcoxon's Sign-Rank-Test was applied in MATLAB. The unblocked ears condition was found significantly different from all other conditions ("System OFF": $p < 0.002$, "System ON": $p < 0.001$, "Open HD": $p < 0.001$). Between "System OFF" and "Open HD" the effect was significant ($p = 0.003$) as well. Only the *transparent hearing* setup showed no significant difference to "System OFF" and "Open HD", indicating that there was no improvement in terms of localization error when using the system.

Proband	No System	System Off	System On	Open HD
1	0.9355	0.8580	0.5219	0.8963
2	0.9292	0.8532	0.7539	0.7412
3	0.9569	0.6925	0.6785	0.7977
4	0.7874	0.6828	0.7050	0.6737
5	0.9617	0.8663	0.7418	0.7159
6	0.9168	0.8755	0.8586	0.7723
7	0.7824	0.7926	0.6919	0.7573
8	0.8615	0.7620	0.9038	0.6520
9	0.9531	0.8757	0.7386	0.6919
10	0.8997	0.8105	0.7324	0.6409
11	0.9362	0.6904	0.7639	0.5739
12	0.9062	0.8317	0.7172	0.8066
MEAN	0.9022	0.7993	0.7340	0.7266

Table 2: Spherical correlation coefficient for all participants and systems

	left ear	right ear
system OFF	-0.30	-0.24
system ON	-0.19	0.16

	left ear	right ear
system OFF	-0.36	-0.25
system ON	0.02	0.51

Table 3: Correlation coefficients between localization error and quantification error E at the evaluated positions

3.5 Correlation between Quantification Error and Localization Error

It was of interest too see if the amount of localization error, as observed in the listening experiment, corresponds to the quantification error of the HRTF curves determined from the preceding acoustic impulse response measurements. Table 3 summarizes the correlation coefficients between the two quantities at the evaluated positions for both "System On" and "System Off" cases. The coefficient was calculated for each ear individually. Furthermore, for the localization error two different data sets were taken into account. One where the source confusion errors have been removed, the other without any correction. None of the cases show any considerable correlation. For the blocked ear canal the correlation coefficient is negative but more or less similar for all cases. On the other hand, when the system is on, the coefficient for both ears is almost zero for the corrected localization dataset. However, if the source confusions are not removed from the data, the correlation coefficient is very different for both ears.

3.6 Discussion

With regard to front-back and up-down confusions, it is interesting to observe that the transparent systems performance is better or worse than the blocked ears case depending on the participant. At the same time, there was no significant discrepancy observed for

the same participants for the open-ears case. This suggests that the existing configuration suits some individuals better than others, the reasons should be investigated in future work. Apart from that, the *transparent hearing* poses a considerable improvement over open headphones in terms of confusions.
Concerning the localization error, apparently "easy" directions, namely 1, 6, 9, 12, 14 show a reasonable localization accuracy throughout all system conditions in azimuth, however, elevation errors are considerably higher for all systems apart from the unblocked ear case. Since HRTF cues play a very important role in localization of elevated sources, a possible explanation could be that the cues are destroyed by all systems to a certain degree. The fact, that locations between 40° and 140° azimuth have the tendency to be shifted towards 90° could be explained by the apparently higher ITDs that result from the placement of the microphones further outside from the ear-canal. One observation that should be mentioned is that the judgement centroid of the 180° location is located in the right hemisphere for all conditions. This might be caused by some bias in the experiment setup due to loudspeaker placement or suboptimal room acoustics. In general, the performance was similar with *System On* and *System Off*, suggesting that the particular configuration states no improvement, when considering plain localization results.
Comparison of quantification error and localization error shows that the two quantities (see section 3.5) do not correlate as one would expect. One possible reason could be the fact that the acoustic measurements were carried out in a different spatial environment than the listening tests. Apart from that, limitations in the experiment design could be responsible for the discrepancy (see 5). It is worth observing the fact that the correlation coefficient is different for each ear when the system is on. This indicates that the individual earplugs have not exactly the same acoustic properties assuming a well calibrated measurement setup. Inspection of the quantication error graphs in Figure 9 supports this observation.

4 Sound Quality Test

Although the localization experiment does not show a real improvement regarding the ability to identify the direction of incoming sound events compared to the case where the ear canal is blocked by earplugs, it is evident that such events are represented in a more natural way with transparent hearing. This fact suggests that such a system can provide more conformable communication/interaction with the environment. Besides that, it is of particular interest how people rate the transparent hearing in the presence of music. A second experiment was designed and carried out to evaluate the subjective judgement of such a system. The method of choice was a comparative blind test, where test subjects had to rate different configurations and put them in a certain order according to their preference.

4.1 Inverse Filtering of the Earphone Transfer Function

As described in section 1.2, the influence of earphone presence can be described by a transfer function. Since the sound environment should be displayed by the transparent hearing system as naturally as possible, one way to improve the pseudo-acoustics further is to equalize this earphone transfer function by an inverse filter. A basic approach was used for the inverse filtering and evaluated in the listening test. Figure 27 illustrates the idea of earphone equalization: if it is possible to obtain an inverse for transfer function of the earphone-ear-canal transmission path (the ETF), it can be used to cancel out the earphone's impact on the overall frequency response.

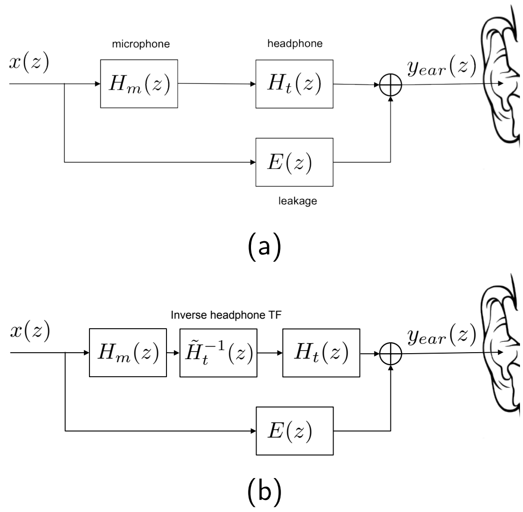


Figure 27: Signal flow of the system (a) without headphone equalization, (b) with headphone equalization.

Such a transfer function can be measured by placing a microphone near the eardrum, however such ETFs appear to be highly individual [HJT+04] and therefore should be obtained for every user separately. For the purpose of this evaluation, ETFs of the Roland system were measured on a dummy head in order to see how well a generic transfer function works for equalization.

$$|H_{inv}(j\omega)| = |H(j\omega)|^{-1} \tag{12}$$

$$\arg[H_{inv}(j\omega)] = -\arg[H(j\omega)] \tag{13}$$

First of all, the impulse response of the earphone mounted on a Br uel & Kjær HATS dummy head was measured. The measurement was carried out via the exponential sweep method [Far00], where the excitation signal was played back via the earphones. The acquired impulse response then was modified in MATLAB by simply inverting both the magnitude-response and the phase-response in order to obtain the inverse counterpart of the ETF (Eq. 12-13). Figure 28 shows the original ETF, the inverse and the resulting equalized frequency response. It can be seen that the inversion works out quite well between 1 kHz and 8.5kHz. However, it appears that the phase inversion evokes a constant level boost of 5 dB through out that frequency range. Such a direct inversion of mixed-phase systems yields an a-causal, infinite and potentially unstable impulse response [OSB99]. The ripples in the magnitude curve appear, because the infinite impulse response undergoes rectangular windowing since it's coefficients had to be implemented directly in a FIR structure. In order to obtain a better linear filter, especially for lower frequencies, more complex algorithms are required.

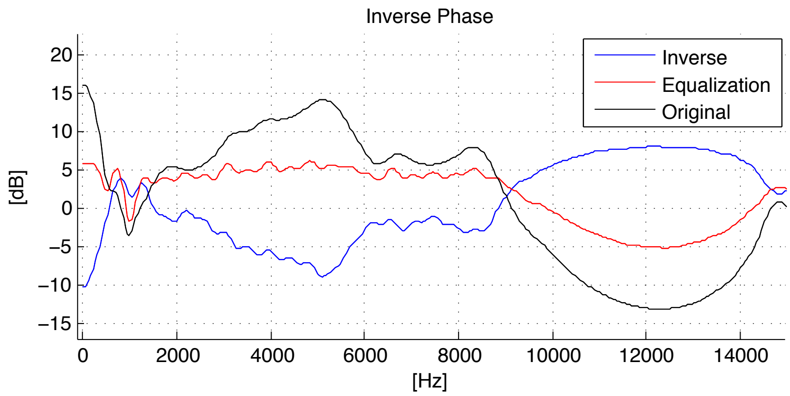


Figure 28: Inverse Filter with resulting equalization curve.

4.2 Test Setup

Test subjects were seated in front of a laptop computer. The basic audio configuration for transparent hearing was the same as for the localization experiment. FIR filter objects and music playback were added to the patch. The GUI, as depicted in Figure 29, was implemented in Pure Data as well and consisted of three buttons for switching between the test conditions in real time and a toggle for music playback control.

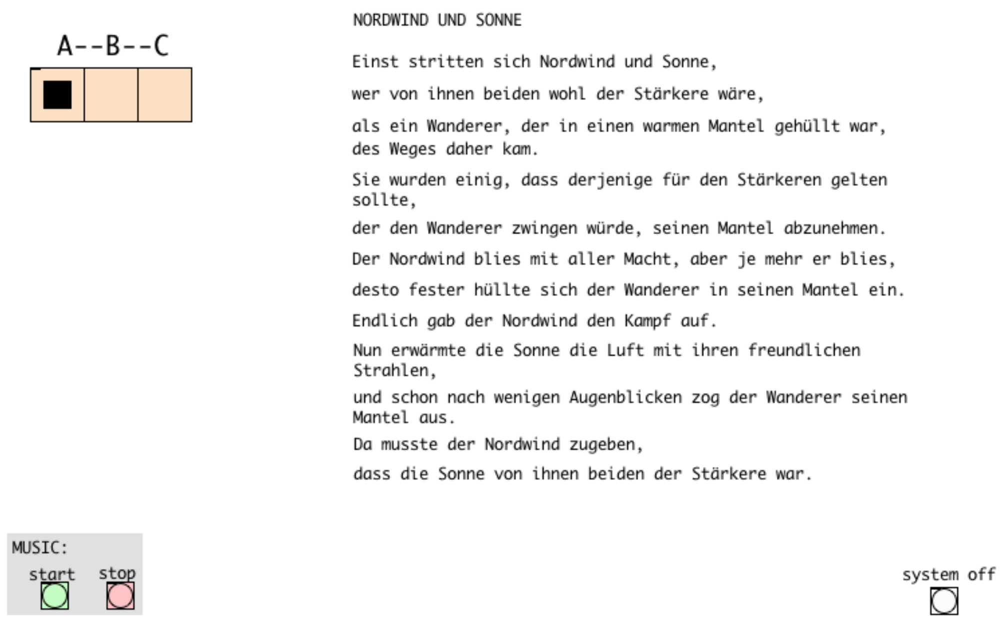


Figure 29: User Interface for comparative blind test.

4.3 Participants

Like for the localization experiment, a group of 12 people took part in this study. Subjects were aged between 22 and 48, half of them being male, the other half female. Nobody had previous experience with listening tests and they participated voluntarily.

4.4 Test Procedure

The task was to compare the set of three different configurations and asses them according to several criteria. Participants were advised to put the three options in an order by placing them on a modified semantic differential scale, like depicted in Figure 30.

The tested conditions in terms of system configurations were:
A Transparent Hearing with ETF equalization
B transparent Hearing without equalization
C Transparent Hearing Off, as control condition

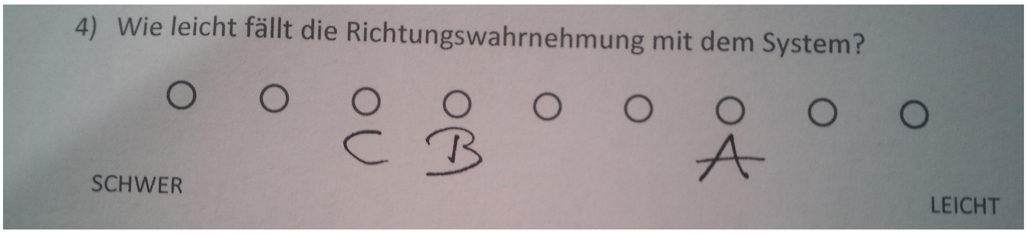
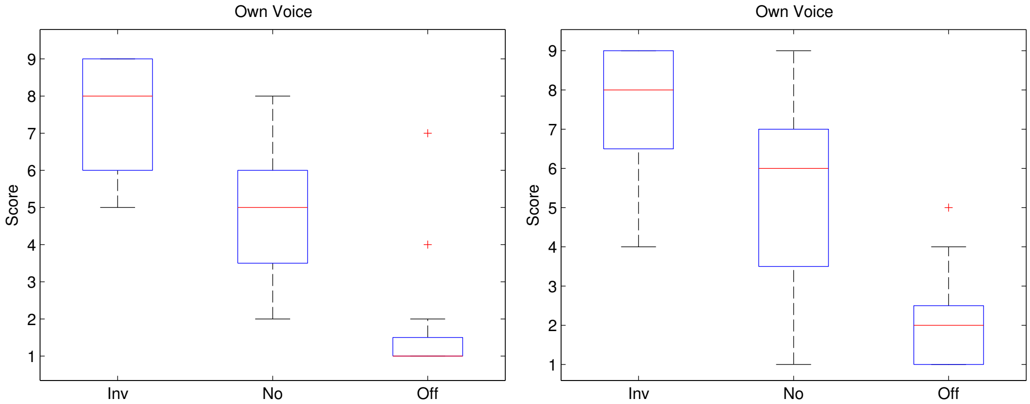


Figure 30: Example of the judgement scale used in the test

The questionnaire consisted of 4 questions:
1) "How natural is the sound of your own voice?"
2) "How natural is the sound of a voice from outside?"
3) "How easy is sound localization with the system?"
4) "Please put the systems in an order according to overall preference.",
and an additional comment field for possible remarks. For the first task the test subjects were instructed to read aloud the text "Nordwind und Sonne" which is widely used for language tests in the German-speaking part. At the same time they could switch between the system configurations on the GUI as often as necessary. For the next two questions, the same text was read by the supervisor for as long as needed to accomplish the task. During the question about localization quality, the supervisor was walking around in the room while reading. In order to rate the overall preference, test subjects had the opportunity to repeat any of the preceding procedures. In a further step, the same test procedure was carried out with an additional music signal, namely the song "Jamming" by Bob Marley. The song was chosen because it's musical structure provides more or less constant dynamics over time, and is of unobtrusive nature from the authors point of view. The signal level was determined in advance in order to find a comfortable ratio between pseudo-acoustics level and music.

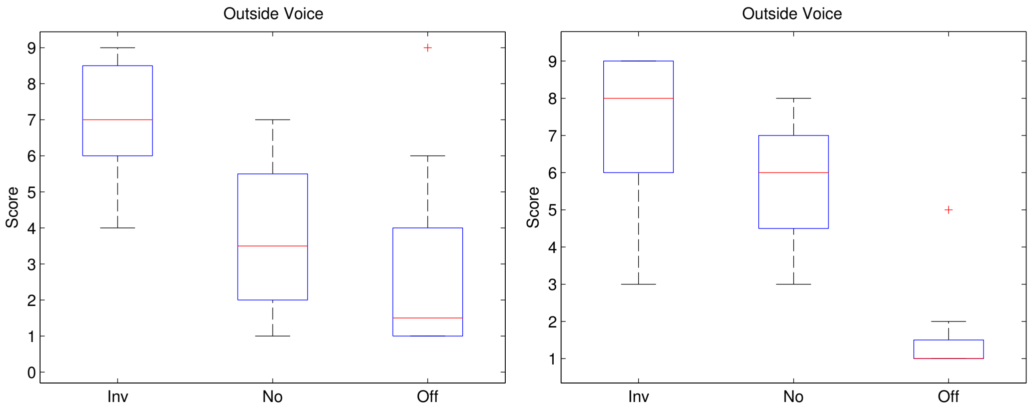
4.5 Results

Figures 31-34 visualize the results for every question using box-plots, where the red line represents the median of the score, the blue box represents the upper and lower quadrille and the whiskers indicate the lowest and highest score. In terms of voice quality, whether own voice or voice from outside, the equalized system is rated notably better. At the same time it seems, that the presence of music does not effect the impression severely. Concerning outside voice, the system without equalization is rated worse on average without music (Figure 32). For the localization task, again, the equalized system was



(a) No Music (b) With Music

Figure 31: Scores for all conditions regarding "Own Voice Quality".



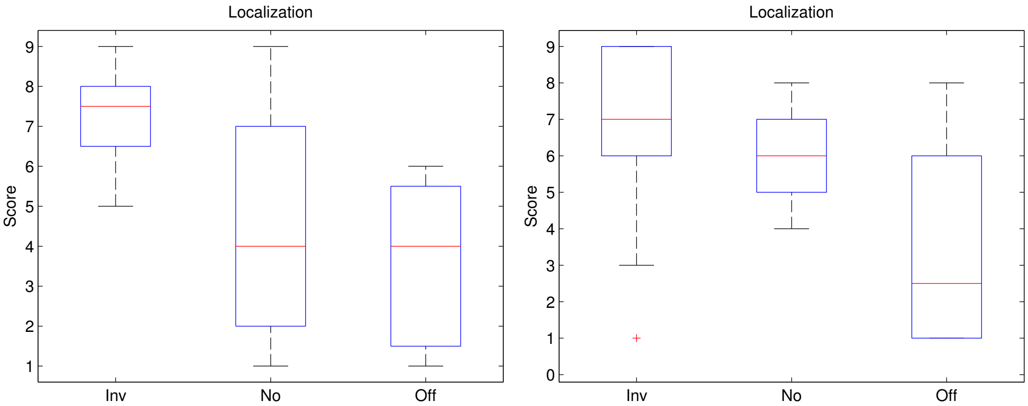
(a) No Music (b) With Music

Figure 32: Scores for all conditions regarding "Outside Voice Quality".

rated best for both conditions. Finally, in terms of overall preference, the equalized system is rated considerably by the participants.

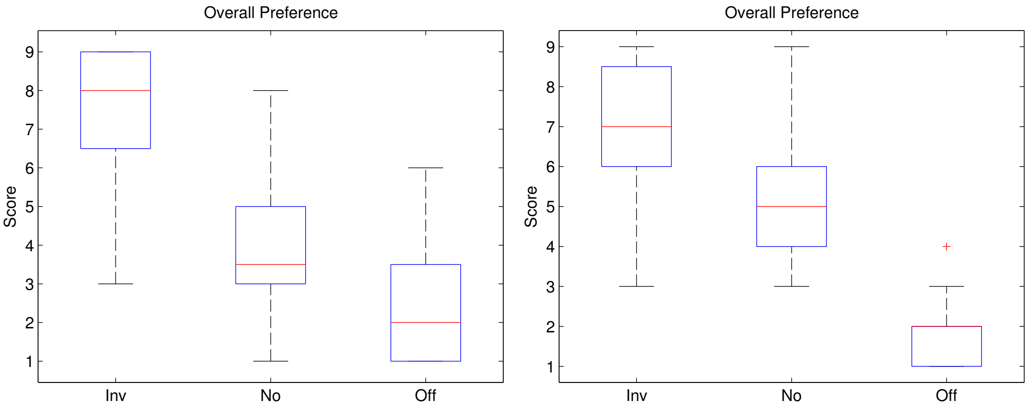
4.5.1 Statistical Analysis

A two-way ANOVA was used to test the results for significant main effects and significant interaction of factors "system" and "music" for each question individually. It turned out, that there are no significant main effects of "Music" or interaction effects throughout the questions. On the other hand, the factor "System" yielded significant effects for every question, as summarized in Table 4. Pairwise comparisons of "system" show significant differences between all configurations, except the localization question. Here the conditions "System On" and "System Off" have no significant effect. As can be seen in Figure 33, for the case without music, both configurations have very similar results,



(a) No Music (b) With Music

Figure 33: Scores for all conditions regarding "Localization Quality".



(a) No Music (b) With Music

Figure 34: Scores for all conditions regarding "Overall Preference".

Question	F-measure	sig. level
Own Voice	$F(3,33) = 39.264$	$p < 0.001$
Outside Voice	$F(3,33) = 48.704$	$p < 0.001$
Localization	$F(3,33) = 11.976$	$p < 0.001$
Preference	$F(3,33) = 39.856$	$p < 0.001$

Table 4: significant main effects of "system" for all questions

the median is the same.

4.6 Discussion

Concerning outside voice, the system without equalization is rated worse on average without music (Figure 32). This seems kind of obvious, since it can be expected, that music is masking the system's noise to a certain degree, making harder to hear a difference at all. For the localization task, again, the equalized system was rated the best for both conditions. However, the particular results have to be treated with caution, since the participants were allowed to move their head and were able to see the actual movement of the supervisor and thus use visual feedback. Finally, the overall preference rating shows that the equalized system can be considered an improvement for the given task, both with presence of music and without. It should be noted that while the present box-plots visualize the overall tendencies quite well, they have to be treated with caution, since for representative information about data distribution the number of participants should be higher.

5 Limitations

Some limitations of the experiment design should be taken into account when interpreting the results of the location experiment. First, the room was an issue, since the experiment had to be set up in a corner. This fact implied the disadvantage to have two sound-reflecting walls in the vicinity of some speakers. It was tried to reduce possible reflections by attaching some absorbing material behind the respective speakers. Still, it could not be guaranteed to overcome possible reverberation issues thoroughly. The user input interface could have been more intuitive. One cannot deny that it might be quite challenging for the participants to execute two tasks in order to enter a single direction. Nevertheless, it seemed to be the most promising solution under the circumstances described above. Apart from that, the question arises if expert listeners would have been more suitable for this kind of experiment, since it appears challenging to map a perceived location to actual angular values, as demanded by the particular interface.

Concerning the second experiment, results of the question about localization quality should be handled with care. The very good rating of the equalized system could be caused by bias, since the participants may already have had a positive impression of the equalized systems sound quality. Nevertheless, at least for the "music on" condition, also the non-equalized system was rated remarkably better than the blocked ear canal. This rises questions around to what degree head-movement and visual feedback improve the performance of the *transparent hearing*, since the isolated localization experiment showed no significant differences between "System ON" and "System OFF". Furthermore, it seems that the sole presence of amplified environmental sound gives the impression of improved localization. In the future, such a device should be evaluated in an experiment where those factors are considered as well.
Finally, the headset configuration itself could have been more sophisticated concerning sound quality. In a next step, a prototype headset with well known acoustic properties should be designed, making sure that HRTF cues are preserved in the recorded signal. Possible influence of system latency could be eliminated by implementing mixer and filters using analog hardware.

6 Conclusion

The concept of *transparent hearing* as a possible solution for issues of traffic danger and limited interaction with the environment when using portable audio appliances was presented. A basic prototype system was implemented using an existing binaural microphone-headphone headset and evaluated against its localization performance and sound quality.
Listening tests show no significant improvement compared to a blocked ear canal in terms of localization of single auditory events. This suggests, that the frequency response of the particular headset system is not flat enough to preserve HRTF cues, as the results of the acoustic measurements indicate.
Concerning source confusions (front-back, up-down) it is remarkable that the system shows significantly better results than the blocked ears case for a certain group of people, while for others it performs significantly worse. It appears likely that individual physiognomic differences influence the performance. Furthermore, there are significantly less confusion errors compared to open headphones, consequently the system brings some improvement over existing solutions.
In terms of sound quality the system brings significant improvement over normal headphones, in particular considering the perception of human voice. This can be enhanced even further by applying basic headphone equalization filters. When subjects are asked about the systems localization capability, results show a significant improvement when the system is on during music playback. Obviously, even the presence of audible envi-

ronmental sound events provides a better impression of localization, despite the results of the localization study. Therefore the current system can be beneficial in non-critical situations where it can enhance the users preference and facilitate social interaction while listening to music. On the other hand, the transparent hearing system is not suitable for situations where proper localization capability is crucial for safety like in urban traffic.
For that reason, future work should cover the design of a headset system capable of preserving the relevant localization cues and the proper evaluation of such a design that takes into account factors like visual feedback, head-movement and the loudness level of environmental sound.

References

[Ben95] B. Benderson, "Audio augmented reality: A prototype automated tour guide," in *ACM Human Computer in Computing Systems conference companion*, 1995.

[BKES03] D. Brungart, A. Kordik, C. Eades, and B. Simpson, "The effect of microphone placement on localization accuracy with electronic pass-through earplugs," in *Applications of Signal Processing to Audio and Acoustics, 2003 IEEE Workshop on.*, oct. 2003, pp. 149 – 152.

[Bla12] S. Blamberger, "Development of a modular system for speaker array prototyping," University of Music and Dramatic Arts - IEM, Tech. Rep., 2012.

[BWB08] H. Benko, A. D. Wilson, and R. Balakrishnan, "Sphere: Multi-touch interactions on a spherical display," 2008.

[Car96] S. Carlile, *Virtual Auditory Space: Generation and Applications*, 1st ed. Landes Bioscience, 1996.

[CLH97] S. Carlile, P. Leong, and S. Hyams, "The nature and distribution of errors in sound localization by human listeners." *Hear Res*, vol. 114, no. 1-2, pp. 179–96, 1997.

[Far00] A. Farina, "Simultaneous measurement of impulse response and distortion with a swept-sine technique," in *Audio Engineering Society Convention 108*, 2 2000.

[GG96] M. D. Good and R. H. Gilkey, "Sound localization in noise: The effect of signal-to-noise ratio," *The Journal of the Acoustical Society of America*, vol. 99, no. 2, pp. 1108–1117, 1996.

[HJT+04] A. Härmä, J. Jakka, M. Tikander, M. Karjalainen, T. Lokki, J. Hiipakka, and G. Lorho, "Augmented reality audio for mobile and wearable appliances," *J. Audio Eng. Soc*, vol. 52, no. 6, pp. 618–639, 2004.

[LC98] P. Leong and S. Carlile, "Methods for spherical data analysis and visualization." *J Neurosci Methods*, vol. 80, no. 2, pp. 191–200, 1998.

[LCB+05] V. Lemaire, F. Clerot, S. Busson, R. Nicol, and V. Choqueuse, "Individualized hrtfs from few measurements: a statistical learning approach," in *Neural Networks, 2005. IJCNN '05. Proceedings. 2005 IEEE International Joint Conference on*, vol. 4, 31 2005-aug. 4 2005, pp. 2041 –2046 vol. 4.

[LSAM12] R. Lichenstein, D. C. Smith, J. L. Ambrose, and L. A. Moody, "Headphone use and pedestrian injury and death in the united states: 2004 - 2011," *Injury Prevention*, 2012.

[MBL07] P. Majdak, P. Balazs, and B. Laback, "Multiple exponential sweep method for fast measurement of head-related transfer functions," *J. Audio*

Eng. Soc, vol. 55, no. 7/8, pp. 623–637, 2007. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=14190

[MK02] F. Mueller and M. Karau, "Transparent hearing," in *CHI 2002 Extended Abstracts on Human Factors in Computing Systems*, 2002.

[MRK+10] J. Mantell, J. Rod, Y. Kage, F. Delmotte, and J. Leu, "Navinko: Audio augmented reality-enabled social navigation for city cyclists," in *Pervasive 2010 W4: Multimodal Location Based Techniques for Extreme Navigation*, Helsinki, Finland, 5 2010.

[OSB99] A. V. Oppenheim, R. W. Schafer, and J. R. Buck, *Discrete-time signal processing (2nd ed.)*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1999.

[SC98] B. Shinn-Cunningham, "Applications of virtual auditory displays," in *Proceedings of the 20th International Conference of the IEEE Engineering in Biology and Medicine Society*, vol. 20, no. 3, 1998.

[Tik09a] M. Tikander, "Development and evaluation of augmented reality audio systems," Ph.D. dissertation, Helsinki University of Technology, 2009.

[Tik09b] ——, "Usability issues in listening to natural sounds with an augmented reality audio headset," *J. Audio Eng. Soc*, vol. 57, no. 6, pp. 430–441, 2009.

[VAB10] Y. Vazquez-Alvarez and S. A. Brewster, "Designing spatial audio interfaces to support multiple audio streams," in *Proceedings of MobileHCI*, 2010.

[VRAT99] C. A. V. R. Algazi and D. Thompson, "Dependence of subject and measurement position in binaural signal acquisition," *J. Audio Eng. Soc*, vol. 47, no. 11, pp. 937–947, 1999.