

György Wersényi

*Széchenyi István University, Department of Telecommunications, Egyetem tér 1., H-9026, Győr, Hungary, [wersenyi@sparc.core.hu](mailto:wersenyi@sparc.core.hu)*

## Localization in a HRTF-based Minimum-Audible-Angle Listening test for GUIB applications

*Received 15.12.2006, published 16.01.2007*

Listening tests were carried out for investigating the localization judgments of 40 untrained subjects through equalized headphones and with HRTF (Head-Related Transfer Function) synthesis. The investigation was made on the basis of the former GUIB. (Graphical User Interface for Blind Persons) project in order to determine the possibilities of a 2D virtual sound screen and headphone playback. Results are presented about the minimum, maximum and average values of discrimination skills. The measurement method includes a special 3-category-forced-choice Minimum Audible Angle report on a screen-like rectangle virtual auditory surface in front of the listener. Average spatial resolution of 7–11° and 15–24° were measured in the horizontal plane and median plane respectively dependent of spectral content of the noise signal excitation. Accessory signal processing is suggested to enhance poor vertical localization performance.

### INTRODUCTION

State-of-the-art virtual reality applications are available using head-tracking device, simulation of room reverberation, individual HRTFs and full auralization. These applications may need special hardware, lot of computation time for real-time rendering, circumstantial measurement and installation procedures restricted to expert users and researchers. On the other hand there is a need for low-cost solutions for everyday users, e.g. for visually disabled person to help to use personal computers with simplicity. This case we have to deal only with a plug-n-play sound card running under Windows OS and a headphone, often without any equalization or individual HRTF synthesis.

This study contributes to the former GUIB (Graphical User Interface for Blind Persons) project. After introducing the goal and former results of this we briefly summarize the known problems and possibilities of virtual audio synthesis and listening tests. Detailed description is given in section 3 about the measurement system we use, including test signal generation and playback method. In section 4 results are presented about the localization blur in vertical and horizontal directions on a 2D virtual audio display. At the end conclusions are drawn for GUIB applications and future works are highlighted.

The aim of the former GUIB project was to find solutions for the help of elderly and visually disabled people by the use of personal computers. Blind persons do not have the advantageous properties of graphical user interfaces (GUI) like MS-Windows, icons and the ability of orientation among multiple visual information [1, 2, 3]. Visual events on the screen,

such as opening files, closing windows, movement of the cursor, etc. have to be replaced or extended by sound events. The former results of this project related to sound reproduction are:

- a collection of sounds, representing visual icons and events of the screen only by acoustical information called “earcons” [4],
- the possibilities of different input media [5, 6, 7] and
- the localization blur using a multi-channel loudspeaker playback system [1].

The surprising finding of the latest test has been: blind persons cannot localize better than people with normal vision; furthermore, loudspeaker playback is not suited for a real-life application. The so-called Sound Screen was a multi-channel array of loudspeakers with low spatial resolution. It was also large, heavy and disturbing for the environment (e.g. in an office). Nevertheless, connecting directional filtering and multichannel recordings played back via commonly used 5.1 loudspeaker arrangement is still an ongoing procedure [8].

It was therefore suggested to determine the localization blur with the same system using headphone playback as well.

## 1. SYNTHESIS OF VIRTUAL AUDIO ENVIRONMENTS

Virtual audio synthesis is usually made through headphones. Some results exist from loudspeaker playback under restricted conditions and use. The goal is to create a virtual sound field for the listener in which he is able to localize sound sources and identify changes in sound source locations. During the simulation, a sound file is transmitted through a processing queue including directional dependent filtering. This can be made real-time in the frequency domain or in the time domain and could result in satisfying localization performance in the virtual audio space.

Sound waves, reaching the eardrums, are affected by directional filtering of the outer ears. This binaural filtering effect determines basically the perception of the direction of sound sources depending on the angle of incidence [9–12]. Monaural cues are responsible for the perception of the elevation in the median plane, front-back directions and distance. The Interaural Time Delays (ITD) and the Interaural Level Differences (ILD) are the basic cues for the localization in the horizontal plane that results in a much better localization performance [13–18]. The directional information, delivered by the filtering effects of the outer ears, is complete at the entrance of the ear canal and this information does not vary along the cavity of the ear canal [10, 11, 19].

The transmission from a point in the free-field to the eardrums is described by the complex Head-Related Transfer Functions (HRTFs). In virtual audio environments the HRTFs have to be reproduced through headphones. We can use individual HRTFs, HRTFs from a “good localizer” or from a dummy-head for sound field reproduction. This fact influences the localization [12, 20–22]. It was shown, that HRTFs from a good localizer and the use of simple methods to make them more individual (like scaling in the frequency) can result in satisfying localization [23]. Nowadays, computational performance allows real-time convolution using the time domain variant of the HRTF called Head-Related Impulse Response (HRIR) [24, 25]. Other basic components for the localization are: spectral content, bandwidth, volume, duration, adaptation and learning, a-priori knowledge, and additional visual information.

According to the statements of the *binaural technique*, if we reproduce the sound pressures at the eardrums exactly, the listener will have the same spatial information as he would get if he was present at the sound field. For the reproduction a proper and individual headphone-equalization is required, as far as possible [26–29]. This technique may contain errors as well, like front-back confusion and in-the-head localization due to headphone playback [19, 30]. Headphone auralization often produces an incorrect localization of sound sources and the most significant problem is in-the-head localization [31]. In general, results from free-field measurements tend to be better than when using headphone playback without individual recordings, but the use of individual HRTFs increases the localization performance [32].

*Localization* means finding the absolute position of the sound source. *Localization blur* is the smallest change in the direction of the sound source that can be perceived. To measure the latest, we have to search for the Minimum Audible Angle (MAA) or the Just Noticeable Difference (JND), where subjects only have to compare two sound sources and identify only the change of the source direction [33–39].

Results in this field are difficult to compare, because experimental designs and methods differ. For a direct comparison of results, similar conditions are needed. Furthermore, better results can be obtained in a MAA measurement in contrast to an absolute measurement.

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

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

These can be reduced by adding head-tracking, room reverberation or using an alternative headphone design [40, 41]. The number of front-back confusions could be larger in case of individual HRTFs than for generic HRTFs, which is surprising because it is usually assumed that the use of individual HRTFs helps resolving front-back reversals [31].

Virtual Acoustic Displays (VADs) are widely used in several applications. A VAD identifies a virtual environment, where sound sources are artificially reproduced and listeners are able to localize and identify them. To realize a VAD, two independent questions have to be answered. First, which sounds correspond best to the visual representation and meaning of the object to be reproduced? In other words, what is the best mapping between sounds and events on the screen? Second, what is the localization blur using headphone playback?

In principle, three-dimensional VADs can be realized by reproducing *depth* or *distance* information as well, e.g. by an object approaching to the listener or by overlapped windows. Overlapped windows are common in a GUI where usually only one window is displayed in full screen (or full sized) and others are running in the background, behind the active window. The perception of distance is relatively poor in human localization. State-of-the-art multimedia computers and applications nowadays allow full auralization. Only recently it became possible to handle huge amounts of computation data, real-time filtering of HRTFs, reverberation and head movement effects [42–49].

## 2. MEASUREMENT METHOD

For a future application with earcons, the localization blur of different signals have to be determined. The earcons are short (about hundreds of milliseconds) tones or special noisy-like sound events. Therefore, in the listening tests, we decided to use sound events of 300 ms duration of unfiltered and filtered versions of broadband noise to match and model in a generic but not too specific way the possible real application of earcons.

Furthermore, the measurement has unusual and novel methods like the 3-category-forced-choice, where subjects have to select from three possible answers, in order to determine the “uncertainty” of the listeners during their localization judgments (see later). A two-direction discrimination will be applied to determine the localization blur, independent of the direction of the moving sound source.

Usually the coordinate system is related to the geometry of a VAD, and it is spanned by the scene [31, 50–52]. Instead of the commonly used method to measure the discrimination of sound sources with constant distance, a “virtual rectangle screen” is simulated. The virtual sound screen is a 2D square surface in the front of the listener. It was selected because only few experimental results exist with non-constant source distances in the front of the listener. Usually, horizontal plane experiments are made with constant source distance around the head.

Secondly, the mapping from a visual screen (PC monitor) is better to a “screen-like” 2D virtual sound screen for the orientation with the mouse (see Fig. 1). The maximum range of simulated sound sources is  $\pm 60^\circ$ , both horizontal and vertical. Because the distance of the source is not constant, sources beyond  $60^\circ$  are “too far away”, and it is assumed that subjects make their localization judgments based only on this distance information. In addition, we assume that the listener in a real life application would be able to adjust volume, so the parameter “distance” is neglected. We have to mention that externalization is maybe related to the perception of auditory distance, so control of signal level is maybe helpful to reduce in-the-head localization [31].

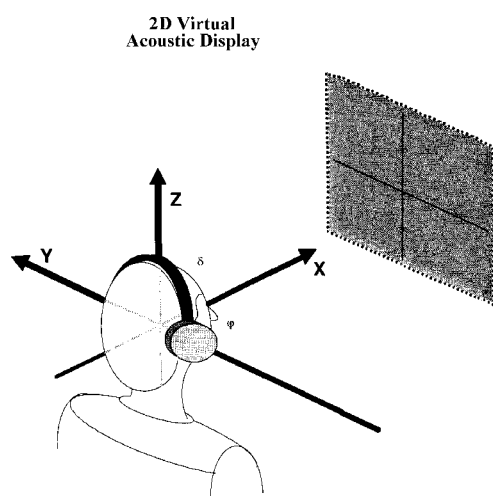


Figure 1.

Illustration of the 2D VAD [1]. The virtual acoustic surface is parallel with the Z-Y-plane. The origin is in the front of the listener:  $\varphi=\delta=0^\circ$ . Virtual objects move during the measurement parallel with the Y- or the Z-axis, in the horizontal or median plane respectively. The black cross on the virtual screen illustrates possible sound source locations

## 2.1. Setup

The measurement setup is based on a PC with the Beachtron DSP Board. Real-time convolution of the mono input signal and the HRTFs is made in the time-domain (16 bit; 44,1 kHz). The system is precisely equalized for the circumaural, open-dynamic Sennheiser HD540 headphone. The HRTFs originate from a good localizer in a measurement of *Wightman and Kistler* [24, 53, 54]. 72 measured HRTFs are available in a form of a 75-point minimum-phase-FIR-filter set with 30° spatial resolution. Synthesis of motion is achieved by linear interpolation between impulse responses derived from the four nearest minimum-phase HRTFs, with the interaural delays interpolated separately and inserted at the end of the filtering process [24, 53]. Linear interpolation of HRIRs and the use of ITD information in addition is widely investigated. This method was found to be the best method in case of missing measured HRTFs [55–57]. Duration and volume of the test signals were determined during a pre-test with 7 subjects. The existence of headphone playback errors was also determined [2, 3]. The main test was made with 40 untrained subjects, all with normal hearing. Untrained subjects are widely used in listening tests because their localization performance is worse than it is of trained subjects [58]. The individual modifications of the HRTFs correspond to measure the size of the head (distance of the ear canal entrances). Setting the ear canal distance could decrease the angular error [59].

Localization depends on the signal frequency (bandwidth), duration, loudness and a-priori knowledge. To reduce the parameters, we work with constant signal volume and duration. Excitation signals for the MAA-measurement are 300 ms noise burst impulse-pairs: white noise (signal A), 1500 Hz low-pass filtered noise (signal B) and a 7000 Hz high-pass filtered version of white noise (signal C). The SPL of signal B is by 10 dB, the level of signal C is by 6 dB greater than the level of signal A for an almost constant sensation of loudness. These values are averaged according to the subjects' opinion during a preliminary test. They had to determine the SPL for signal B and C to be as loud as signal A.

Based on the literature, broadband noise bursts must exceed 100 ms to be the length independent from the sensation of loudness [60]. Stimulus frequency and duration are widely investigated in this context [61].

Signal A, B and C were chosen to fulfill the following requirements:

- In length they have to match the earcons.
- They have to be generic (no speech, music or earcons yet).
- And similar to the earcons, which are tones or noisy-like broadband sound events.
- For investigating the dependency of frequency they do not have to be too specific.
- Subjective loudness should be the same.
- Signals exceeding 40–80 dB and 250 ms are to localize the best. Above 50 dB and 250 ms length the localization is independent of duration and loudness, so these values must be exceeded [9].
- Signal A is seen as “original” signal, and signal B and C are derived from it by low-pass and high-pass filtering respectively. It is suggested due to poor vertical localization to

use simple frequency filtering later to model elevation [2, 3], so the localization blur of the filtered versions of signal A also have to be determined.

- Cut-off frequencies for the filtering were chosen to be far away from each other in the frequency for a good separation between signal A (whatever it is) and its filtered versions.
- We have taken into account that MAA was found to be optimal for signals that either are below 1000 Hz or above 4000 Hz. MAA has a minimum between 250 and 1000 Hz, and above increases to a maximum. Between 3 and 6 kHz there is another minimum. [35–39, 62–68].

All these lead us to choose 1500 Hz and 7000 Hz for cut-off frequencies, but signals with different spectra or filtering may have different MAA results.

## 2.2. General conditions

Novelties and general conditions in our measurement are:

- The use of a 2D virtual sound screen in the front of the listener. Sources can move only in the horizontal (left and right) and in the median plane (up and down) from the origin in 1° resolution. The source distance is not constant and the source is *not moving around the head* as usual.
- Subjects have to report in a 3-category-forced-choice: “no difference between the sources”, “different sound sources” and “I’m not sure”. This is, because the subjects have spatial domains where they are uncertain. The size of this domain can be determined.
- Sound events in pairs have to be compared: there is a stationary reference sound and a moving sound source (producing the same signal). The moving source is moving away from the reference source, after then it moves backward to the reference point. We are looking for the nearest distance from the reference, where listeners are able to discriminate the sources with certainty. This will be chosen for the new reference point. The auditory system has deteriorative accuracy and localization performance in case of an “incoming” sound in contrast to an “outgoing” sound event. This is because it is easier to perceive a sound event if it is loud and fades out as if it fades in from the silence. If we determine the localization blur from both directions of moving, we will get the direction-independent localization performance of the subjects.

The first impulse of the burst-pair is always a fixed reference point, and the second is moving first away, then backward to the reference point. During the MAA measurement, subjects were asked to report in a 3-category-forced-choice. Possible answers were: “no difference” if the subject is not able to discriminate the sources and they seem to come from the same direction. “Different sound sources” means that he is able to distinguish between the signals. He may have the possibility to choose the answer “uncertain”, if he is not sure which is the case.

At the beginning, the reference point is always in the origin. The second source is moving away from this reference. After the subject has reported “different sound sources” the moving

source moves backward. The nearest point where the subject in both direction of moving was able to distinguish the sources will be selected as the new reference point. Maximal total number of reference sources is limited to 13 horizontal (6 left and 6 right).

The same test is made then in the median plane as well, where only two new reference points could be selected up and down respectively. The pause should exceed the 300 ms length of the signal for a correct separation of the burst pairs, so it was chosen 400 ms.

In [62] a similar method was used, but only in a 2-alternative forced choice as the subject's response was used to initiate the next trial. In [63] the subjects had also to report in a forced-choice using pulse-pairs and they had the possibility to be uncertain. But this was not investigated deeply.

### 2.3. Subject selection

20 male and 20 female test persons between 21 and 39 years of age took part in this listening test under the conditions mentioned above. The test was carried out in the anechoic room. Results are presented below, showing average (AVG), maximum (MAX) and minimum (MIN) values of the measured data. Subjects were sitting on a comfortable chair with a signal button in the hands. During the 10 minutes of accommodation time the distance of the ear canal entrances (size of the head) was measured, a detailed explanation of the procedure was given and a trial run was made in one direction. Measuring the head diameter allows to set the ITD information more correctly in case of interpolated HRIRs [56, 57].

Maximum, minimum and averaged values of the measured head diameter and age of the subjects are shown in Table 1 and 2 respectively.

Table 1. AVG, MIN and MAX values of the measured distance between the ear canal entrances. Total average over every subject is 13 cm

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

Table 2. AVG, MIN and MAX values of the ages of the subjects. Total average over every subject is 28 years

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

First, signal A was presented in the directions “down”, “up”, “left” and “right”. After a few minutes break we continued with signal B and signal C. Overall time for the test was about 60 minutes (15 minutes for each test signal on average).

At the end, subjects had to fill out a questionnaire about personal data (gender, age), computer skills (result was: 59% “professional or engineer”; 41% “everyday user”) and headphone user routine (result: 7% “everyday”; 24% “often”; 59% “seldom”; 10% “never”).

### 3. RESULTS

Results were found to be independent of age and computer skills, but little improvement in the localization performance was found by subjects using headphones often. The spatial resolution is as good as independent of the gender. This fact was also observed by *Chen*, in an investigation for evaluating stimulus duration in context with localization [69].

Figures 2–3 show the results. In *vertical* directions, no significant differences appear between the results of female and male subjects. The average resolution for signal A is about 15–17°, 19–24° for signal B and 18–23° for signal C. The maximum values can reach the double of the average value; the minimum values could be 10–50% of the mean value.

In the *horizontal* plane signal A is localized the best with an average resolution of 7–9°, signal B with 9–11° and signal C with 8–10°. In general we can support the finding that broadband sources are localized the best as well as signals with lots of high frequency information, but the differences in our measurements are relatively low: the results of signal A are only 1–2° better than results of signal C and B. The resolution for all signals is similar: differences between nearby reference points are about 10 degrees. This difference is only smaller at reference points beyond 50°, so there could be large individual differences. For example, minimum and maximum values show that signal A could be localized on the left side by the best localizer much better than by others: he was able to detect his sixth reference point at 31 degrees, but some could only distinguish between two sources within the same distance. Large overlapping minimum-maximum areas indicate problems by a real application using the averaged values.

Fig. 2 shows comparative results for Signal A, B and C from the left and the right side. Black spots correspond to new reference points (average value) between maximum and minimum values. Fig. 3 shows the same in the vertical directions.

Our data are comparable with other results from the literature. Table 3 and Table 4 contain comparative results from the median and the horizontal plane, achieved by headphone playback under the given conditions and signals.



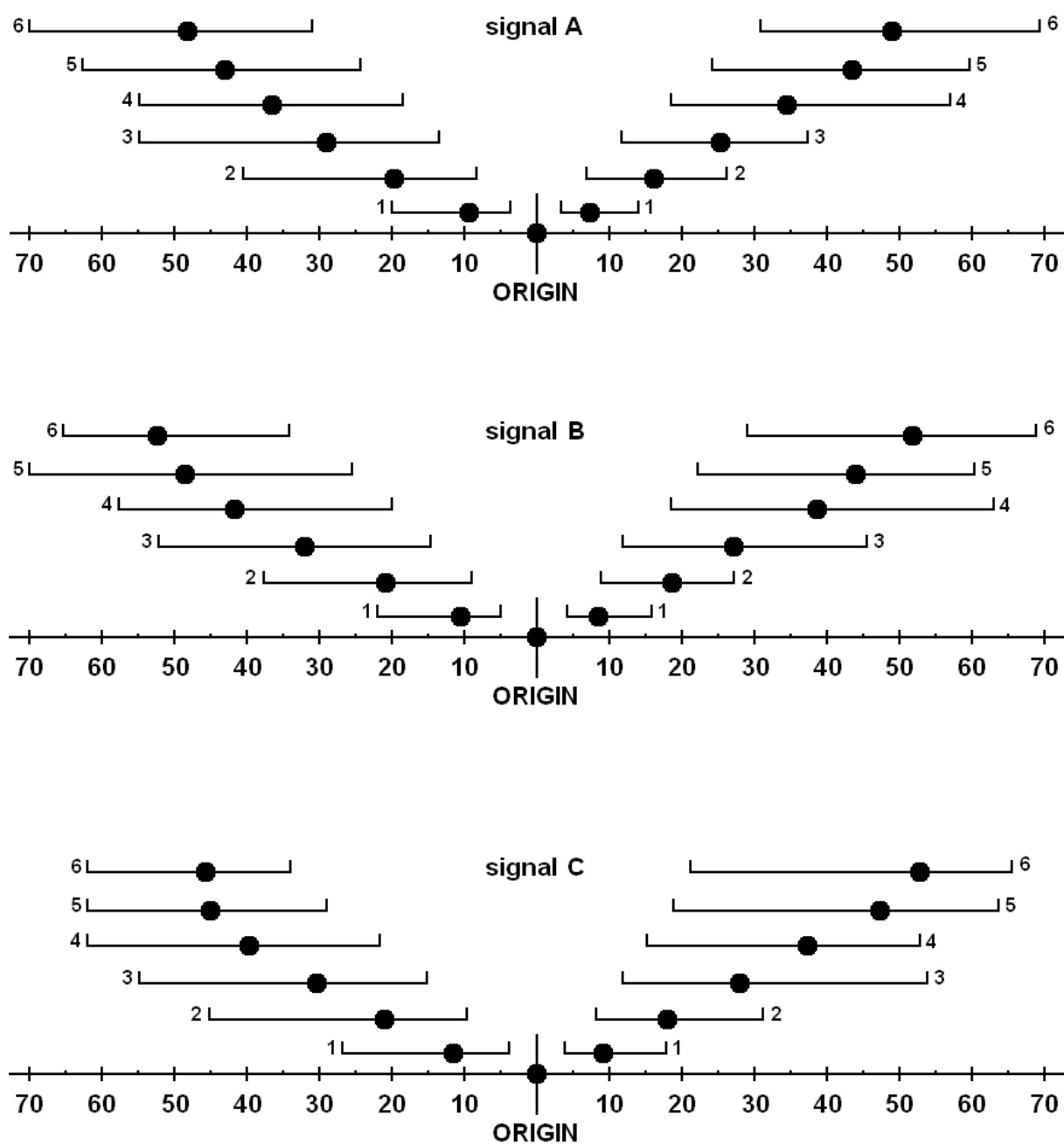


Figure 2. Representation of measured data in the horizontal plane. Black spots correspond to the average values, while solid lines to the deviations (maximum and minimum values)

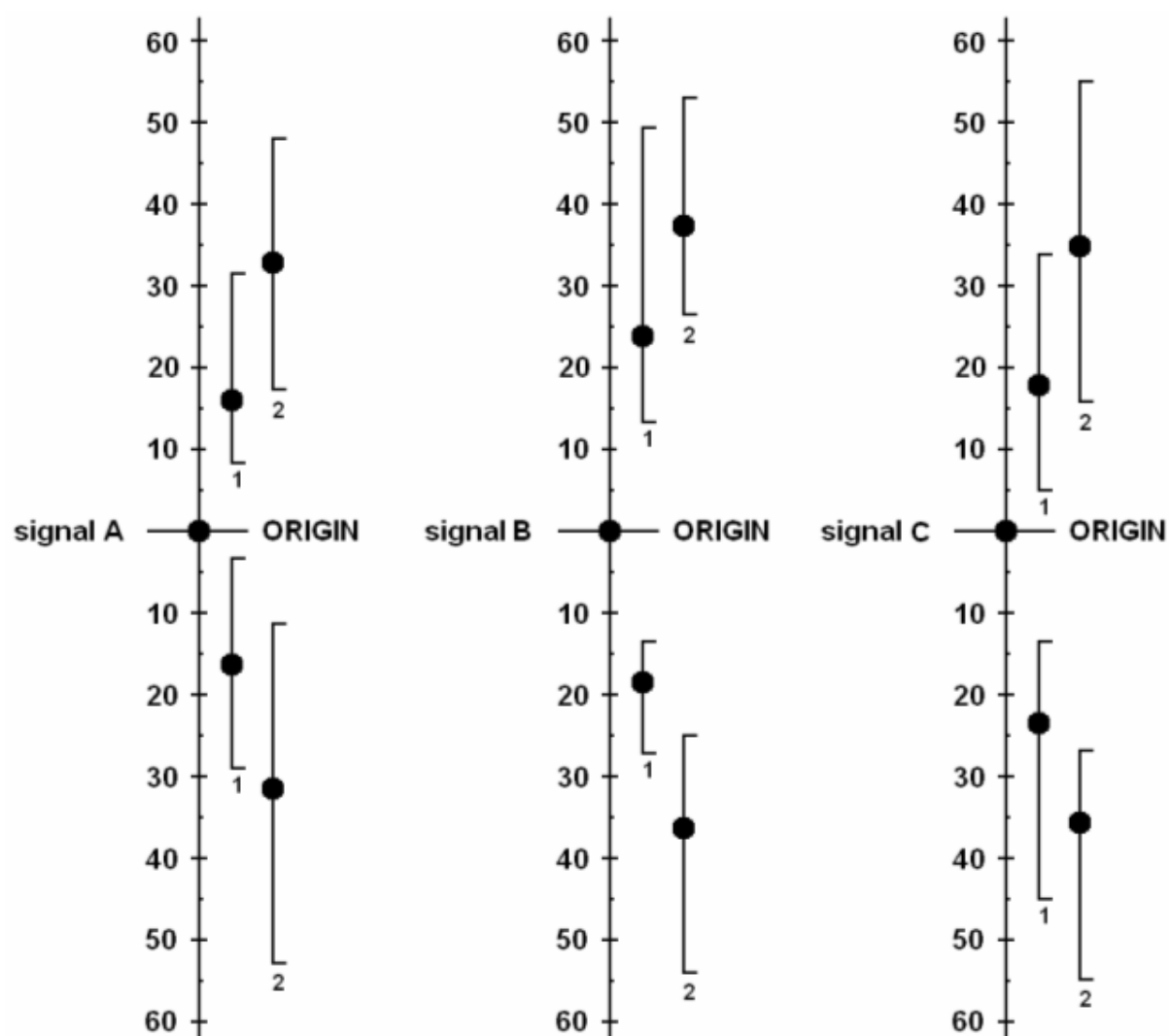


Figure 3. Representation of measured data in the median plane. Black spots correspond to the average values, while solid lines to the deviations (maximum and minimum values)

Table 3. Comparative localization results in the horizontal plane using headphones

AUTHOR	SIGNAL, REMARKS	RESULTS
Oldfield, Parker [70, 71]	azimuthal mean value	9°
	azimuth errors with HRTF filtering	4°–6°
	azimuth errors without HRTF filtering	11,9°
Wersényi	MAA values, 300 ms broadband noise, non-individual HRTFs of a good localizer	7°–10°
McKinley, Ericson [72]	average error, MAA value	5°
Middlebrooks [23]	average error, non-individual HRTFs (other-ear-condition)	17,1°
	average error, individual HRTFs (own-ear-condition)	14,7°
Duda [73]	average error with human HRTFs	4,5°
	average error for broadband signals (12kHz)	3,4°
Gardner [74]	average angle error, pink noise bursts of 250 ms	14,3°
Begault, Wenzel [75]	average error (generic HRTF)	21,7°–23°
	average error (individual HRTF)	20°
Martin [76]	average error, 328 ms noise signal	9,6°–9,7°
	maximal error	13,1°

Table 4. Comparative localization results in the median plane using headphones

AUTHOR	SIGNAL, REMARKS	RESULTS
Oldfield, Parker [70, 71]	elevational mean value	12°
	elevational error with HRTF filtering	6°–8°
	elevational error without HRTF filtering	21,9°
Wenzel, Foster [77]	non-individual HRTFs, 16 subjects lower elevations, front	ca. 24°
	lower elevations, side	ca. 23° (side)
Wightman, Kistler [54]	average error lower elevations, front	ca. 21°
	lower elevations, side	ca. 20°
McKinley, Ericson [72]	MAA value, dummy-head HRTF	30°–35°
Wersényi	MAA values, 300 ms broadband noise, non-individual HRTFs of a good localizer	15°–24°
Duda [73]	average error with human HRTFs	19,2°
	average error for broadband signals (12kHz)	17,2°
Gardner [74]	average angle error, pink noise bursts of 250 ms	34,2°
Begault, Wenzel [75]	average error (individual HRTF)	17–19°

#### 4. DISCUSSION

The goal was to determine how many virtual sound sources can be placed in the horizontal and in the median plane respectively (spatial resolution). The summarized findings are:

- As expected, localization is poorer in the median plane than in the horizontal plane.
- In the median plane, one third of the subjects could not localize the sources at all [2]
- Age, gender and computer skills do not influence the localization, but subjects wearing often headphones delivered better results.
- Broadband signals can be localized best, followed by high frequency stimulus and low frequency tones as the last, but differences are not very significant.
- The hearing system is not symmetrical: different resolution can be measured on the left and the right side as well as up and down. Decreased resolution was observed on the left side in the horizontal plane for all signals.
- The 2D virtual acoustic display is suited for replacing or extending the screen and visual information for blind and visually impaired people in case of proper mapping between acoustic and visual events, thus these results can be the basis for further GUIB applications and investigations.
- Average resolutions of  $7\text{--}11^\circ$  and  $15\text{--}24^\circ$  were measured in the horizontal plane and median plane respectively dependent on the spectral content of the signals.
- It is also suggested for a GUIB application to use broadband noisy-like sound events and/or tones with more high frequency content.

Some earcons are already available derived from the decisions of blind people. Based on these results, for a GUIB-based simulation it is recommended

- not to use vertical displacement of simulated objects, because one third of the users are not able at all to localize virtual sound sources in the median plane. One possible solution could be timbre or pitch modulation, frequency filtering based on psychoacoustic observations: signals having higher frequency components are “above”; signals with lower frequency elements are “below”,
- to segment the horizontal plane for maximum 9 source positions in a spatial resolution of 10 degrees.

Our current investigation is about to test the average spatial resolution by means of simulating steady sound sources in the positions marked with black filled dots on Figures 2 and 3 [78].

## 5. CONCLUSION

Minimum Audible Angle measurements were made in order to determine the localization blur for signals with different spectral content. 40 untrained subjects reported in a 3-category-forced choice using headphone playback and synthesized HRTFs. The Beachtron system is suited for listening tests and for low-cost solutions for everyday users: it offers real-time filtering of HRTFs, user-friendly applications and programming, headphone equalization and even individual settings of the HRTFs through the measurement of the head diameter. We found this system suitable for GUIB applications. On the other hand, the preliminary test showed that a low-cost real-time system with many efforts to a correct binaural reproduction has all kinds of headphone playback errors and vertical localization has to be increased.

The results offer new material for further investigations. We have got the average, best-case and worst-case resolution as the function of the stimulus frequency. Suggested listening tests are:

- The evaluation of the “average” spatial resolution. It is expected that not all subjects will be able to discriminate so many source locations as shown on Fig. 2 and 3 [78].
- As a special test, vertical localization can be tested using timbre or pitch modulation and by simple low-pass and high-pass filtering of the stimuli to create acoustic images “above” and “below”. It is expected that one third of the subjects will not be able to localize in the median plane without additional signal processing on the original signal.
- Special test can be made for investigating left-right and/or up-down asymmetries in connection with right and left-handed persons.
- Blind humans have to be asked to report about the excitation signals, preferred earcons and the way how the movement of the cursor has to be simulated. After the revision and extension of the set of the earcons they can be used in the simulation instead of the noise signals.
- With the help of the visually impaired in the listening test we can determine whether blind persons can better localize in a virtual audio synthesis or not.

## REFERENCES

- [1] K. Crispin, H. Petrie. Providing Access to GUI's Using Multimedia System – Based on Spatial Audio Representation. Audio Eng. Soc. 95th Convention Preprint, New York, 1993.
- [2] G. Wersényi. Localization in a HRTF-based Minimum Audible Angle Listening Test on a 2D Sound Screen for GUIB Applications. Audio Engineering Society (AES) 115th Convention Preprint, New York, USA 2003.
- [3] G. Wersényi. HRTFs in Human Localization: Measurement, Spectral Evaluation and Practical Use in Virtual Audio Environment. Ph.D. dissertation, Brandenburgische Technische Universität, Cottbus, Germany, 2002.
- [4] M. M. Blattner, D. A. Sumikawa, R. M. Greenberg. Earcons and Icons: their structure and common design principles. Human-Computer Interaction 1989, 4(1), 11–44.
- [5] K. Crispin, K. Fellbaum. Use of Acoustic Information in Screen Reader Programs for Blind Computer Users: Results from the TIDE Project GUIB, The European Context for Assistive Technology (I. Porrero, R. Bellacasa), IOS Press Amsterdam, 1995.
- [6] G. Awad. Ein Beitrag zur Mensch-Maschine-Kommunikation für Blinde und hochgradig Sehbehinderte. Ph.D. dissertation, Technical University Berlin, Berlin, 1986.

- [7] D. Burger, C. Mazurier, S. Cesarano, J. Sagot. The design of interactive auditory learning tools. *Non-visual Human-Computer Interactions* 1993, 228, 97–114.
- [8] M. O. J. Hawksford. Scalable Multichannel Coding with HRTF Enhancement for DVD and Virtual Sound Systems. *J. Audio Eng. Soc.* 2002, 50(11), 894–913.
- [9] J. Blauert. *Spatial Hearing*. The MIT Press, MA, 1983.
- [10] E. A. G. Shaw. Transformation of sound pressure level from the free-field to the eardrum in the horizontal plane. *J. Acoust. Soc. Am.* 1974, 56, 1848–1861.
- [11] S. Mehrgart, V. Mellert. Transformation characteristics of the external human ear. *J. Acoust. Soc. Am.* 1977, 61(6), 1567–1576.
- [12] D. Hammershøi, H. Møller. Free-field sound transmission to the external ear; a model and some measurement. *Proc. of DAGA'91, Bochum*, 1991, 473–476.
- [13] A. J. Watkins. Psychoacoustical aspects of synthesized vertical locale cues. *J. Acoust. Soc. Am.* 1978, 63, 1152–1165.
- [14] R. A. Butler, K. Belendiuk. Spectral cues utilized in the localization of sound in the median sagittal plane. *J. Acoust. Soc. Am.* 1977, 61, 1264–1269.
- [15] S. K. Roffler, R. A. Butler. Factors that influence the localization of sound in the vertical plane. *J. Acoust. Soc. Am.* 1968, 43, 1255–1259.
- [16] J. Blauert. Sound localization in the median plane. *Acoustica* 1969/1970, 22, 205–213.
- [17] L. R. Bernstein, C. Trahiotis, M. A. Akeroyd, K. Hartung. Sensitivity to brief changes of interaural time and interaural intensity. *J. Acoust. Soc. Am.* 2001, 109(4), 1604–1616.
- [18] D. McFadden, E. G. Pasanen. Lateralization at high frequencies based on interaural time differences. *J. Acoust. Soc. Am.* 1976, 59, 634–639.
- [19] C. B. Jensen, M. F. Sorensen, D. Hammershøi, H. Møller. Head-Related Transfer Functions: Measurements on 40 human subjects. *Proc. of 6th Int. FASE Conference, Zürich*, 1992, 225–228.
- [20] H. Møller, M. F. Sorensen, D. Hammershøi, C. B. Jensen. Head-Related Transfer Functions of human subjects. *J. Audio Eng. Soc.* 1995, 43(5), 300–321.
- [21] D. Hammershøi, H. Møller. Sound transmission to and within the human ear canal. *J. Acoust. Soc. Am.* 1996, 100(1), 408–427.
- [22] E. M. Wenzel, M. Arruda, D. J. Kistler, F. L. Wightman. Localization using nonindividualized head-related transfer functions. *J. Acoust. Soc. Am.* 1993, 94(1), 111–123.
- [23] J. C. Middlebrooks. Virtual localisation improved by scaling nonindividualized external-ear transfer function in frequency. *J. Acoust. Soc. Am.* 1999, 106(3), 1493–1510.
- [24] BEACHTRON – Technical Manual, Rev.C., Crystal River Engineering, Inc 1993.
- [25] M. A. Senova, K. I. McAnally, R.L. Martin. Localization of Virtual Sound as a Function of Head-Related Impulse Response Duration. *J. Audio Eng. Soc.* 2002, 50(1/2), 57–66.
- [26] H. Møller. Fundamentals of binaural technology. *Applied Acoustics* 1992, 36, 171–218.
- [27] H. Møller. On the quality of artificial head recording systems. *Proceedings of Inter-Noise 97, Budapest*, 1997, 1139–1142.
- [28] P. Majjala. Better binaural recordings using the real human head. *Proceedings of Inter-Noise 97, Budapest*, 1997, 1135–1138.
- [29] H. Møller, D. Hammershøi, C. B. Jensen, M. F. Sorensen. Evaluation of artificial heads in listening tests. *J. Acoust. Soc. Am.* 1999, 106(3), 83–100.
- [30] W. M. Hartmann. How we localize sound. *Physics Today* 1999, 11, 24–29.
- [31] A. Härmä, J. Jakka, M. Tikander, M. Karjalainen, T. Lokki, J. Hiipakka, G. Lorho. Augmented Reality Audio for Mobile and Wearable Appliances. *J. Audio Eng. Soc.* 2004, 52, 618–639.
- [32] H. Møller, M. F. Sorensen, C. B. Jensen, D. Hammershøi. Binaural Technique: Do We Need Individual Recordings. *J. Audio Eng. Soc.* 1996, 44(6), 451–469.
- [33] J. C. Middlebrooks. Narrow-band sound localization related to external ear acoustics. *J. Acoust. Soc. Am.* 1992, (92), 2607–2624.

- [34] H. Fisher, S. J. Freedman. The role of the pinna in auditory localization. *J. Audiol. Research* 1968, 8, 15–26.
- [35] W. M. Hartmann, B. Rakerd. On the minimum audible angle – A decision theory approach. *J. Acoust. Soc. Am.* 1989, 85, 2031–2041.
- [36] T. Z. Strybel, C. L. Manlingas, D. R. Perrott. Minimum Audible Movement Angle as a function of azimuth and elevation of the source. *Human Factors* 1992, 34(3), 267–275.
- [37] D. R. Perrott, A. D. Musicant. Minimum auditory movement angle: binaural localization of moving sources. *J. Acoust. Soc. Am.* 1977, 62, 1463–1466.
- [38] J. Zwislocki, R. S. Feldman. Just noticeable differences in dichotic phase. *J. Acoust. Soc. Am.* 1956, 28, 860–864.
- [39] P. A. Campbell. Just noticeable differences of changes of interaural time differences as a function of interaural time differences. *J. Acoust. Soc. Am.* 1959, 31, 917–922.
- [40] D. R. Begault, E. Wenzel, M. Anderson. Direct Comparison of the Impact of Head Tracking Reverberation, and Individualized Head-Related Transfer Functions on the Spatial Perception of a Virtual Speech Source. *J. Audio Eng. Soc.* 2001, 49(10), 904–917.
- [41] A. Harma, J. Jakka, M. Tikander, M. Karjalainen, T. Lokki, J. Hiipakka, G. Lorho. Augmented Reality Audio for Mobile and Wearable Appliances. *J. Audio Eng. Soc.* 2004, 52(6), 618–639.
- [42] J. Blauert. Localization and the law of the first wavefront in the median plane. *J. Acoust. Soc. Am.* 1971, 50, 466–470.
- [43] M. Cohen, E. Wenzel. The design of Multidimensional Sound Intrefaces. In W. Barfield, T.A. Furness III (Editors) “Virtual Environments and Advanced Interface Design” Oxford Univ. Press, New York, 1995, Oxford, 291–346.
- [44] J. Sandvad, D. Hammershøi. Binaural auralization. Comparison of FIR and IIR filter representation of HIR. *Proc. of 96th Convention of the Audio Eng. Soc.*, Amsterdam, 1994.
- [45] M. Kleiner, B. I. Dalenbäck, P. Svensson. Auralization — an overview. *J. Audio Eng. Soc.* 1993, 41, 861–875.
- [46] K. D. Jacob, M. Jorgensen, C. B. Ickler. Verifying the accuracy of audible simulation (auralization) systems. *J. Acoust. Soc. Am.* 1992, 92, p. 2395.
- [47] J. Blauert, H. Lehnert, J. Sahrhage, H. Strauss. An Interactive Virtul-environment Generator for Psychoacoustic Research I: Architecture and Implementation. *Acoustica* 2000, 86, 94–102.
- [48] D. R. Begault, 3-D Sound for Virtual Reality and Multimedia. Academic Press, London, UK, 1994.
- [49] K. Brinkmann, U. Richter. Zur Messunsicherheit bei psychoakustischen Messungen. *Proc. of DAGA’87*, Aachen, 1987, 593–596.
- [50] K. Crispian, H. Petrie. Providing Access to Graphical-Based User Interfaces for Blind People: Using Multimedia System Based on Spatial Audio Representation. 95th AES Convention, *J. Audio Eng. Soc.* (Abstracts), 1993, 41, p. 1060.
- [51] E. Mynatt, W. K. Edwards. Mapping GUIs to Auditory Interfaces. *Proc. ACM Symposium on User Interface Software Technology*, Monterey, November 1992, 61–70.
- [52] E. Mynatt, G. Weber. Nonvisual Presentation of Graphical User Interfaces: Contrasting Two Approaches. *Proc. 1994 ACM Conference on Human Factors in Computing Systems*, Boston, April 1994, 66–172.
- [53] S. H. Foster, E. M. Wenzel. Virtual Acoustic Environments: The Convolvotron. Demo system presentation at SIGGRAPH’91, 18th ACM Conference on Computer Graphics and Interactive Techniques, Las Vegas, NV, ACM Press, New York, 1991.
- [54] F. L. Wightman, D. J. Kistler. Headphone Simulation of Free-Field Listening I.–II. *J. Acoust. Soc. Am.* 1989, 85, 858–878.

- [55] M. Matsumoto, S. Yamanaka, M. Tohyama, H. Nomura. Effect of Arrival Time Correction on the Accuracy of Binaural Impulse Response Interpolation. *J. Audio Eng. Soc.* 2004, 52(1/2), 56–61.
- [56] F. P. Freeland, L. W. P. Biscainho, P. S. R. Diniz. Interpositional Transfer Function for 3D-Sound Generation. *J. Audio Eng. Soc.* 2004, 52(9), 915–930.
- [57] P. Minnaar, J. Plogsties, F. Christensen. Directional Resolution of Head-Related Transfer Functions Required in Binaural Synthesis. *J. Audio Eng. Soc.* 2005, 53(10), 919–929.
- [58] S. E. Olive. Differences in Performance and Preference of Trained versus Untrained Listeners in Loudspeaker Tests: A Case Study. *J. Audio Eng. Soc.* 2003, 51(9), 806–825.
- [59] V. R. Algazi, C. Avendano, R. O. Duda. Estimation of a spherical-head model from anthropometry. *J. Audio Eng. Soc.* 2001, 49(6), 472–479.
- [60] E. Zwicker, R. Feldtkeller. *Das Ohr als Nachrichtenempfänger*. S. Hirzel Verlag, Stuttgart, 1967, p. 181.
- [61] R. A. Butler, R. F. Naunton. Role of stimulus frequency and duration in the phenomenon of localization shifts. *J. Acoust. Soc. Am.* 1964, 36(5), 917–922.
- [62] D. R. Perrott, J. Tucker. Minimum Audible Movement angle as a function of signal frequency and the velocity of the source. *J. Acoust. Soc. Am.* 1988, 83, 1522–1527.
- [63] W. Mills. On the minimum audible angle. *J. Acoust. Soc. Am.* 1958, 30, 237–246.
- [64] M. Kinkel, B. Kollmeier. Diskrimination interauraler Parameter bei Schmalbandrauschen. *Proc. of DAGA'87*, Aachen, 1987, 537–540.
- [65] J. L. Hall. Minimum detectable change in interaural time or intensity difference for brief impulsive stimuli. *J. Acoust. Soc. Am.* 1964, 36, 2411–2413.
- [66] D. W. Grantham. Detection and discrimination of simulated motion of auditory targets in the horizontal plane. *J. Acoust. Soc. Am.* 1986, 79, 1939–1949.
- [67] J. M. Chowning. The simulation of Moving Sound Sources. *J. Audio Eng. Soc.* 1971, 19, 2–6.
- [68] S. M. Abel, C. Giguere, A. Consoli, B. C. Papsin. Front/Back Mirror Image Reversal Errors and Left/Right Asymmetry in Sound Localization. *Acoustica* 1999, 85, 378–389.
- [69] F. Chen. Localization of 3-D Sound Presented through Headphone - Duration of Sound Presentation and Localization Accuracy. *J. Audio Eng. Soc.* 2003, 51(12), 1163–1171.
- [70] S. R. Oldfield, S. P. A. Parker. Acuity of sound localisation: a topography of auditory space I-II. *Perception* 1984, 13, 581–617.
- [71] S. R. Oldfield, S. P. A. Parker. Acuity of sound localisation: a topography of auditory space III. *Perception* 1986, 15, 67–81.
- [72] R. L. McKinley, M. A. Ericson. Flight Demonstration of a 3-D Auditory Display. In *Binaural and Spatial Hearing in Real and Virtual Environments* (edited by R.H. Gilkey and T.R. Anderson), Lawrence Erlbaum Ass., Mahwah, New Jersey, 1997, 683–699.
- [73] R. O. Duda. Elevation Dependence of the Interaural Transfer Function. in *Binaural and Spatial Hearing in Real and Virtual Environments* (edited by R. H. Gilkey and T. R. Anderson), Lawrence Erlbaum Ass., Mahwah, New Jersey, 1997, 49–75.
- [74] W. G. Gardner, *3-D Audio Using Loudspeakers*. Kluwer Academic Publ., Boston, 1998.
- [75] D. R. Begault, E. Wenzel, M. Anderson. Direct Comparison of the Impact of Head Tracking Reverberation, and Individualized Head-Related Transfer Functions on the Spatial Perception of a Virtual Speech Source. *J. Audio Eng. Soc.* 2001, 49(10), 904–917.
- [76] R. L. Martin, K. I. McAnally, M. A. Senova. Free-Field Equivalent Localization of Virtual Audio. *J. Audio Eng. Soc.* 2001, 49(1/2), 14–22.
- [77] E. M. Wenzel, S. H. Foster. Perceptual consequences of interpolating head-related transfer functions during spatial synthesis. *Proceedings of the ASSP Workshop on Applications of Signal Processing to Audio and Acoustics*, New York, USA, 1993.
- [78] G. Wersényi. What Virtual Audio Synthesis Could Do for Visually Disabled Humans in the New Era? *Proceedings of 12th AES Regional conference*, Tokyo, 2005, 180–183.