Citation:

WERSÉNYI, GY.,
Acoustic Signal Processing for Listening Tests in Virtual Audio.

Proceedings of the 2001 Polish-Czech-Hungarian Workshop on Circuit Theory, Signal Processing, and Telecommunication Networks. Budapest, Hungary, 14-17 September 2001., pp. **175-181**.

*"2001 Polish-Czech-Hungarian Workshop on Circuit Theory,
Signal Processing, and Telecommunication Networks",
Budapest, Hungary, 14-17 Sep. 2001*

Acoustic Signal Processing for Listening Tests in Virtual Audio

WERSÉNYI, György
University of Technology and Economics, Budapest, Hungary
Technical University of Brandenburg, Cottbus, Germany
wersenyi@sparc.core.hu

Abstract

Working with virtual audio representations using headphones with the reproduction of the correct spatial and directional information is a difficult task. There are lots of different systems with common signal processing algorithms, like equalization of the headphones, canceling the undesired transfer functions, and the reproduction of the directional filtering of the human outer ears. All of these are basically made in the time-domain with real-time convolution of FIR-filter sets.
This paper presents the possibilities and performance of a low-cost system based on the BEACHTRON™ DSP card during listening tests for measuring the localization blur in a virtual audio environment. Some preliminary results are shown to compare the spatial resolution and interesting evaluation of directional judgments of a moving source with this system.

I. Introduction

Virtual Acoustic Displays (VAD) are widely used. Flight-simulators, computer games or blind computer users need the possibility to move and act in virtual environments with or without visual help of the eyes. These acoustic displays produce sound waves from different directions. The goal is to get the best spatial resolution, discrimination, and sound quality during the playback. The sound sources have to be identified and localized precisely.
This is especially important for blind PC users, who do not have the visual display and the common used Graphical User Interfaces (like MS-Windows). The former GUIB-Project (Graphical User Interface for Blind Persons) had the goal to develop the proper playback media to replace the screen [1].
During the creation of an acoustic virtual environment two basic questions have to be answered: how can we make the best mapping between visual events and acoustic sounds, and how is the spatial resolution during the playback.
Generally a 3D VAD can be explained so that the information „distance“ or „depth“ will be realized by the volume of the sound (overlapping windows). In case of a 2D VAD users are able to adjust the volume, so it is not necessary to work with distance hearing.

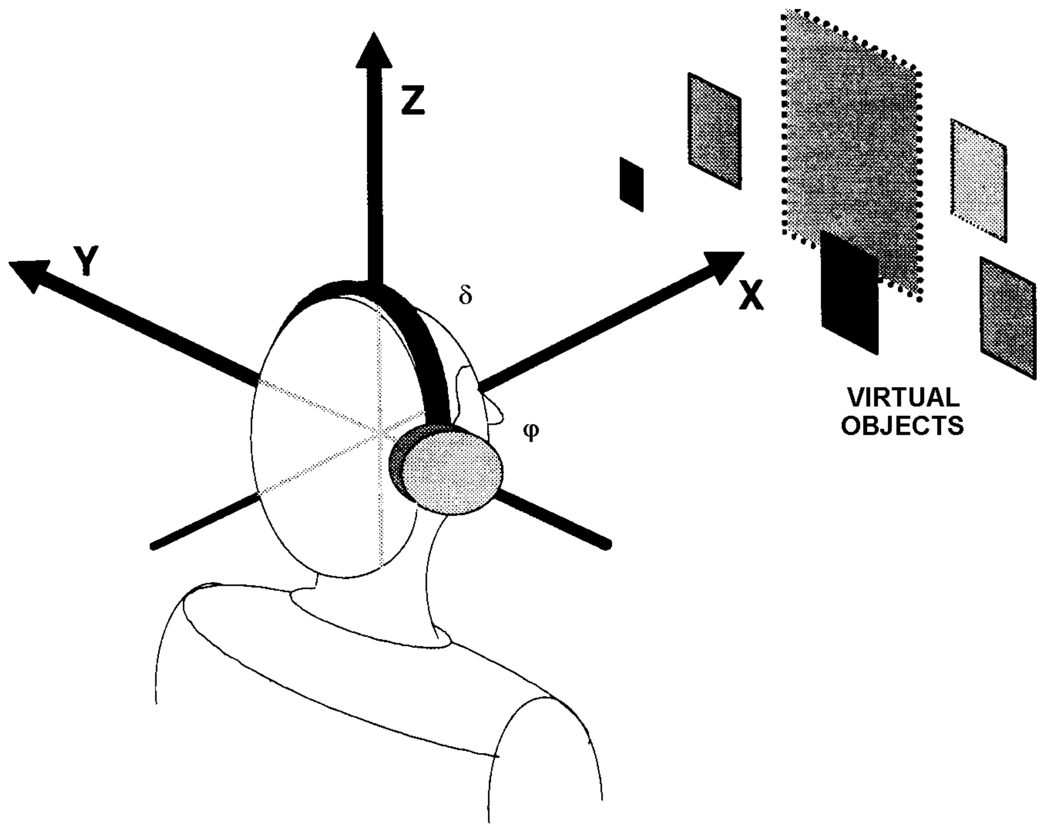


Fig.1. 2D acoustic display parallel with the frontal plane. The origin is in the „forward“ direction φ=δ=0°. Virtual sources in this measurement move left, right, up or down from the origin.

The Head-Related Transfer Functions (HRTFs) are complex functions describing the transmission from the free-field to the eardrum. This directional dependent filtering is the basis of the human spatial and directional hearing [2, 3, 4] (see Fig.2.).

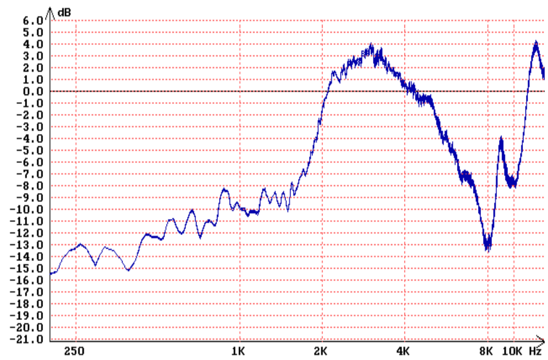


Fig.2. Typical HRTF in the frontal direction of a dummy-head [7].

Every human has his/her own individual HRTFs. The best way to reproduce exact spatial information through headphones is to measure and synthesize a person's own HRTFs. A cheaper and easier way is the use of the HRTFs of a good localizer. Of course, for the proper binaural reproduction headphone equalization is required [5]. The Head-Related Impulse Responses (HRIR) are the time-domain equivalent of the HRTFs. They can be measured with impulse excitation or calculated by Invers-Fourier transforming the HRTFs. It is possible to make a convolution between the input signals and the HRIRs in real-time.

The playback medium could be loudspeaker or headphones. Loudspeakers are insufficient, because they are large, disturb the environment, and even more channel systems have decreased spatial resolution. Headphones are better suited but head movements and the HRTF filtering effects are missing. Playback through headphones may results in well-known errors like front-back-confusion, elevation shift, or sources too near to the listener independent of the signal processing [5, 6].

II. Psychoacoustic measurements

Psychoacoustic measurements can be made basically in two different ways. During the so called „absolute“ measurements the subjects have to point on the sound source and identify its location. This is for loudspeaker playback. The question relating to **localization** is: where is the sound source? The other solution is to search for the Minimum Audible Angle (MAA) or the Just Noticeable Difference (JND), where the subjects only have to compare two sound sources and identify only the change of the source direction. The question relating to **localization blur** is: what is the smallest change of the position of the sound source that produces a noticeable change of the auditory event?

The localization performance depends on many parameters. First of all on the monaural and interaural parameters. In the median plane no interaural differences are present so the localization is made on the HRTF filtering only. The same signal reaches both ears. Outside the median plane Interaural Intensity Differences and Interaural Time Delays appear together with the HRTF filtering which results in an increased localization performance (spatial resolution). The spectral shape, bandwidth, cut-off frequency of the signal also influences the localization. Broadband signals and high frequency components are better to localize. Increase of the volume and the duration increase the localization performance: sound sources between 40-80 dB SPL and signals over 250 ms are localized the best [6].

The localization of a human is time variant. It needs adaptation, learning phases, and it is influenced by prior expectation and fatigue as well.

III. Measurement setup and comparing of the results

III.1. SET UP

Based on former results of the GUIB-Project we are going to use headphone playback system 2D flat virtual acoustic horizon in the front of the listener and simple signals [1].

The measurement software is running on a PC under MS-DOS. The real-time HRTF filtering is made by the Beachtron™ DSP Card. 72 measured HRTFs for each ear are present from the measurement of *Wightman and Kistler* from a good localizer person between –36° and +54° in elevation and in every 30° horizontal direction [8]. The rest of the HRTFs are interpolated from the nearest four. They are stored as minimalphase FIR-Filters of 75 points and the DSP card makes the real time convolution in the time-domain with the monaural sound files in 44100 Hz and 16 bit resolution [9, 10]. The HRTFs can be set more individually by measuring the distance between the earcanal entrances of the subjects. One Beachtron can handle with two sound sources at a time but 8 can be connected in cascade. The output socket and the software playback are optimized and equalized for the open-dynamic circumauaral Sennheiser HD540 headphone. This system only needs 33 MHz CPU speed and can programmed in high-level instructions of C++ language.

III.2. METHOD

We decided to use a square 2D virtual surface in the front of the listener. The sound sources can be placed only in the horizontal or in the median plane in 1º resolution. The distance of the virtual sources from the listener is not constant. The test signals are 300 ms of white noise, 1500 Hz low-pass and 7000 Hz high-pass filtered version of the white noise.
During the MAA measurement the task is to discriminate the reference source from the moving source in a so called "three-categorie-forced-choice". The possible answers are: „no difference" if the subject is not able to discriminate the sources and they seem to come from the same direction. „Different sound sources" means that he or she is able to distinguish between the signals. He or she may have the option of choosing „uncertain" as the answer if he or she is not sure which is the case.

III.3

RESULTS

Before measuring the MAA values a simple test was made with a female speech signal. The average SPL at the eardrum for the "most comfortable volume" is 58,2 dB (measured with the head and torso simulator). Answers from 25 subjects were evaluated as follows.

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

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

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

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

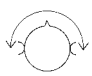
In both cases 80% of the subjects reported in-the-head localization.

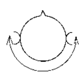
III. The source is moving around the head to the left in the horizontal plane.

"How does the sound source move?" BEHIND THE HEAD: 43%
AROUND THE HEAD TO THE LEFT: 50%
OTHER: 7%

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

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

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

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

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

IV. The source is moving *up*.
"Direction of the moving?" UP: 92% DOWN: 8%
"Reverse direction is possible?" YES: 8% NO: 92%

V. The source is moving *down*.
"Direction of the moving?" UP: 12% DOWN: 88%
"Reverse direction is possible?" YES: 41% NO: 59%

69% of the subjects had front-back confusion. If the source is in the front, only one third was able to localize the source at its correct position. The subjects were very confused and undecided due to in-the-head localisation, and 58% believed the reverse direction of their answers as well. If the source was in the back, there was almost no confusion and the uncertainity decreased.
During the third section, the source was always moving around the head in the horizontal plane, only the questions were different in order to see if the subjects could be influenced. Many of the subjects reported the source moving behind the head (in the back hemisphere). 58% belived that the turning-movement is above the horizontal plane (elevation shift). Only 11% was able to detect the source moving in the frontal hempishpere. It is interesting that for 58% this movement was acceptable as a moving source in the frontal plane. Changing of the elevation was suprisingly easy to recognise, mostly up.
All of the subjects were easily influenced and they reported all kinds of answers by the same signal reproduction, which suggests low quality localization in the median plane and the existence of all well-known errors during headphone playback.

Table 1. and Table 2. contain results from the median and the horizontal plane comparing our preliminary results from 7 subjects with other tests. The exhaustive subjective listening test will be performed in the future with 40 subjects.

AUTHOR	SIGNAL	SPATIAL RESOLUTION
Blauert [6]	all signals, absolute minimum	1°
Haustein, Schirmer [11]	Broadband noise	3,2°
	100 ms white noise Impulse	± 3,6° forward ± 9 to 10° left/right
Heffner, Heffner [12]	Noise (MAA)	1,3°-1,8° forward 9°-10° left/right
Oldfield, Parker [13, 14]	φ = 0°-80°	4°-6°
Middlebrooks [15]	AVG Error of 150 ms broadband noise	5,8°
Middlebrooks [16]	AVG Error with individual HRTFs AVG Error with individual HRTFs	14,7° 17,1°
Wersényi	AVG for white noise impulse (MAA) AVG for 1500 Hz low pass filtered white noise impulse (MAA)	8,8° 10,7°

Table 1. Localization in the horizontal plane

AUTHOR	SIGNAL	SPATIAL RESOLUTION
Blauert [6]	white noise	4° forward
Wettschurek [17]	MAA for white noise (8-10° Standard Deviation)	± 4° forward ± 10° up
	Low-pass noise with 4 kHz cut-off frequency	± 8° forward ± 20° up
Oldfield, Parker [13, 14]	φ = 0°-80°	6°-8°
Wenzel, Foster [18]	AVG Error	ca. 24° forward
Wightman, Kistler [8]	AVG Error	ca. 21° forward
Middlebrooks [15]	AVG Error of 150 ms broadband noise	5,7°
Wersényi	AVG for white noise impulse (MAA) AVG for 1500 Hz low pass filtered white noise impulse (MAA)	16,5° 18,3°

Table 2. Localization in the median plane.

IV. Conclusion

During the preliminary phase of the listening test we measured the localization blur of 7 subjects in a virtual audio environment. The input signal is convolved in real-time with the needed HRTF directional filtering. Results were compared with former results from other tests. The average localization blur for white noise impulse is 8,8° and 16,5° in the horizontal plane and in the median plane respectively.

The directional judgments of the subjects show that well-known errors of the headphone playback are present. In-the-head localization and front-back instances of confusion are more significant than elevation shifts. The subjects can be very easily influenced by their eyes and will. This phenomenon is independent of the signal processing and suggests alternative headphone design and elementary problems with headphone playback systems [19].

V. REFERENCES

[1] K. CRISPIEN, H. PETRIE, „Providing Access to GUI's Using Multimedia System – Based on Spatial Audio Representation", AES 95th Convention Preprint, New York, 1993.
[2] J. C. MAKOUS, J. C. MIDDLEBROOKS, „Two-dimensional sound localization by human listeners", *J. Acoust. Soc. Am.*, **87(5)**, pp. 2188-2200, 1990 May.
[3] S. HEHRGART, V. MELLERT, „Transformation characteristics of the external human ear", *J. Acoust. Soc. Am.* **61(6)**, pp.1567-1576, 1977 June.
[4] H. MOLLER, M. F. SORENSEN, D. HAMMERSHOI, C. B. JENSEN, „Head-Related Transfer Functions of human subjects", *J. Audio Eng. Soc.* **43(5)**, pp. 300-321, 1995.
[5] H. MOLLER, „Fundamentals of binaural technology", *Applied Acoustics* **36**, pp. 171-218, 1992.
[6] J. BLAUERT, *Spatial Hearing,* The MIT Press, MA, 1983.
[7] GY. WERSENYI, „Measurement system upgrading for more precise measuring of the Head-Related Transfer Functions", in *Proceedings of Inter-Noise 2000*, Vol.II., Nice, pp. 1173-1176, 2000.
[8] F.L. WIGHTMAN, D.J. KISTLER, „Headphone Simulation of Free-Field Listening I.-II.", *J. Acoust. Soc. Am.* **85**., pp. 858-878, 1989 February.
[9] Crystal River Engineering, Inc. : BEACHTRON – Technical Manual, Rev.C., 1993.
[10] S.H. FOSTER, E.M. WENZEL, „Virtual Acoustic Environments: The Convolvotron", Demos system presentation at SIGGRAPH'91, 18th ACM Conference on Computer Graphics and Interactive Techniques, Las Vegas, NV (ACM Press, New York), 1991.
[11] B.G. HAUSTEIN, W. SCHIRMER, „Messeinrichtung zur Untersuchung des Richtungslokalisationsvermögens", Hochfrequenztech. und Elektroakustik **79**, pp. 96-101, 1970.
[12] R.S. HEFFNER, H.E. HEFFNER, "Sound localization acuity in the cat: Effect of azimuth, signal duration and test procedure", Hear. Res. **36**, pp. 221-232, 1988.
[13] S.R. OLDFIELD, S.P.A. PARKER, „Acuity of sound localisation: a topography of auditory space I-II.", Perception **13**, pp. 581-617, 1984.

[14] S.R. OLDFIELD, S.P.A. PARKER, „Acuity of sound localisation: a topography of auditory space III.", Perception **15**, pp. 67-81, 1986.
[15] J.C. MIDDLEBROOKS, „Spectral Shape Cues for Sound Localization", in Binaural and Spatial Hearing in Real and Virtual Environments (edited by R.H. Gilkey and T.R. Anderson), Lawrence Erlbaum Ass., Mahwah, New Jersey, pp. 77-97, 1997.
[16] J. C. MIDDLEBROOKS, „Virtual localisation improved by scaling nonindividualized external-ear transfer function in frequency", *J. Acoust. Soc. Am.* **106(3)**, pp. 1493-1510, 1999.
[17] R. WETTSCHUREK, „Die absoluten Unterschiedswellen der Richtungswahrnehmung in der Medianebene beim natürlichen Hören sowie beim Hören über ein Kunstkopf-Übertragungssystem", Acoustica **28**, pp. 197-208, 1973.
[18] E.M. WENZEL, S.H. FOSTER, „Perceptual consequences of interpolating head-related transfer functions during spatial synthesis", Proc. of the ASSP Workshop on Applications of Signal Processing to Audio and Acoustics, New York, IEEE Press, 1993.
[19] F. M. KÖNIG, "A new supra-aural Dynamic Headphone System for in-front localization and surround reproduction of sound", AES Convention Preprint **4495**, München, 1997.