SIMULATION OF SMALL HEAD-MOVEMENTS ON A VIRTUAL AUDIO DISPLAY USING HEADPHONE PLAYBACK AND HRTF SYNTHESIS

György Wersényi

Széchenyi István University
Department of Telecommunications
Egyetem tér 1, H-9024, Győr, Hungary
wersenyi@sze.hu

ABSTRACT

Correct determination of sound source location often fails during headphone playback in virtual simulation. Among other cues, small movements of the head are considered to be important in free-field listening to avoid in-the-head localization and/or front-back reversals. Up-to-date virtual reality simulators are able to locate the head's actual position, and through proper feedback, real-time update of the actual HRTFs can be realized for a better spatial simulation. This study uses the BEACHTRON sound card and its HRTFs for simulating small head-movements by randomly moving the simulated sound source to emulate head movements. This method does not need any additional equipment or feedback. Results of a listening test with 50 subjects demonstrate the applicability of this procedure focusing on resolving in-the-head localization and front-back reversals. The investigation was made on the basis of the former GUIB (Graphical User Interface for Blind persons) project.

[Keywords: Spatialization, HRTF, localization error, GUIB]

1. INTRODUCTION

The former GUIB Project (Graphical User Interface for Blind Persons) was focused on creating a virtual audio display (VAD) for the elderly and the visually disabled [1, 2]. These individuals do not have the possibility to use graphical user interfaces and icons, and they need special tools if they want to use personal computers. This project included a number of experiments, such as finding the proper mapping between icons or events on the screen and sound samples (called Earcons), possibilities of different input media (Touch screen, Braille keyboards), and evaluation of playback systems [3-5]. First, a multi-channel loudspeaker array was tested and was found to be inappropriate. Subsequently, headphone playback through HRTF filtering was applied. Both methods used the BEACHTRON sound card for simulation. Although the GUIB project ended years ago, some psychoacoustic measurements have been made with this system. Those investigations focused e.g., on headphone playback errors, localization blur and spatial resolution of the VAD.

2. HEAD-TRACKING AND VIRTUAL LOCALIZATION

The purpose of our current investigation is to find tools to improve the localization performance with the system mentioned above. Leading to this investigation, we tested additional high-pass and low-pass filtering of sound sources to bias correct localization judgments in the median plane [6].

One of the main goals of this study is to decrease front-back reversals and/or in-the-head localization rates. It is well known that during headphone playback these errors influence localization [7-14]. State-of-the-art multimedia virtual simulators use head-tracking devices, simulation of room reverberation and different methods to create the best fitting HRTF set [15-19]. We focus on head-tracking that has been shown to be important for reducing such errors [20-23]. Furthermore, *small* head movements (often unwanted) of about 1-3 degrees could influence in-the-head localization in free-field listening through small changes in the interaural differences. We assumed that such minute of head movements could reduce in-the-head localization and the small changes in the interaural level and time differences may lead to better results. Dynamic changes through intentionally or unintentionally movements of the head in this order can be relevant [14, 24, 25].

State-of-the-art methods use headphones with a head-tracking device. Such a device has some sort of a feedback and additional hardware (e.g. laser pointer and receivers); typically they also require considerable computational resources. In the simulation it is possible to change the HRTFs dynamically, in order to create a correct spatial event, and compute the appropriate HRTFs synchronized to the listener's head-movements.

In contrast, our system is built on different methods. We try to find out how small head movements influence virtual localization. Instead of moving the head, using feedback and additional equipment, we simply simulate these movements by moving the virtual sound source. This is achieved by small changes in the HRTFs that are not synchronized with the actual position of the head. The goal of the investigation is to explore whether simulation and changes in HRTFs can replace additional hardware and head-tracking devices. These changes in the HRTFs are about 1-4 degrees, in order to simulate only small movements of the head, and to investigate the influence on in-the-head localization and reversals rates.

3. MEASUREMENT SETUP

The virtual audio display is simulated in front of the listener as a 2D sound screen as seen on Fig.1. The BEACHTRON system uses the HRTFs of a "good localizer" from measurements by *Wightman and Kistler* [26-30]. Real-time filtering is made in the time-domain by the HRIRs. Equalization for the Sennheiser

HD540 headphone was also included. Furthermore, it is possible to set the head-diameter to obtain a better interaural time difference simulation.

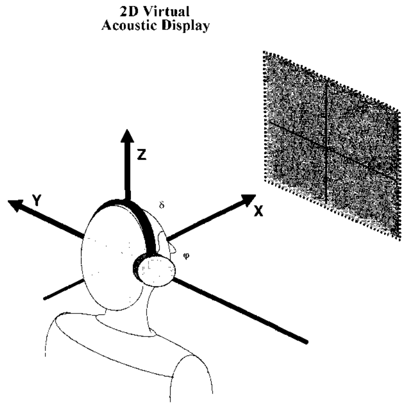

Figure 1. *Illustration of the 2D VAD. The acoustic surface is parallel with the Z-Y-plane. The origin (the reference location) is in the front of the listener ($\varphi=\delta=0°$).*

The investigation was made in an anechoic chamber. Fifty young adults, university students, participated.

The simulation program sets the virtual sound source in the "front" direction (φ=δ=0º), which we considered as the target source location. Without the simulation of head-movements, this is a stationary source, a reference condition. During simulation, this sound source moved randomly by way of changing the following parameters:

(A): direction and extent of the movement (from 0º to 10º) both horizontal and vertical. In case of A=0, no movement will be simulated creating the reference condition of a stationary source.

(B): number of new locations (the number of times the source location is changed, 1 to 100)

(C): presentation per location (the number of times the stimulus is presented in one location, 1 to 1000).

After setting these parameters, white noise signal of 10 ms was played back. The length of the simulation is

$$\text{Total time} = B*C*10 \text{ (ms)}. \quad (1)$$

For example, by setting A=2, B=50 and C=5 the following simulation could be made. A random generator calculates an actual source location within ±2º degrees of the origin that includes (-2, -1, 0, +1, +2) in all directions (see Fig.2). These points represent potential source locations. With B=50, fifty source locations will be determined and in each location the sound file will be presented five times (50 ms). Because the possible number of locations is 25, B=50 means that all of them will be selected twice in a random order. By reducing the number of C and increasing the number of B, we can simulate faster head-movements.

During the simulation subjects are asked to report
- whether the perceived location is in the head
- front-back reversals and
- whether they experience the percept of a stationary or a moving source (perception of movement). This latter question is a control, because our goal is to simulate a sound source that *appears to be* steady, and thus we would like the subjects not to detect any movement.

We assumed that about 1-3 degrees of random movement will be perceived as a stationary source. At the start of the experiment, all subjects were exposed to the reference condition where A=0, corresponding to a stationary source in front of them, followed by stimuli with different A, B and C parameters.

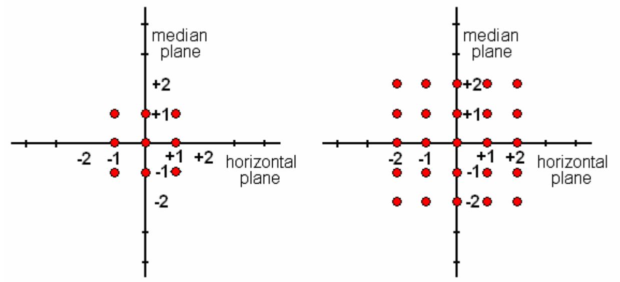

Figure 2. *Simulation for parameters A=1 (left) and A=2 (right). Total number of simulated source positions is $(2A+1)^2$.*

During evaluation, subjects answered the following questions: "Is the sound source externalized or in-the-head?", "Where is the simulated sound source in the virtual space?" and "Do you have the percept of a moving source?". Results were filled in a table (see Table 2 and 3).

4. DISCUSSION

Results are presented here only for parameters B=100 and C=50. Setting A=0 represents a sound source located at the origin in front on the listener. In this case, the same sound source location was selected 100 times and the sound file was repeated 50 times. These values correspond to a relatively slow simulation. By increasing parameter A, the 100 simulated sound source positions were equally distributed among a number of $(2A+1)^2$ source positions. For small A's, more repeats were performed and each source locations was used several times (see Table 1.).

A	Nr. of simulated sources	B	C	Maximal length of simulation [sec]	Possible number of repeats
0	1 (origin)	100	50	50	(100)
1	9	100	50	50	11,11
2	25	100	50	50	4
3	49	100	50	50	2,04
4	81	100	50	50	1,23
5	121	100	50	50	0,82
6	169	100	50	50	0,59
7	225	100	50	50	0,44

Table 1: Different settings and values of parameter A during simulation.

When A=2, the number of possible source locations is 25 (see Fig.2.). By setting B=100, we decided to use each location four times during the simulation. As a consequence, when A is greater than 5, the number of possible source locations is higher than B and only part of all possible source locations could be presented. In that case, a simulation of 50 seconds was too long and the subjects responded before the end of the trial. The actual length of the stimulus is thus not crucial, because it is a consequence of parameter C. Using a 10 ms sound file and a number of 50 for parameter C is the same as using a 50-ms sound file and a value of 10 for parameter C.

Parameter A was increased throughout the experiment and it terminated when subjects reported the percept of a moving sound source (by answering “yes” to the third question). First, they were exposed to the reference situation (A=0). Second, parameter A was set to 1, and the listening test was repeated etc. The answers were filled in the tables. Yellow fields indicate the perception of movement, so the simulation was stopped after that. This means, subjects exceeded the limit of the individual localization blur that influences the measurement and the evaluation [31].

Table 2 shows the results for in-the-head localization. N stays for externalized virtual source (no error) and Y stays for existing in-the-head localization. The first row in the table shows values of parameter A. For example, subject 23 had in-the-localization for the stationary source as well for the moving source in 1 degree steps. As we used A=2, he reported an externalized virtual source without perception of the movement. By A=3 he perceived the movement.

At the evaluation subjects could be classified in the following sets:

- subjects, where the simulation of head movements did help to resolve in-the-head localization (first they have it, later they do not). E.g. subject nr. 23.
- subjects, where the simulation of head movement did not help by resolving in-the-head localization (they have it from the beginning and also with simulation). E.g. subject nr. 14.
- subjects, where the simulation of head movement is not necessary for resolving in-the-head localization (they do not have it even without simulation). E.g. subject nr. 2.

From the 50 subjects 14 found the simulation helpful (28%). Most of them, 28 did not need it because they externalized the sound source from the beginning. For 6 subjects the simulation did not help at all. It is interesting that 2 subjects reported first externalized source then during the simulation in-the-head localization.

The same evaluation can be made for front-back reversals. It is often discussed that in case of a simulated sound source in the front a report of backward direction can be regarded as incorrect localization. Even using HRTFs from a good localizer can lead to a high rate of reversals. Therefore, Table 3 includes answers “front”, “back” and “other direction”. 23 subjects (46%) reported correct localization in the front and 24 (48%) reported back or other source locations independent of the head-movement simulation. 11 of the subjects (22%) reported mainly “other” directions and had never the sensation of a frontal sound source location. Only for two subjects did help the simulation.

Furthermore, the border of the yellow-filled fields shows the limit in degrees where subjects first perceived the movement. Most of the subjects reported this sensation at 3 degrees. Of course, this blue and yellow pattern is the same for both tables.

5. CONCLUSIONS

50 untrained subjects participated in a listening test using HRTF synthesis and headphone playback. A virtual sound source in front of the listener was simulated first stationary, followed by random movements of 1-7 degrees around the reference location in all directions. The goal was to simulate small head movements and to evaluate front-back reversal and in-the-head localization rates. Preliminary results using only one setting of the parameters lead us to conclude that this kind of simulation can be helpful to resolve in-the-head localization if we randomly move the simulated sound source about 1-2 degrees. For 28% this simulation was helpful while 56% of the listeners were not influenced at all. On the other hand, the simulation did not really influence front-back reversals. Correct perception of frontal direction appeared by 46% of the subjects. A further 26% reported about front-back reversals and 22% failed localization. Simulated head-movements more than 4 degrees will be perceived as a moving source.

6. FUTURE WORKS

Currently we are evaluating different settings, especially for parameters B and C. Optimum values for these parameters might be determined. A detailed evaluation and presentation of the results is planned to be made in the near future.

	0 deg	*1 deg*	*2 deg*	*3 deg*	*4 deg*	*5 deg*	*6 deg*	*7 deg*
1	N	N	N	N				
2	N	N	N	N				
3	N	N	N	N	N	N		
4	N	N	N	N				
5	N	N	N	N				
6	N	N	N	N	N	N		
7	N	N	N	N				
8	N	N	N	N				
9	N	N	N	N				
10	N	N	N	N				
11	N	N	N	N	N			
12	N	N	N	N				
13	N	N	N	N				
14	Y	Y	Y	Y				
15	Y	Y	Y	Y				
16	N	N	N	N				
17	Y	Y	N	N				
18	N	N	N	N				
19	Y	N	N	N				
20	N	N	N	N				
21	N	N	N	N				
22	N	N	N	N				
23	Y	Y	N	N				
24	Y	Y	Y	Y				
25	N	N	N	N	N	N		
26	N	N	N	N				
27	Y	Y	N	N				
28	Y	Y	Y	N				
29	N	N	N	N				
30	Y	Y	N	N				
31	Y	Y	Y	Y	Y			
32	N	N	N	N				
33	Y	Y	N	N				
34	N	N	N	Y				
35	N	N	N	N				
36	Y	N	N	N	N			
37	Y	Y	Y	Y				
38	Y	N	N	N				
39	Y	Y	N	N				
40	N	N	N	Y	Y	Y		
41	Y	N	N	N	N			
42	Y	Y	Y	Y	Y	Y		
43	N	N	N	N	N			
44	Y	N	N	N				
45	Y	Y	N	N	N	N	N	
46	N	N	N	N	N	N		
47	N	N	N	N				
48	N	N	N	N				
49	Y	N	N	N	N			
50	N	N	N	N				

Table 2: Individual results about the existence of in-the-head localization for 50 subjects. N means externalized source, Y means in-the-head localization. Blue fields indicate a sound source that is perceived as a steady source, yellow fields indicate perception of movement.

	0 deg	1 deg	2 deg	3 deg	4 deg	5 deg	6 deg	7 deg
1	front	front	Front	front				
2	front	front	other	front				
3	front	front	front	front	front	Front		
4	other	other	other	other				
5	front	front	front	front				
6	front	front	front	front	front	Front		
7	other	front	front	front				
8	front	front	front	front				
9	front	front	front	front				
10	front	front	other	front				
11	front	front	front	front	front			
12	front	front	back	back				
13	other	other	other	other				
14	other	other	back	back				
15	front	front	front	front				
16	front	front	front	front				
17	back	front	front	front				
18	back	back	back	back				
19	back	back	back	back				
20	front	front	front	front				
21	front	front	front	front				
22	other	other	other	other				
23	back	back	back	back				
24	front	front	front	front				
25	back	back	back	back	back	back		
26	back	back	back	back				
27	front	front	front	front				
28	back	back	back	back				
29	other	back	back	back				
30	front	front	front	front				
31	front	front	front	front	front			
32	back	back	back	back				
33	back	back	back	back				
34	front	front	front	front				
35	front	front	front	front				
36	other	other	back	back	back			
37	front	front	front	front				
38	back	back	back	back				
39	other	back	back	back				
40	front	front	front	front	front	front		
41	back	back	other	back	back			
42	front	front	front	front	front	front		
43	back	back	back	back	back			
44	back	back	back	back				
45	other	other	other	other	other	other	other	
46	other	other	other	other	other	other		
47	back	back	back	back				
48	other	back	other	other				
49	front	front	front	front	front			
50	other	other	other	other				

Table 3: Individual results about front-back reversals for 50 subjects. "Back" and "other" indicate error in localization. Blue fields indicate a sound source that is perceived as a steady source, yellow fields indicate perception of movement.

7. REFERENCES

[1] K. Crispien, and H. Petrie, "Providing Access to GUI's Using Multimedia System – Based on Spatial Audio Representation," *Audio Eng. Soc. 95th Convention Preprint*, New York, USA, 1993.

[2] Gy. Wersényi, "Localization in a HRTF-based Minimum Audible Angle Listening Test on a 2D Sound Screen for GUIB Applications," *Audio Eng. Soc. 115th Convention Preprint*, New York, USA, October 2003.

[3] D. Burger, C. Mazurier, S. Cesarano, and J. Sagot, "The design of interactive auditory learning tools," *Non-visual Human-Computer Interaction.*, vol. 228, pp. 97–114, 1993.

[4] M. M. Blattner, D. A. Sumikawa, and R. M. Greenberg "Earcons and Icons: their structure and common design principles," *Human-Computer Interaction*, vol. 4, no. 1, pp. 11-44, 1989.

[5] G. Awad, *Ein Beitrag zur Mensch-Maschine-Kommunikation für Blinde und Hochgradig Sehbehinderte*. Dissertation, TU-Berlin, Berlin, Germany, 1986.

[6] Gy. Wersényi, "What Virtual Audio Synthesis Could Do for Visually Disabled Humans in the New Era?," in *Proc. of the 12th AES Regional Convention Tokyo,* Tokyo, Japan, June 2005, pp. 180-183.

[7] J. Blauert, *Spatial Hearing*. MIT Press, MA, USA, 1983.

[8] F. E. Toole, "In-head localization of acoustic images," *J. Acoust. Soc. Am.*, vol. 48, pp. 943-949, 1969.

[9] P. Laws, *Zum Problem des Entfernungshören und der Im-Kopf-Lokalisertheit von Hörerignissen*. Dissertation, TU-Aachen, Aachen, Germany, 1972.

[10] G. Plenge, "Über das Problem der Im-Kopf-Lokalisation," *Acoustica*, vol. 26, pp. 241–252, 1972.

[11] N. Sakamoto, T. Gotoh, and Y. Kimura, "On „out-of-head localization" in headphone listening," *J. Audio Eng. Soc.*, vol. 24, pp. 710–716, 1976.

[12] J. Kawaura, Y. Suzuki, F. Asano, and T. Sone, "Sound localization in headphone reproduction by simulating transfer functions from the sound source to the external ear," *J. Acoustic Soc. Japan.*, vol. 12, pp. 203–215, 1991.

[13] P. A. Hill, P. A. Nelson, and O. Kirkeby, "Resolution of front-back confusion in virtual acoustic imaging systems," *J. Acoust. Soc. Am.*, vol. 108, no. 6, pp. 2901–2910, 2000.

[14] A. Harma, 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.

[15] J. Blauert, H. Lehnert, J. Sahrhage, and H. Strauss, "An Interactive Virtul-environment Generator for Psychoacoustic Research I: Architecture and Implementation," *Acoustica*, vol. 86, pp. 94-102, 2000.

[16] D. R. Begault, *3-D Sound for Virtual Reality and Multimedia*. Academic Press, London, UK, 1983.

[17] R. L. McKinley, and M. A. Ericson, *Flight Demonstration of a 3-D Auditory Display. in Binaural and Spatial Hearing in Real and Virtual Environments*. (Edited by R.H. Gilkey and T.R. Anderson), Lawrence Erlbaum Ass., Mahwah, New Jersey, pp. 683-699, 1997.

[18] M. Cohen, and E. Wenzel, *The design of Multidimensional Sound Interfaces*. in W. Barfield, T.A. Furness III (Editors) „Virtual Environments and Advanced Interface Design", Oxford Univ. Press, New York, pp. 291-346, 1995.

[19] F. Chen, "Localization of 3-D Sound Presented through Headphone - Duration of Sound Presentation and Localization Accuracy," *J. Audio Eng. Soc.*, vol. 51, no. 12, pp. 1163–1171, 2003.

[20] D. R. Begault, E. Wenzel, and 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.*, vol. 49, no. 10, pp. 904–917, 2001.

[21] M. Kleiner, B. I. Dalenbäck, and P. Svensson, "Auralization – an overview," *J. Audio Eng. Soc.*, vol. 41, pp. 861–875, 1993.

[22] R. L. Martin, K. I. McAnally, and M. A. Senova, "Free-Field Equivalent Localization of Virtual Audio," *J. Audio Eng. Soc.*, vol. 49, no. 1/2, pp. 14-22, 2001.

[23] W. Noble, "Auditory localization in the vertical plane: Accuracy and constraint on bodily movement," *J. Acoust. Soc. Am.*, vol. 82, pp. 1631-1636, 1987.

[24] P. Minnaar, S. K. Olesen, F. Christensen, and H. Moller, "The importance of Head movements for binaural room synthesis," in *Proc. Int. Conf. on Auditory Display*, Espoo, Finland, July 2001, pp. 21-25.

[25] D. R. Perrott, H. Ambarsoom, and J. Tucker, "Changes in Head Position as a measure of auditory localization performance: auditory psychomotor coordination under monaural and binaural listening conditions," *J. Acoust. Soc. Am.*, vol. 82, pp. 1637-1645, 1987.

[26] Crystal River Engineering, Inc., *BEACHTRON – Technical Manual*. Rev. C, 1993.

[27] S. H. Foster, and 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.

[28] E. M. Wenzel, M. Arruda, D. J. Kistler, and F. L. Wightman, "Localization using nonindividualized head-related transfer functions," *J. Acoust. Soc. Am.*, vol. 94, no. 1, pp. 111-123, 1993.

[29] D. J. Kistler, and F. L. Wightman, "Principal Component Analysis of Head-Related Transfer Functions," *J. Acoust. Soc. Am.*, vol. 88, pp. 98, 1990.

[30] F. L. Wightman, and D. J. Kistler, "Headphone Simulation of Free-Field Listening I-II," *J. Acoust. Soc. Am.*, vol. 85, pp. 858-878, 1989.

[31] P. Minnaar, J. Plogsties, and F. Christensen, "Directional Resolution of Head-Related Transfer Functions Required in Binaural Synthesis," *J. Audio Eng. Soc.*, vol. 53, no. 10, pp. 919–929, 2005.