What Virtual Audio Synthesis Could Do for Visually Disabled Humans in the New Era?

György Wersényi

Széchenyi István University, Department of Telecommunications, Egyetem tér 1., H-9026, Győr, Hungary

Proceedings of the AES 12th Regional Convention, Tokyo, June 12-14, 2005, Japan, pp. 180-183, 2005.

ABSTRACT

Listening tests were carried out for investigating the localization performance of 42 untrained subjects using noise stimuli in a 2D virtual acoustic display (VAD). Measurements were made on the basis of the former GUIB project (Graphical User Interface for Blind Persons). Results of the evaluation of the average spatial resolution will be presented. Suggestions for optimal partitioning the virtual space is made on a rectangle 2D VAD in front of the listener, focusing on vertical localization.

0 Introduction

The GUIB (Graphical User Interface for Blind Persons) project was founded to create a proper virtual environment for blind persons to help them by the use of personal computers [1, 2]. They do not have the possibility of GUIs and thus, events on the screen have to be replaced or extended by sound events [3-6]. A simple, low-cost method is wanted to make visually disabled people able to orientate himself e.g. on a usually PC with MS-Windows.

1 Measurement method

The playback system includes the Beachtron DSP card that produces virtual sound events on a 2D rectangle virtual acoustic display in front of the listener. At the first step 40 subjects determined the average, best-case and worst-case individual spatial resolution in the horizontal and median plane respectively [7, 8]. We used white noise and filtered noise stimuli (Fig.1). Test signals were selected to model real sound events in length and loudness in a generic way but in the same time allowing testing localization depending on spectral content. Cut-off frequencies for the filtering were chosen to be drastic and far from each other in the frequency in order for a good separation between Signal A (whatever it will be later) and the filtered signals from it.

Listeners reported in a three-category-forced-choice Minimum-Audible-Angle (MAA) measurement and determined a directional-independent average spatial resolution along the horizontal and vertical axes (Fig.2). 300 ms burst-impulse pairs were used and subjects had to discriminate them as the second noise burst was moving away or toward the first (reference) noise burst signal in 1° steps. MAA was found to be optimal for signals that are below 1000 Hz and/or above 4000 Hz. Real-time HRTF filtering originating from a "good localizer" together with proper headphone equalization is made by the DSP card that is necessary for a virtual sound field simulation [9-14].

For a user-friendly mapping between visual and sound events of the screen (e.g. for using the mouse) a rectangle 2D "screen-like surface" is simulated as an extension or replacement of the display.

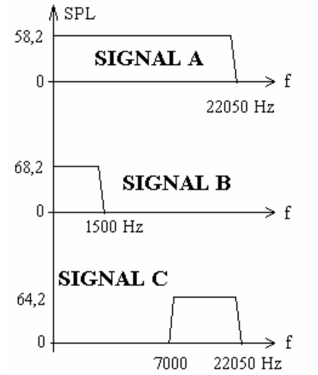

Fig.1. Spectra of the noise signal excitation

1.1 Results of the current investigation

The results of 40 subjects delivered an average resolution as shown on Fig.2. Black filled dots correspond to virtual source locations on the 2D VAD as a total average over all subjects and test signals [8, 15].

Based on Fig.2 the average resolution was simulated using the same system, measurement method and stimuli. The goal was to test this resolution and determine how many subjects could actually use a resolution of 13x5. We assumed that 13 sources horizontally (in a resolution of about 7-10°) and 5 vertically (in a resolution of about 15°) will be "too much" and unusable for a real application.

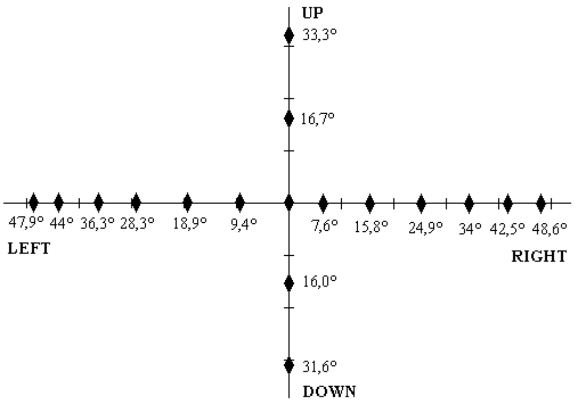

Fig.2. Average values as possible source locations from a former measurement

Instead of the method described earlier now a simplified method is used for the listening test. Listeners were now asked to report only in a 2-category-forced-choice as sound sources moving from one source location to another. Possible answers were "no difference between source locations" and "different source locations" depending on the sensation. E.g. a reference noise impulse was simulated at 7.6° and the second at 15.8°. If the subject was able to discriminate them, the reference point was moved in 15.8° etc. If he could not make a spatial separation, the second source was moving one step further (24.9°). A new reference point was initiated by the listening test as the listener was able to discriminate the sound sources.

Table 1 shows results for Signal A, B and C as well as the total average horizontally and vertically. Only about 21% of the subjects were able to perceive all 13 simulated sources in the horizontal plane and 29% all five in the median plane (the origin is always included and is a simulated sound source location).

Median plane localization is much better for white noise than for filtered noise stimuli, but both seems to be inappropriate in contrast to horizontal plane localization. We were also searching for a source number limit that can be localized by about 80% of the users.

	Signal A (white noise)	Signal B (1500 Hz LPF white noise)	Signal C (7000 Hz HPF white noise)	Signal A, B, C Total
Horizontal				
all 13 locations	28%	24%	12%	21%
at least 9 locations	83%	83%	81%	**82%**
Vertical				
all 5 locations	54%	19%	14%	29%
at least 3 locations	95%	78%	81%	**85%**

Table 1. Evaluation of the average resolution of 13x5 based on a MAA listening test of 42 subjects. Signal A is white noise, Signal B and Signal C are LPF and HPF filtered versions of Signal A respectively.

Referring to Table 1 82% of the subjects were able to discriminate 4 sources left and right from origin respectively independent of signal content. This 82% includes all subjects who could discriminate 9, 10, 11, 12 or 13 sound source locations horizontally. 85% could discriminate at least one source location above and below the origin (3 or 5). This evaluation assumes that subjects who can localize 5 vertical sound source positions are also able to handle less than five.

%	**1**	**2**	**3**	**4**	**5**	**6**	**Sum**
Right	74	88	82	80	86	74	80.6
Left	70	93	90	77	83	63	79.3
Up	60	81	-	-	-	-	70.5
Down	58	62	-	-	-	-	60

Table 2. Evaluation of individual source locations in every direction (average for all signals).

It is interesting to see the evaluation from the side of the sound source locations. Table 2 shows that the first reference point on the left side could be identified by 74% of the listeners (in contrast to the origin) and skipped by 26%. Individual source location could be identified and localized by about 80% of the listeners in the horizontal plane. This also means that there is a steady number of 20% where a source location was skipped by users, because they were not able to discriminate them from the neighbored source location. The summarized result for vertical localization is much worse: only 70.5% and 60% of the source locations could be identified up and down respectively.

All this suggests a resolution of virtual sources of 9x3 instead of 13x5 (Fig.3).

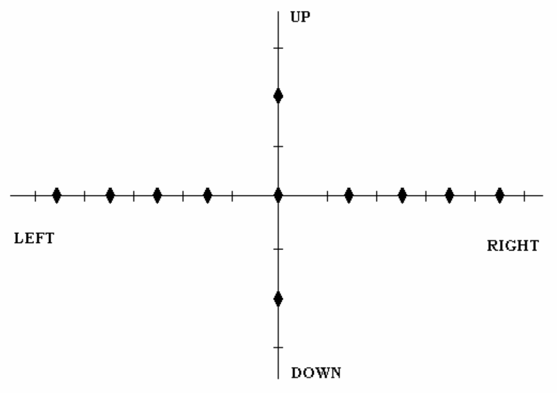

Fig.3. Suggested sound source locations in the median and in the horizontal plane that could be suitable for about 80% of the users

1.2 Vertical localization

Our previous study showed that using this playback system and method about 33% of the listeners could not localize vertically at all [8]. They make their MAA judgments based on the spectral distortion of the HRTFs (as sound sources "sound different") without real localization. Subjects were asked this time as well to determine the movement of the sound source (up, down or left, right). False answers indicate the lack of real localization.

In the horizontal plane all subjects answered correct for Signal A and C, and only 2 false answers appeared for Signal B. This is almost 100% of correct answers.

The correct answers in the median plane for Signal A were 63%, for Signal B 58% and for Signal C only 52%. As supposed, vertical localization is poorer than horizontal localization: one-third of the subjects could not tell whether the sound source is "up" or "down" (Signal A), and for Signal B and C this number reaches about the half of the subjects. For Signal C 73% could determine the locations "up", but only 30% the locations "down". This phenomenon is known as elevation shift and we support the fact that in vertical localization plays the signal content a significant role: broadband signals could be localized the best, low-frequency signals rather "down" and high-frequency stimulus rather "up".

1.3 Errors during localization

For controlling the subjects' answers, sometimes the second noise impulse did not move at all, thus both impulses were steady at the same source location. Indeed, in 95% of the simulation subjects did not observe any change as expected. Surprisingly, about 5% of the answers indicated sensation of different source locations in the horizontal plane and about 4% in the median plane. But there was no pattern to recognize in the errors in dependence of direction or signal frequency.

Our first investigation showed regular asymmetry of localization results on the left and right side in the horizontal plane. Sources on the left side were harder to localize by 2-4° on average. Fig.2 reflects this fact (7.6° in contrast to 9.4° for the first source location) and due to this asymmetrical simulation the left-right asymmetry disappeared. There was no convincing difference among localization judgments from the left and the right side.

2 Discussion

Results from our first investigation with the system described above lead us to have an average spatial resolution of 13x5. This average was calculated over 40 subjects and signals with different spectra.

In a similar listening test the average resolution was presented to a control group of 42 students at the university. Results show that about 80% of the users would be able to use a resolution of 9x3.

Due to the poor vertical localization and increasing errors in the horizontal plane as we move further from the origin, an even more simplified 2D resolution will be simulated in the next test runs. Instead of moving only along the horizontal and vertical axes, the whole surface will be used as shown on Fig.4.

Our current investigation uses 3x3 and 5x2 spatial resolution on the 2D virtual surface (Fig.4). Preliminary results using all three signals show that about 40-48% could use a 3x3 and 38-50% a 5x2 resolution depending on spectral content. These results are due to poor vertical localization while subjects mostly fault vertically. At a resolution of 3x3 only 2-5% of the subjects make horizontal errors but about 48-58% make vertical errors. For the resolution of 5x2 20-27% had horizontal errors and 30-35% vertical errors. This suggests that even 5 horizontal source locations instead of 3 increased the horizontal errors up to 27%, and 2 vertical locations instead of 3 reduced the vertical errors only down to 35%.

A1	A2	A3
B1	B2	B3
C1	C2	C3

A1	A2	A3	A4	A5
B1	B2	B3	B4	B5

Fig.4. Partitioning of a 2D VAD in front of the listener for 3x3 and 5x2 for listening tests.

These results drive us to investigate this deeply and apply additional methods to increase vertical localization. As mentioned above, high-pass filtering and low-pass filtering of the input signal (sound event) additional to the HRTF filtering may increase the number of correct judgments during vertical localization. This is based on the observation that signals with more high-frequency content often are localized in the upper hemisphere and signals with more low-frequency content at lower elevations [10]. Our results also support this fact by comparing localization judgments in the directions up and down. In addition, a-priori knowledge about this filtering-method could bias listeners toward correct judgments.

3 Summary

Listening tests were carried out with 42 untrained subjects in a virtual audio simulation using a 2D virtual audio display and different noise excitation signals. Based on former results an average spatial resolution of 5 locations vertically and 13 horizontally was simulated. Only 21%-29% could actually use this spatial resolution, but about 82%-85% were able to localize 9 horizontal and 3 vertical positions independent of signal content. This suggests that an average spatial resolution over subjects is insufficient for 70-80% of possible users. Results suggest a partitioning for 9x3, but preliminary results of a current investigation using a 3x3 and 5x2 simulation over the whole 2D surface refer to insufficient localization even using 2 vertical and 5 horizontal source locations.

4 Future works

Future works includes final listening tests to apply simple signal processing tools for increasing quality and vertical localization. Existing programs are suitable to "enhance" the sensation of vertical displacement by using additional signal processing methods.

5 References

[1] K. Crispien, 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).

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

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

[4] M. M. Blattner, D. A. Sumikawa, R. M. Greenberg, "Earcons and Icons: their structure and common design principles," Human-Computer Interaction, vol. 4(1), pp. 11-44, (1989).

[5] D. R. Begault, "3-D Sound for Virtual Reality and Multimedia" Academic Press, London, UK, (1994).

[6] 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).

[7] Crystal River Engineering, Inc.: BEACHTRON – Technical Manual, Rev.C., (1993).

[8] Gy. Wersényi, "Localization in a HRTF-based Minimum Audible Angle Listening Test on a 2D Sound Screen for GUIB Applications," AES Convention Preprint Paper, Nr.5902, Presented at the 115th Convention, New York, USA (2003).

[9] J. Kawaura, Y. Suzuki, F. Asano, T. Sone, "Sound localization in headphone reproduction by simulating transfer functions from the sound source to the external ear," J. Acoust. Soc. Japan E 12, pp. 203-215, (1991).

[10] J. Blauert, "Spatial Hearing" The MIT Press, MA, (1983).

[11] H. Møller, "Fundamentals of binaural technology," Applied Acoustics, vol. 36, pp. 171-218, (1992).

[12] E. M. Wenzel, M. Arruda, D. J. Kistler, F. L. Wightman, "Localization using nonindividualized head-related transfer functions," J. Acoust. Soc. Am., vol. 94(1), pp. 111-123, (1993).

[13] W. M. Hartmann, B. Rakerd, "On the minimum audible angle – A decision theory approach," J. Acoust. Soc. Am., vol. 85, pp. 2031-2041, (1989).

[14] W. Mills, "On the minimum audible angle," J. Acoust. Soc. Am., vol. 30, pp. 237-246, (1958).

[15] Gy. Wersényi, "HRTFs in Human Localization: Measurement, Spectral Evaluation and Practical Use in Virtaul Audio Environment" PhD thesis, BTU Cottbus, 2002.

György Wersényi was born in Hungary, in 1975. He received the MSc degree in electrical engineering from the Technical University of Budapest in 1998 and the PhD degree in 2002 from the BTU Cottbus, Germany. He is now assistant professor at the Széchenyi István University. His areas of interest include spatial hearing, acoustic measurements, psychological acoustics and auditory modeling.

http://rs1.szif.hu/~wersenyi/
wersenyi@sparc.core.hu