Acoust. Sci. & Tech. **28**, 4 (2007)

PAPER

Localization in a Head-Related Transfer Function-based virtual audio synthesis using additional high-pass and low-pass filtering of sound sources

György Wersényi

Department of Telecommunications, Széchenyi István University, H-9024, Györ, Egyetem tér 1., Hungary

(*Received 10 July 2006, Accepted for publication 1 November 2006*)

Abstract: Listening tests were carried out for investigating the localization judgments of untrained subjects through equalized headphones and with HRTF synthesis. The investigation was made on the basis of the former 'Graphical User Interface for Blind Persons' project in order to determine the possibilities of a 2D virtual sound screen and headphone playback. 50 untrained subjects evaluated a virtual audio display in front of the listener using different horizontal and vertical resolutions on a 2D surface. A listening test using white and filtered noise signals was followed by a special investigation using simple high-pass and low-pass filtering of the original sound in order to increase correct vertical localization judgments. The simulation uses high-pass filtering for higher elevations and low-pass filtering for lower elevations in a 5×2 and a 3×3 spatial resolution. Results of the listening test will be presented and the efficiency of the filtering in correct localization judgments will be discussed.

Keywords: Localization, HRTF, Virtual, Headphone

PACS number: 43.66.Qp [doi:10.1250/ast.28.244]

1. INTRODUCTION

Virtual Acoustic Displays (VAD) are widely used in several applications. Virtual sound sources are artificially reproduced and they have to be localized and identified by human listeners. The abbreviation GUIB stays for Graphical User Interface for Blind Persons. In this international project the goal was to create a virtual environment for elderly and visually disabled people to help those using personal computers. Blind persons do not have the advantageous properties of Graphical User Interfaces and the ability of orientation among multiple visual information (icons, windows) [1–11].

In this case events and icons of a screen have to be replaced and/or extended by sound events (so called Earcons). The simulation includes the spatial distribution of the sound sources in a 2D virtual audio display in front of the listener using headphone playback. The former results of this project are: a collection of sound sources (Earcons), the possibilities of different input media (Braille-displays, touch-pads) and the localization blur of a multi-channel loudspeaker array [2,9,12]. In this first part only the loudspeaker playback was introduced.

Next, the localization results with the same system were presented using headphone playback: 40 untrained subjects determined the worst-case, best-case and average spatial resolution in the horizontal and median plane respectively [13,14]. Trained subjects are experienced in listening tests, thus, they would deliver better results than inexperienced subjects. The investigation included a special three-category-forced-choice Minimum-Audible-Movement-Angle (MAMA) measurement using white noise burst stimulus and filtered version of white noise impulse pairs. MAMA is the smallest change between moving audio sources that cause audible difference and allow directional separation of the sound sources.

Later, the third part of the investigation realized a control measurement with another group of subjects. They evaluated the 'average' results, the applicability of different resolutions both in a MAMA as well as in an absolute measurement [15]. Results showed that a 2D rectangle screen in front of the listener may have a resolution of 3×1 (horizontal $\times$ vertical) up to 5×2 or 3×3. Results also showed that localization judgments and errors are mainly due to vertical errors and poor vertical localization.

For increasing the correct judgments of vertical displacement of sound sources a simple method is now introduced to enhance correct answers during the tests. Based on the psychoacoustic fact, that signals having more high frequency information are often localized "high" as long signals having more low frequency information are often localized "low," high-pass and low-pass filtering is

applied on the input signal. This HPF and LPF filtering is additional to the HRTF filtering. E.g. by choosing three vertical locations, unfiltered white noise means horizontal plane, high-pass filtered version of it means "above horizontal plane" while low-pass filtered version means "below horizontal plane."

Now, 50 untrained subjects evaluated spatial resolutions of 3×3 and 5×2 (horizontal $\times$ vertical) with and without this additional HPF/LPF filtering. Results show that this simple HPF and LPF filtering increase the number of correct judgments in vertical localization. Discussion is made on the applicability of this method in real applications using Earcons or other input signals instead of white noise.

2. SYSTEM SETUP

The measurement system includes the Beachtron DSP-board [2,5,16]. The Beachtron is an ISA-slot based sound card that performs real-time spatialization up to two separate audio sources. The system delivers Left and Right outputs, which are mixed and played through conventional headphones. The card includes a Motorola DSP56001 clocked at 40 MHz, and high-performance AD converters using 44,100 samples per second and 16 bits. The Left and Right filters (HRTFs) can be changed dynamically, as often as every 46 ms, or about, 22 times per second from a data-set containing 74 filters each, per Left and Right ear. The HRTFs are implemented as FIR filters of length 75 taps (per ear). The playback system is equalized for the Sennheiser HD 540 headphone.

The supplied HRTFs originate from the measurement of *Wightman and Kistler* [4,6]. They were measured in the ear canal of a female head. This subject was chosen as a representative of 8 subjects, whose data showed good localization accuracy and an average rate of front-back confusions. Details of this measurement technique are described in [6]. It was also shown that using the HRTFs of such a "good localizer" results in sufficient directional simulation in the horizontal plane and in only minor increase of front-back confusions in the median plane [4]. The Beachtron let to set the diameter (size) of the head as well. This produces more accurate simulation by setting the exact time of arrival of the virtual sound at the eardrums.

Figure 1 shows the coordinate system and signal presentation. Figure 2 shows the input signals: white noise (A), LPF white noise with 1,500 Hz cut-off frequency (B) and HPF white noise with 7,000 Hz cut-off frequency (C). These are 300 ms long impulses. Cut-off frequencies were chosen to be fairly away from each other to create a significant audible filtering effect. The value about 1,500 Hz is well known in the literature as a limit that separates localization based on the time delays in the fine structure and in the envelope. The value of 7,000 Hz was found

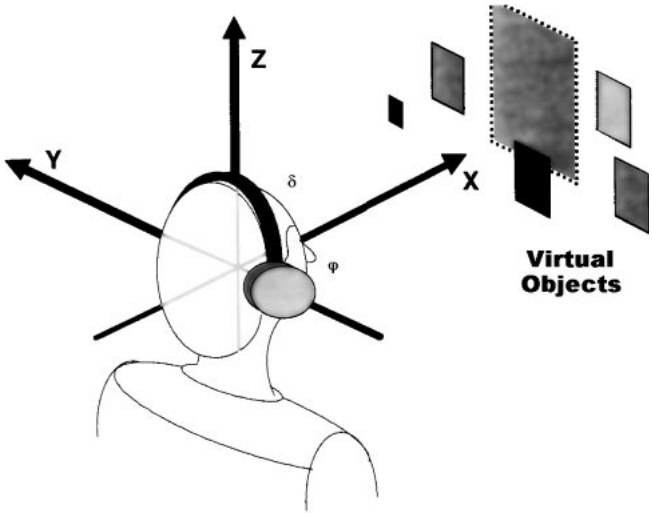

Fig. 1 Illustration of virtual sources in a 2D representation. The virtual acoustic surface is parallel with the *Z-Y*-plane. The origin is in the front of the listener: $\varphi = \delta = 0°$.

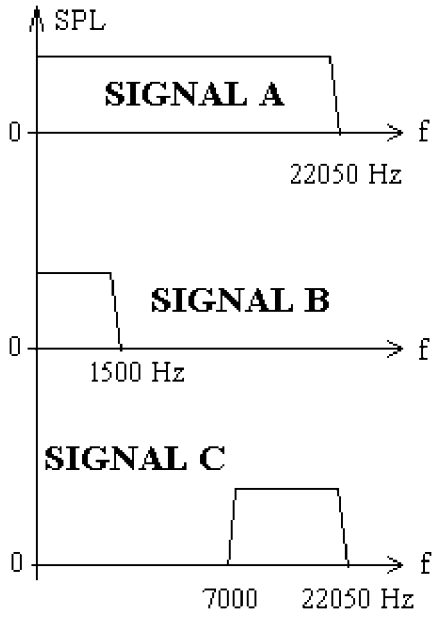

Fig. 2 Spectral representation of the input signals using rectangle filtering. Subjects determined the level of Signals B and C to be as loud as Signal A.

appropriate by filtering of various types of sound sources, even for sound events having more high frequency content. Filtering is made by a rectangle filter characteristics and Blackman-Harris windowing using CoolEdit Software. Listeners determined during a preliminary test the overall signal levels for the stimuli. First they set the most comfortable level for signal A that resulted in an average value of 58.2 dB. This was followed by the setting of signal B and signal C to be as loud as signal A. In order to do this, signal B has to be 10 dB louder and signal C 6 dB louder than signal A, using the equalized Sennheiser

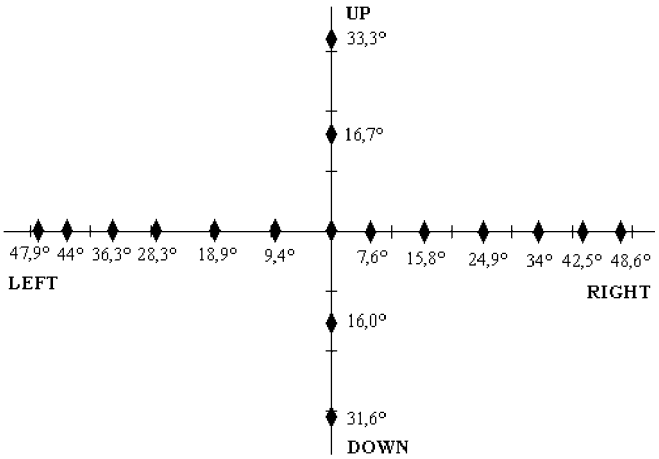

Fig. 3 Average values as possible source locations for signal A. Black filled dots correspond virtual source locations on the 2D VAD.

headphone. Considering that users are able to adjust volume in applications these values are only used for the listening tests as a common platform.

The first part of a series of investigation was about to determine the best, the worst and the averaged spatial resolution with this system [13,14]. A preliminary test delivered results about the values of typical headphone playback errors like front-back reversals, in-the-head-localization and movements symmetrical to the median plane.

The measurement method included a special 3-categorie-forced-choice in order to determine the "uncertainty domains" during the Minimum Audible Movement Angle (MAMA) localization task. The VAD is a 2D "screen-like" sound screen in front of the listener. Simulated distance of the VAD is 50 inches in front of the listener and the virtual screen is spanned $\pm 60^\circ$ from the origin. Sound sources were simulated along the median and horizontal plane only. There is only less experimental result in such an environment. Localization blur results were compared with former results. This investigation ended with suggestions for a GUIB application: how to partitioning a 2D VAD depending on the spectrum of the test signal. The results of 40 subjects delivered an average resolution as shown on Fig. 3. All investigations so far were executed on sighted subjects to relieve the blind people. Future investigations include the blind as well.

3. MEASUREMENT METHOD AND RESULTS

The next part of the investigation included a control group of another 40 subjects [15]. Subjects have been young adults between 21 and 39 years of age, half male, half female all with normal hearing, verified by standard audiometric screening. Based on Fig. 3. 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 13×5 along both axes. 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. Outside of laboratory conditions we assume decreased spatial resolution and limited possibilities (in a standard living room with commercial headphones).

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 are 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 subjects answer if the listener was able to discriminate the sound sources.

Table 1 shows results for Signals 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. 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. All this suggests a resolution of virtual sources of 9×3 instead of 13×5 (Fig. 4).

3.1. Vertical Localization

Our previous study showed that using this playback system and method about 33% of the listeners could not localize vertically at all [13]. They make their MAMA 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).

Table 1 Evaluation of the average resolution of 13 × 5 based on a MAMA listening test of 40 subjects. Signal A is white noise, Signals B and Signal C are LPF and HPF filtered versions of Signal A respectively.

	Signal A (white noise)	Signal B (1,500 Hz LPF white noise)	Signal C (7,000 Hz HPF white noise)	Signals A, B, C Total
Horizontal				
all 13 locations at least 9 locations	28%	24%	12%	21%
	83%	83%	81%	**82%**
Vertical				
all 5 locations at least 3 locations	54%	19%	14%	29%
	95%	78%	81%	**85%**

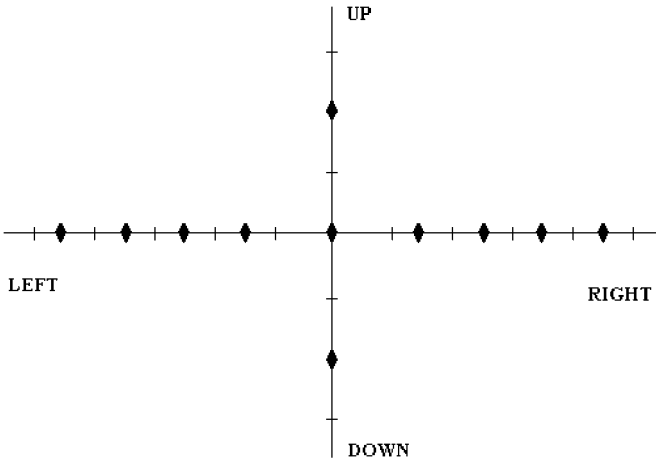

Fig. 4 Suggested sound source locations in the median and in the horizontal plane that could be suitable for about 80% of the users in a MAMA measurement. Compare with Fig. 3 and note the decreased number of possible source locations.

False answers indicate the lack of real localization. In this investigation all subjects answered correct for Signals A and C, and only 2 false answers appeared for Signal B for the horizontal plane. 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 Signals 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."

3.2. Localization Behaviour

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. Figure 3 reflects this fact (7.6° in contrast to 9.4° for the first source location) but now 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.

3.3. Absolute Measurement

The third part of the investigation uses the whole 2D VAD instead of the axes only. 50 young adults, students of the university between 20 and 23 years of age participated in this investigation. All have been male with normal hearing. Subjects were sitting in a chair in the anechoic room. During the adaptation time the test signals were presented them and trial runs were made. A detailed description of the goal of the investigation and of the procedure was also given. The measurement used the same equipment as before, the same signals (A, B and C), level of loudness and headphone.

Sources can be simulated by different spatial resolution as shown on Fig. 5. Sound sources are in the middle of the blocks and subjects have to identify them by calling the appropriate letter and number. The goal of this absolute measurement is to test different resolutions on the whole surface, vertical and horizontal localization performance and possible improvement by filtering methods using the same signal excitation and equipment.

One sound source contained only one burst-pair: two 300 ms of the same signal separated by 400 ms silence.

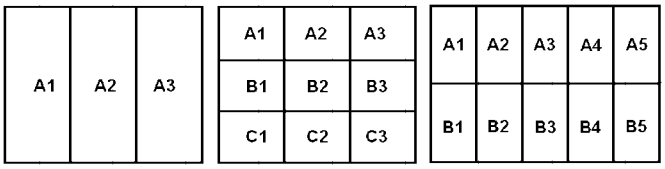

Fig. 5 Different spatial resolution of the 2D VAD: 3×1, 3×3 and 5×2. Sound sources are simulated in the middle of the surface elements.

Table 2 Resolution of 5×2, Signals A, B, C.

Signal A

True	Neighbor	False
50.71%	29.29%	20.00%

Signal B

True	Neighbor	False
46.96%	32.41%	20.89%

Signal C

True	Neighbor	False
37.86%	35.36%	26.79%

Every block was initiated exactly twice (two rounds) in randomized order. That means e.g. for 3×3 and for signal A, every subject delivered 18 answers. Using all three signals it is 3×18 for the resolution 3×3, and 3×20 for 5×2. All together averages are calculated for all 50 subjects, each with 114 answers (5,700).

Two different spatial resolutions were chosen for evaluation. It is supposed (and it was verified before this investigation) that a partitioning of 3×1 is acceptable for everybody. Three horizontal positions (front, left and right) without any vertical simulation can be used with 100% accuracy.

Therefore, we decided to increase vertical positions up to 3 and the number of possible horizontal locations up to 5. The resolution 3×3 is suited for testing the vertical errors while 5×2 the angular errors. Figures 6 and 7 present summarized results for all signals based on Tables 2 and 3 .

Answers of the subjects can be "true" if they hit the correct block where the sound source is being simulated. The answer is "neighbor" if the localization is correct horizontally but false vertically. Finally, the answer is false, if localization fails both horizontally and vertically. In deed, the sum of neighbored and false answers is the total number of incorrect localization.

For resolution 3×3 only 38–48% of the answers are correct depending on spectral content. Incorrect localization is mainly due to poor vertical localization.

Table 3 Resolution of 3×3, Signals A, B, C.

Signal A

True	Neighbor	False
48.28%	47.49%	4.21%

Signal B

True	Neighbor	False
39.46%	58.43%	2.11%

Signal C

True	Neighbor	False
37.73%	56.89%	5.36%

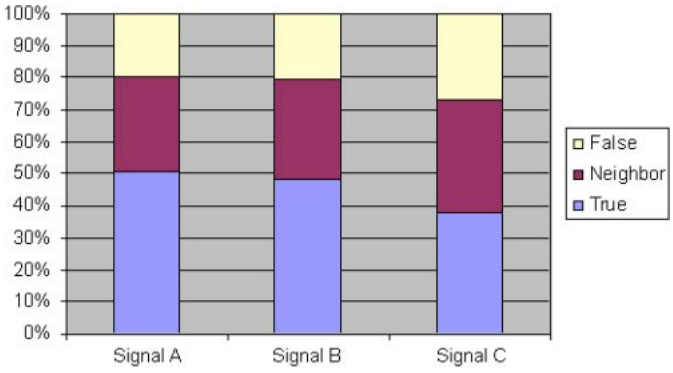

Fig. 6 Measurement results for resolution 5×2 for Signals A, B and C respectively. Compare with Fig. 7.

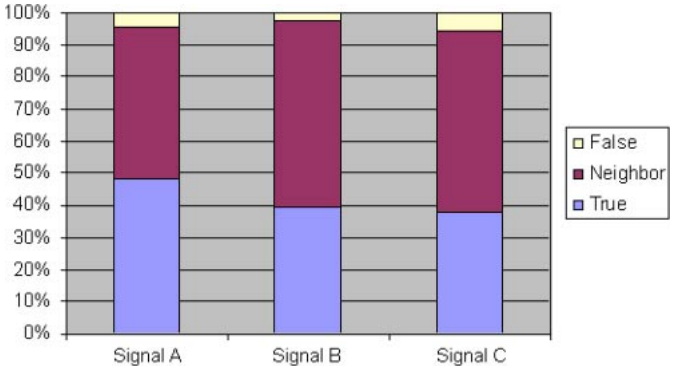

Fig. 7 Measurement results for resolution 3×3 for Signals A, B and C respectively. Compare with Fig. 6.

For resolution 5×2 only the rate of correct answers is about the same. Because of the less vertical positions and of the five possible horizontal locations, more false answers appear.

These results suggest better performance by increasing vertical localization and using 3–5 horizontal locations. Subjects were undecided even during giving correct answers by the far left and far right column: columns 1 and 2 as well as columns 4 and 5 are very hard to

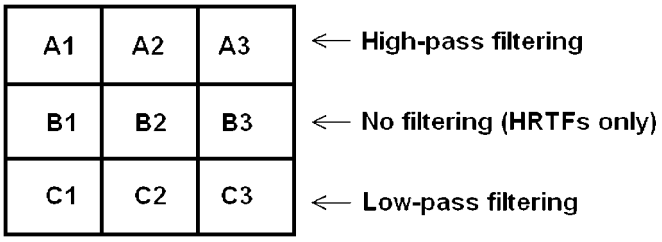

Fig. 8 A possibility for increasing vertical localization. Input signals (wave) can be filtered by HPF and LPF filters before or after the HRTF filtering.

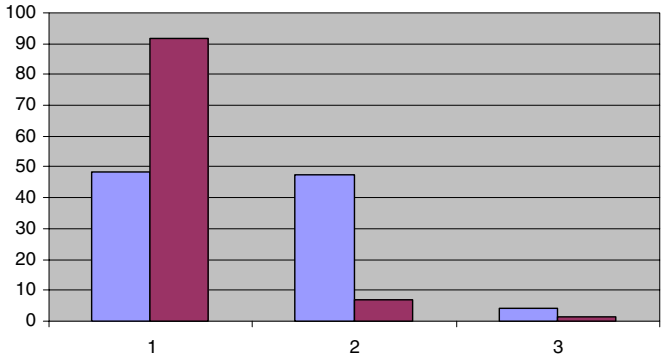

Fig. 9 Measurement data for resolution 3 × 3 based on Tables 3 and 4. Left columns show true (1), neighbor (2) and false (3) answers using only HRTF filtering, right columns show results using additional HPF and LPF filtering. Note the increased rate of correct and decreased rate of incorrect answers.

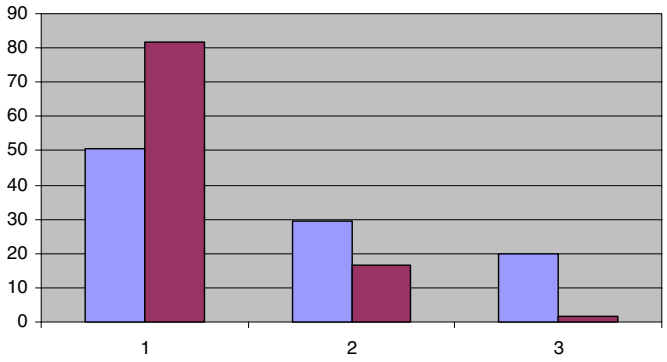

Fig. 10 Measurement data for resolution 5 × 2 based on Tables 2 and 4. Left columns show true (1), neighbor (2) and false (3) answers using only HRTF filtering, right columns show results using additional HPF and LPF filtering. Note the increased rate of correct and decreased rate of incorrect answers.

Table 4 Results using additional LPF and HPF filtering using white noise.

3 × 3

True	Neighbor	False
91.84%	6.94%	1.22%

5 × 2

True	Neighbor	False
81.88%	16.62%	1.50%

discriminate. Signal A is to be localized best, followed by signals C and B.

3.4. Simulation Using Additional Filtering

It was shown in previous measurements that signals containing more high frequency information are often localized "above," while signals with more low-frequency information "below" the horizontal plane [7,17,18].

We repeated the same investigation of section 4 with another 50 subjects the same way except signal presentation. To increase correct answers during vertical localization, a very simple low-pass and high-pass filtering was used to bias incorrect judgments. It was supposed that additional filtering may increase the identification of the blocks. The filtering was realized by using signal A as broadband noise and signal B and C as HPF and LPF versions of it respectively. Figure 8 shows an example for the resolution 3 × 3. HRTF filtering is always included by the DSP card, so for the blocks A1–A3 and C1–C3 both HRTF *and* additional filtering is applied.

The subjects did know the fact that filtering is applied as they got familiar with the signals. We tried also without this a-priori knowledge but subjects figured out what the measurement is about, so they realized that filtering should help them to solve the localization problem.

Evaluation is made the same way as before. Because only one signal is used (Signal A is the excitation, signals B and C is only seen as its filtered version), Figure 9, Fig. 10 and Table 4 present results for different resolutions and for comparison.

Results show significant increasement of correct answers: 80–90% of the answers were correct. It was surprisingly that incorrect answers appeared. It is very important to emphasize that correct answers in vertical localization are due to the spectral filtering of the HPF and LPF filter and not due to real localization. The same observed *Mills*: subjects reported that the difference between the stimuli seemed to be in the loudness or quality of the sound rather than its location [19,20]. Subject can distinguish between signals because they sound different rather than based on their location.

4. SUMMARY

Listening test were carried out to investigate the localization judgments of 50 untrained subjects using a 2D virtual audio surface in front of the listener. The

simulation is made by real-time HRTF filtering through equalized headphones using broadband and filtered noise.

The investigation is made on the basis of the GUIB Project to test the possibilities of a 2D acoustic display. Different spatial resolutions were evaluated. First, the average spatial resolution that is delivered from a former measurement was presented. This suggested a spatial resolution of 9×3 can be suitable for 80% along the horizontal and vertical axes in a MAMA measurement.

The next part contained an absolute measurement of two different resolutions of the 2D surface. Using only the HRTF filtering only about the half of the subjects were able to use a 3×3 and a 5×2 resolution due to poor vertical localization. The extended simulation using additional LPF and HPF filtering to the HRTF filtering increased the rate of correct answers up to 80–90%.

It is suggested to avoid using vertical displacement of sources or to use additional filtering to increase correct judgments. Subjects are able to discriminate three different elevations by additional filtering even without real localization. A-priori knowledge and spectral modification helps to resolve spatial ambiguity. It is supposed that a 3×3 resolution with additional filtering is suitable for about 90% of the users.

Future works include simulation using different kind of sound sources such as the Earcons, narrow-band noises or speech.

REFERENCES

[1] M. Cohen and E. Wenzel, "The design of multidimensional sound interfaces," in *Virtual Environments and Advanced Interface Design*, W. Barfield and T. A. Furness III, Eds. (Oxford University Press, New York, Oxford, 1995), pp. 291–346.

[2] K. Crispien and H. Petrie, "Providing access to GUI's using multimedia system — Based on spatial audio representation," *95th Conv. Audio Eng. Soc. Prepr.*, 3738 (1993).

[3] D. Burger, C. Mazurier, S. Cesarano and J. Sagot, "The design of interactive auditory learning tools," in *Non-Visual Human-Computer Interactions: Prospects for the Visually Handicapped*, D. Burger and J.-C. Sperandio, Eds. (John Libbey Eurotext Publisher, Montrouge, 1993), pp. 97–114.

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

[5] S. H. Foster and E. M. Wenzel, "Virtual acoustic environments: The convolvotron," Demo system presentation at SIGGRAPH'91, *18th ACM Conf. Computer Graphics and Interactive Techniques*, Las Vegas, Nev. (ACM Press, New York, 1991).

[6] F. L. Wightman and D. J. Kistler, "Headphone simulation of free-field listening I.–II.," *J. Acoust. Soc. Am.*, **85**, 858–878 (1989).

[7] J. Blauert, *Spatial Hearing* (The MIT Press, Cambridge, Mass., 1983).

[8] H. Møller, "Fundamentals of binaural technology," *Appl. Acoust.*, **36**, 171–218 (1992).

[9] K. Crispien and 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 and R. Bellacasa, Eds. (IOS Press, Amsterdam, 1995).

[10] 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. Acoust. Soc. Jpn.* (*E*), **12**, 203–215 (1991).

[11] K. Fukodome, "Equalization for the dummy-head-headphone system capable of reproducing true directional information," *J. Acoust. Soc. Jpn.* (*E*), **1**, 59–67 (1980).

[12] M. M. Blattner, D. A. Sumikawa and R. M. Greenberg, "Earcons and icons: their structure and common design principles," *Hum.-Comput. Interaction*, **4**, 11–44 (1989).

[13] Gy. Wersényi, "Localization in a HRTF-based minimum audible angle listening test on a 2D sound screen for GUIB applications," *115th Conv. Audio Eng. Soc. Prepr.*, 5902 (2003).

[14] Gy. Wersényi, *HRTFs in Human Localization: Measurement, Spectral Evaluation and Practical Use in Virtaul Audio Environment*, PhD Thesis, BTU Cottbus (2002).

[15] Gy. Wersényi, "What virtual audio synthesis could do for visually disabled humans in the new era?," *Proc. 12th AES Reg. Conv. Tokyo*, June 12–14, pp. 180–183 (2005).

[16] Crystal River Engineering, Inc., BEACHTRON — Technical Manual, Rev.C. (1993).

[17] C. Tan and W. Gan, "Direct concha excitation for the introduction of individualized hearing cues," *J. Audio Eng. Soc.*, **48**, 642–653 (2000).

[18] M. Morimoto and H. Aokata, "Localization cues of sound sources in the upper hemisphere," *J. Acoust. Soc. Jpn.* (*E*), **5**, 165–173 (1984).

[19] W. Mills, "On the minimum audible angle," *J. Acoust. Soc. Am.*, **30**, 237–246 (1958).

[20] A. W. Mills, "Lateralization of high frequency tones," *J. Acoust. Soc. Am.*, **32**, 132–134 (1960).