19th INTERNATIONAL CONGRESS ON ACOUSTICS
MADRID, 2-7 SEPTEMBER 2007

ON THE IMPROVEMENT OF VIRTUAL LOCALIZATION IN VERTICAL DIRECTIONS USING HRTF SYNTHESIS AND ADDITIONAL FILTERING

PACS: 43.66.Qp

Wersényi, György[1]
[1]Széchenyi István University; Egyetem t. 1, Győr, H-9026, Hungary; wersenyi@sze.hu

ABSTRACT
50 untrained subjects participated in a virtual localization task using equalized headphone and HRTF synthesis. A 2D sound screen was simulated in front of the listener with spatial resolutions of 3x1, 3x3 and 5x2. Broadband noise stimuli were used and subjects reported in an absolute measurement about sound source directions. Based on the first results, a simple method was developed to decrease vertical errors. The simulation has been extended by additional high-pass and low-pass filtering that resulted in a rate of 80-90% of correct answers.

INTRODUCTION
Localization on Virtual Acoustic Displays (VAD) is widely investigated. Virtual simulation includes the equalization of the transmission line (first of all the headphone) and HRTF synthesis. The filtering of the Head-Related Transfer Functions are responsible for basic localization cues in free-field listening. Therefore, the measurement and synthesis of HRTFs can be critical in virtual audio environments [1, 2]. During simulation, subjects are wearing headphones and report about their sensation. These can be basic localization tasks (determining sound source locations) and/or several errors related to headphone playback (front-back reversals, in-the-head localization, elevation shift). It is well known that virtual localization is inferior to free-field listening. Furthermore, horizontal plane localization is superior to vertical localization due to existing interaural time and level differences. In the median plane, earsignals are very similar and it is difficult to localize sound sources correctly.
Our investigation is part of a project where icons and events of a computer screen are replaced or extended by sound events using appropriate directional simulation as well [3, 4]. Sound events have to be spatially distributed on a 2D virtual sound screen in front of the listener (Figure 1). Possibilities and localization behavior is investigated here focusing on vertical localization. Former results were presented already about spatial resolution in a series of minimum-audible-angle measurements as well as in an absolute measurement [4, 5].

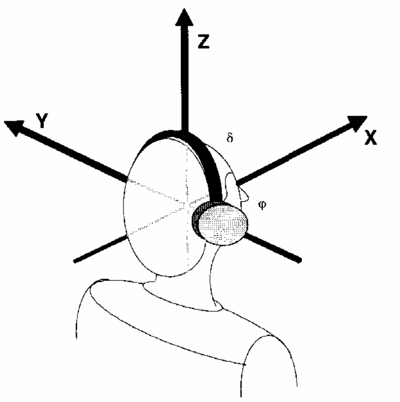

Figure 1.- The head-related coordinate system. The 2D virtual acoustic display is simulated in front of the listener parallel with the Z-Y plane [3-5].

MEASUREMENT SETUP

In the current investigation 50 untrained subjects with normal hearing participated. Measurements were made in the anechoic chamber. Excitation signal was a white noise burst-pair of 300 ms separated by 400 ms silence. Simulation is made through the equalized Sennheiser HD 540 headphone using the BEACHTRON sound card. This includes real-time filtering of Head-Related Impulse Responses (HRIRs) of a "good localizer" from the measurement of *Wightman and Kistler* [6, 7, 8]. Fig.2. presents the actual spatial resolutions of 3x1, 3x3 and 5x2.

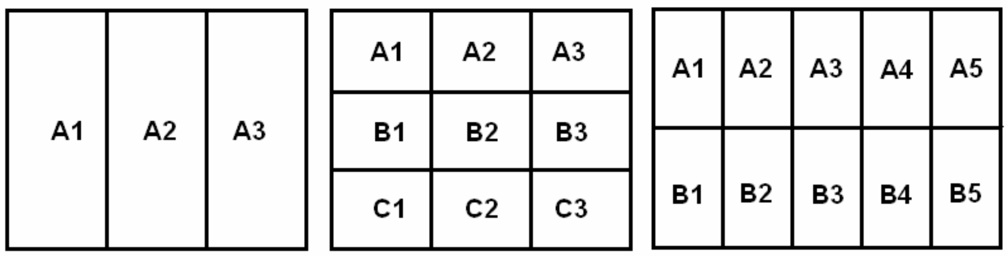

Figure 2.- Different spatial resolution of the 2D VAD: 3x1, 3x3 and 5x2. Sound sources are simulated in the middle of the surface elements.

During the measurement sound sources were presented randomly by simulating them in the middle of one of the surface elements marked with letters and numbers. Two rounds were completed for each subject (e.g. for 3x3 that means 18 answers for one subject and 20 answers for 5x2). Subjects were asked to call the corresponding field where the sound seems to be come from. Evaluation of the answers can be "true", "only vertical error" or "horizontal error". The latest means, they miss the target horizontally or horizontally and vertically.

In the second part a simple filtering was applied *additional* to the HRTF filtering as shown in Fig.3. High-pass filtering with a cut-off frequency of 7000 Hz and low-pass filtering with a cut-off frequency of 1500 Hz was applied on the white noise excitation (Fig. 4.). The idea is based on the psychoacoustical phenomenon that sounds having more high frequency content are often localized "high" as long signals with more low-frequency content "low" [1, 9, 10]. With this simple filtering we tried to bias correct answers.

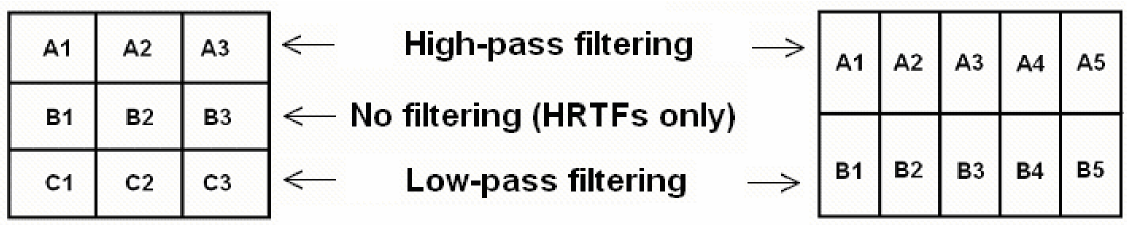

Figure 3.- A possibility for increasing correct answers. The input signal can be filtered by HPF and LPF filters before or after the HRTF filtering.

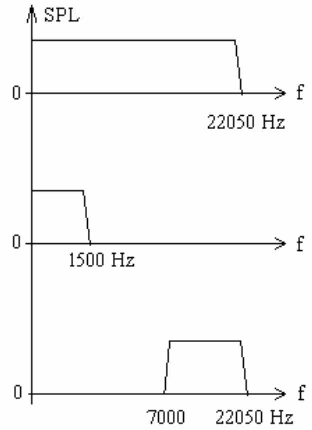

Figure 4.- Spectral representation of the applied signals: white noise, low-pass and high-pass filtered versions respectively.

RESULTS

Our previous study showed that using this playback system and HRTF filtering about 33% of the listeners could not localize vertically at all [11]. They make their judgments based on the spectral distortion of the HRTFs (as sound sources “sound different”) without real localization.

The resolution of 3x1 can be used in all cases, no errors appeared. The resolution 3x3 is suited for testing the vertical errors while 5x2 the angular errors. Table 1 and 2 summarize the results for white noise. For resolution 3x3 only 48% of the answers are correct, incorrect localization is mainly due to poor vertical localization. For resolution 5x2 only the rate of correct answers is about the same. Because of the two vertical positions and of the five possible horizontal locations, more false answers appear horizontally. Subjects were undecided even during giving correct answers by the far left and far right column: column 1 and 2 as well as column 4 and 5 are very hard to discriminate.

Table I.- Results for unfiltered white noise

3x3		
True	Only vertical error	Horizontal error
48%	48%	4%

5x2		
True	Only vertical error	Horizontal error
51%	29 %	20%

These results suggest better performance by increasing vertical localization and using 3-5 horizontal locations.

In the second part, we did not tell the subjects why filtering had been applied. However, they figured out soon what the measurement is about, so they realized that filtering should help them to solve the localization problem.

Results show significant improvement: 80-90% of the answers were correct (Table 2.). 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 necessarily 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 [12, 13]. Subject can distinguish between signals because they sound different rather than based on their location.

Table 2.- Results using additional LPF and HPF filtering on white noise

3x3		
True	Only vertical error	Horizontal error
92%	7%	1%

5x2		
True	Only vertical error	Horizontal error
82%	17%	1%

Figure 5 and Figure 6 show the results from a different perspective. They are designed based on Tables 1 and 2 and represent the effect for each resolution with and without additional filtering. Blue columns show measurement data using the HRTF synthesis only, red columns show results using the additional filtering. As expected, the rates for “true” answers increased significantly while the number of false answers decreased.

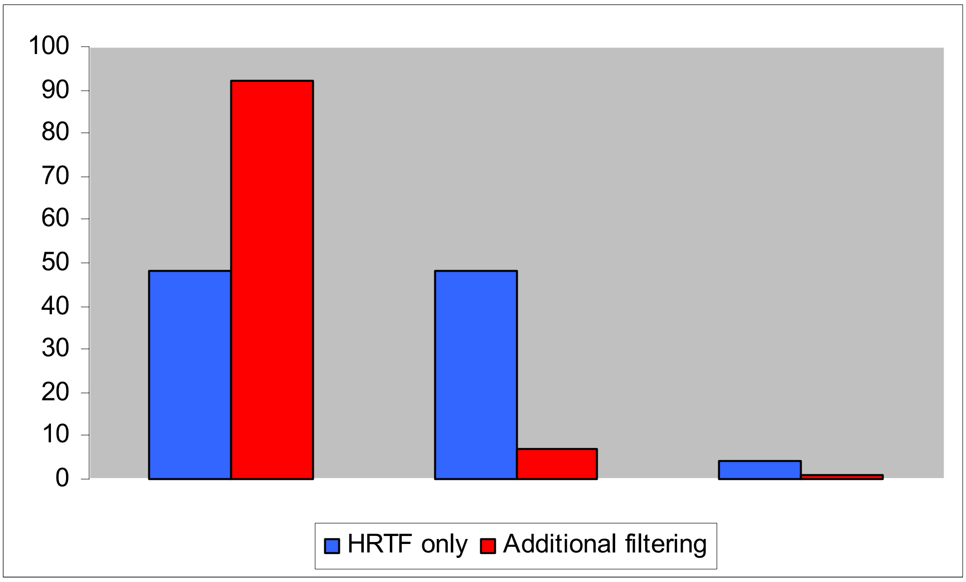

Figure 5 -Results for resolution 3x3: correct answers (left), only vertical error (middle) and horizontal error (right) with and without additional filtering based on Tables 1 and 2.

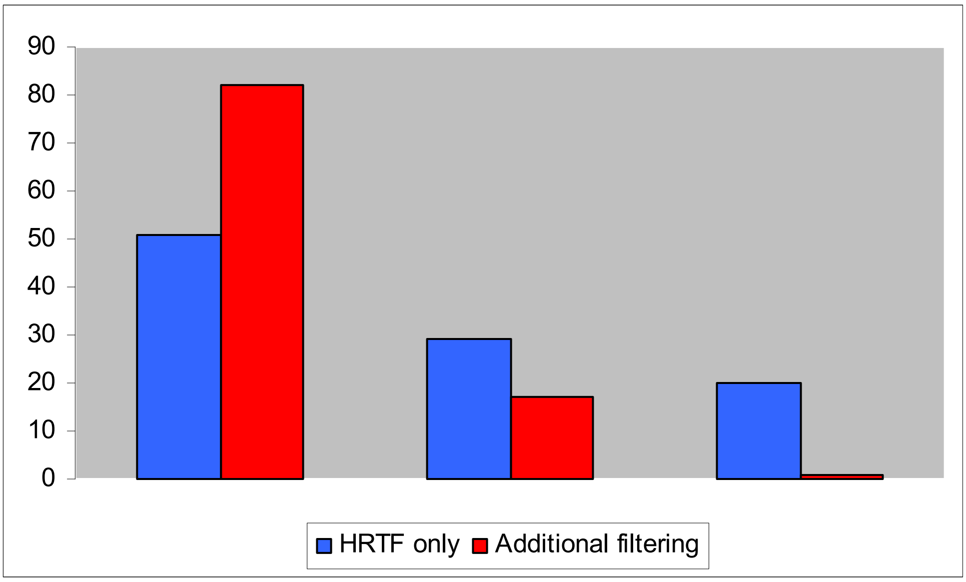

Figure 6 - Results for resolution 5x2: correct answers (left), only vertical error (middle) and horizontal error (right) with and without additional filtering based on Tables 1 and 2.

SUMMARY

Listening tests were carried out using headphone and HRTF synthesis on a 2D virtual acoustic display in front of the listener. Different spatial resolutions were evaluated of 3x1, 3x3 and 5x2 focusing on vertical localization. Using only the HRTF filtering, only about half of the subjects delivered correct answers. By applying a simple high-pass and low-pass filter on the broadband noise stimulus, correct answers can be biased up to 80-90%. This improvement in the answers is rather due to the spectral distortions of the filtering without real localization.

References:

[1] J. Blauert: Spatial Hearing. The MIT Press, MA, (1983)
[2] H. Møller: Fundamentals of binaural technology. Applied Acoustics **36** (1992) 171-218
[3] M. M. Blattner, D. A. Sumikawa, R. M. Greenberg: Earcons and Icons: their structure and common design principles. Human-Computer Interaction **4, No.1** (1989) 11-44
[4] K. Crispien, H. Petrie: Providing Access to GUI's Using Multimedia System – Based on Spatial Audio Representation. Audio Eng. Soc. 95th Convention Preprint (1993)
[5] Gy. Wersényi: What Virtual Audio Synthesis Could Do for Visually Disabled Humans in the New Era? Proceedings of the 12th AES Regional Convention Tokyo, (2005) 180-183
[6] Crystal River Engineering, Inc., BEACHTRON – Technical Manual, Rev.C., (1993)
[7] E. M. Wenzel, M. Arruda, D. J. Kistler, F. L. Wightman: Localization using nonindividualized head-related transfer functions. Journal of the Acoustical Society of America **94** (1993) 111-123
[8] F. L. Wightman, D. J. Kistler: Headphone Simulation of Free-Field Listening I.-II. Journal of the Acoustical Society of America **85** (1989) 858-878
[9] C. Tan, W. Gan: Direct concha excitation for the introduction of individualized hearing cues. Journal of the Audio Engineering Society **48, No.7-8** (2000) 642-653
[10] M. Morimoto, H. Aokata: Localization cues of sound sources in the upper hemisphere. Journal of the Acoustical Society of Japan **E5** (1984) 165-173
[11] Gy. Wersényi: Localization in a HRTF-based Minimum Audible Angle Listening Test on a 2D Sound Screen for GUIB Applications. Audio Engineering Society (AES) Convention Preprint Paper, Nr.5902, Presented at the 115th Convention, New York, (2003)
[12] W. Mills: On the minimum audible angle. Journal of the Acoustical Society of America **30** (1958) 237-246
[13] A. W. Mills: Lateralization of high frequency tones. Journal of the Acoustical Society of America **32** (1960) 132-134