



Science Arts & Métiers (SAM)

is an open access repository that collects the work of Arts et Métiers Institute of Technology researchers and makes it freely available over the web where possible.

This is an author-deposited version published in: <https://sam.ensam.eu>
Handle ID: <http://hdl.handle.net/10985/10826>

To cite this version :

Vladimir ORTEGA-GONZALEZ, Frédéric MERIENNE, Samir GARBAYA - Reducing reversal errors in localizing the source of sound in virtual environment without head tracking - In: 5th International Workshop on Haptic and Interaction Design HAID' 10, Denmark, 2010-09-16 - Haptic and Audio Interaction Design - 2010

Any correspondence concerning this service should be sent to the repository

Administrator : scienceouverte@ensam.eu



Reducing Reversal Errors in Localizing the Source of Sound in Virtual Environment without Head Tracking

Vladimir Ortega-González*, Samir Garbaya, and Frédéric Merienne

Arts et Metiers ParisTech, CNRS, Le2i
Institut Image, 2 rue T. Dumorey, Chalon-sur-Saône 71000, France
erikvladimir@gmail.com

Abstract. This paper presents a study about the effect of using additional audio cueing and Head-Related Transfer Function (HRTF) on human performance in sound source localization task without using head movement. The existing techniques of sound spatialization generate reversal errors. We intend to reduce these errors by introducing sensory cues based on sound effects. We conducted an experimental study to evaluate the impact of additional cues in sound source localization task. The results showed the benefit of combining the additional cues and HRTF in terms of the localization accuracy and the reduction of reversal errors. This technique allows significant reduction of reversal errors compared to the use of the HRTF separately. For instance, this technique could be used to improve audio spatial alerting, spatial tracking and target detection in simulation applications when head movement is not included.

1 Introduction

In this paper we present an approach to improve the accuracy of sound localization without human head tracking. Human auditory system has poor resolution in source localization [4]. However, according to the published research, the existing techniques of sound spatialization do not allow localizing sound sources with a resolution higher than provided by human natural hearing. The resolution could be very close to natural conditions depending on the employed model [21]. This leads to the conclusion that the existing techniques of spatial sound do not allow localizing sound source accurately.

Spatialized sound is used in different virtual reality applications for two main purposes: the simulation of acoustics in virtual environment. The second purpose is to enhance the quality interaction by means of spatial audio feedback.

The impact of 3D sound on presence and realism within the context of acoustics simulation has been already investigated (e.g. Hendrix and Barfield [9] and Larsson et al. [11]). The related studies concluded that the contribution of 3D sound is significant for the sensation of presence but not for realism.

* Corresponding author.

The use of 3D sound in interactive applications has been studied by different authors. The main related applications are concerned with the audio spatial tracking, the audio spatial alerting and the assisted navigation.

The audio-assisted navigation refers to the use of 3D sound to provide the user with information to assist the navigation in unknown environment. This application of 3D sound has been subject of different research works (e.g. Lumbreras and Sánchez [13], Lokki et al. [12] and Walker and Lindsay [20]).

The audio spatial tracking refers to the capacity of detecting and following a moving target represented by a sound source. The effect of 3D sound for spatial tracking was studied by Mckinley et al. [15]. The authors found that the use of spatial auditory information contributes to better target detection. Bronkhorst et al. [3] presented a study of the effect of 3D sound for tracking targets in a flight simulator. They found that 3D sound is suitable for complementing and for replacing visual displays such as the radar.

The audio spatial alerting refers to the displaying of spatially localized audio alerts. These spatial audio alerts can be more easily differentiated and they can convey spatial information. Ho and Spence [10] investigated on the benefits of audio spatial alerting in potentially dangerous situations in driving. According to their results, the use of spatial audio allows an effective way of capturing the attention of the driver better than non-spatial alarms.

The approach presented in this paper is based in the combination of the commonly used spatialization technique (HRTF: Head-Related Transfer Function) with additional sensory cues. The result is that 3D sound is enriched with additional audio cues. The existing techniques related to audio spatial tracking and alerting allow acceptable results but the reduction of reversal errors and the improvement of localization accuracy could produce better interaction quality.

The presence of reversal errors in audio spatial tracking and alerting can induce errors in the interaction and reduce human performance. These errors can be potentially important depending on the application. For instance, a target or an alert which is identified by the user as coming from a wrong direction. In these applications, the listener normally do not have time to make use of head movement to localize sound sources, for these reasons we focus our study on localization task when information about head movement is not provided.

2 Related Work

In this section two main issues are presented: the technique of reference of sound spatialization based on the HRTFs and the existing approaches to improve sound source localization by human user in virtual environment.

2.1 Sound Source Localization and the Head Related Transfer Function

The Head Related Transfer Function (HRTF) is a filter which expresses how an acoustic wave is affected by the anatomy of the listener. Begault [1] stated

that HRTF represents the spectral filtering which occurs before the arrival of the sound to the internal ear drum. These filters recreate the natural altering effect on listened sounds caused by the morphology of each auricle as well as by the diffraction and reflection effects due to the head and shoulders respectively. These effects vary from individual to individual forcing systems designers to choose between the use of individualized or generalized filters.

Since interaural time and intensity differences are ambiguous under certain conditions, generalized HRTFs and more particularly customized HRTFs, are useful for reducing the presence of front-back and up/down (reversals) errors and of cones of confusion [1]. However, the implementation of customized HRTF requires specialized facilities, specialized equipment and specialized audio treatment processes. This is time consuming and relatively expensive. These elements make difficult the use of customized HRTF in virtual environment particularly considering that the cost of other immersive components is also high.

Wenzel et al. [21] conducted a series of experiments to evaluate the pertinence of non-individualized HRTF for virtual acoustic displaying. Authors observed that generalized HRTF makes users obtain a very similar angular accuracy in both real conditions and with 3D sound rendering. They also noticed that the use of generalized functions increases the rate of reversal errors.

Begault et al. [2] conducted an experiment to compare the effect of magnetic head-tracking, reverberation and generalized and individualized HRTF with non-speech sound for spatial displaying. They worked with a group of untrained subjects that indicated their judgements by using a graphic interface. According to their results, head tracking significantly helps to reduce angle errors and particularly to avoid reversals; reverberation helps azimuth precision; and the use of generic HRTF only affects slightly the perception compared to individualized HRTF. The reported mean angle errors vary from 15 to 25 degrees for both azimuth and elevation. This work is relevant to our study because the experimental conditions of the study described in this paper are similar.

2.2 Improving Sound Localization in Virtual Environment

There exist different approaches aiming at improving the localization accuracy of sound sources in virtual environment. Durlach et al. [4] presented the approach of supernormal auditory localization. This proposal is based on altering the azimuth of the used HRTF based on a transformation (mapping) function. A complement of this work can be found in Shinn-Cunningham et al. ([18] and [19]). This work is restricted to sound sources localized in the horizontal plane.

Gronh and Takala [7,8] presented an approach named MagicMikes for the sonification of localized data sets into a navigation context. This approach considers the use of audio cueing and audio spatialization. Gronh et al. [6] presented a study for evaluating localization accuracy in immersive environment with multichannel systems. The obtained results vary from 11 to 14 degrees in azimuth and from 22 to 27 in elevation.

Marentakis and Brewster [14] conducted an experiment of comparing different interactive audio cues for improving the efficiency of localization of sound

sources. They tested different sources positioned in the horizontal plane. They obtained an accuracy error vary from 4 to 15 degrees in azimuth.

Lokki et al. [12] presented a study of the effect of spatial audio cueing on user performance for the navigation of virtual environments. The authors found that audio cueing have significant effect on performance (in terms of execution time) and that it is possible to navigate using audio cues only. One interesting point of this work is the combination of parametric cues with sound spatialization for improving sound source localization in elevation.

In contrast to the approach presented in this paper, the existing techniques relating to the improvement of sound localization are limited to the case where sound sources are located in the horizontal plane. The objective of our approach is to assist localization in both elevation and azimuth.

3 The Approach of HRTF and Additional Cues

The approach is based on spatial sound sources enriched with specialized cues. These cues provide the user with information which is intended to assist them in the execution of specific tasks. Each audio cue is associated to a parametrized sound effect. This association is determined by a mapping function defined by a behavior curve.

The proposed approach considers that the additional cues enrich the signal before the application of the HRTF model. In order to avoid the technical requirements of individualization, a generalized HRTFs is used. The HRTF database used in the work described in this paper is the KEMAR HRTF database created by Gardner [5] at the MIT Media Lab Machine Listening Group.

The used additional cues are detailed in Table 1. For each cue the associated audio effect as well as the corresponding parameter are specified. These parameters depend on the angular difference between the user head and the direction of origin of the sound source. The mapping of each effect and the corresponding parameter is specified by a behavior curve (Figure 1). These curves were defined empirically and adjusted during the pre-tests. The values of the parameters vary dynamically depending on the user activity.

The term of frontality refers to the ability to distinguish whether the sound source is localized in the front or at the back of the user. The technique is to use a lowpass filter that changes the character of the sound when it is localized at the back of the user. Thus, subjects perceive an occluded sound for absolute azimuth angles bigger than 90 degrees. The corresponding behavior curve is specified in figure 1a.

Table 1. Additional cues

Cue	Effect	Variable	Parameter	Range
Verticality	Reverb	Level	Elevation	$[-10k, 2k]mdB$
Horizontality	Attenuation	Factor	Azimuth	$[-40, 0]dB$
Frontality	Highpass	Cutoff Freq.	Azimuth	$[1, 22k]Hz$
Angular Proximity	Lowpass	Cutoff Freq.	Overall angle	$[200, 22k]Hz$

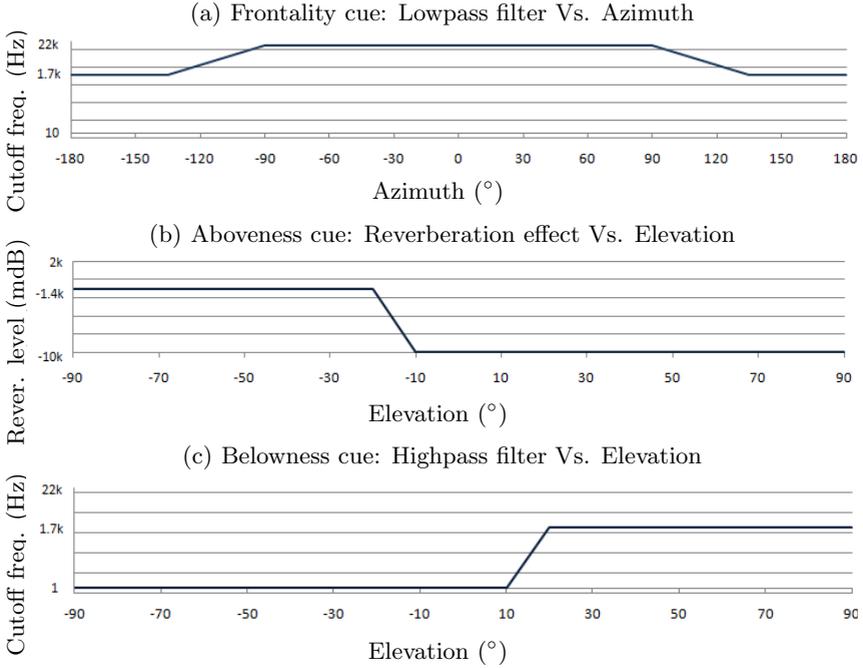


Fig. 1. Behavior curves characterizing the response of the different applied effects in terms of the defined parameters of the additional cues. The vertical axis of subfigures (a) and (c) are plot in logarithmic scale.

The term belowness refers to whenever the sound source is localized below the horizontal reference plane of the listener. For cueing this information we use the reverberation effect. In this manner, the sound sources in low elevations will be distinguishable by this extra reverberation. The corresponding behavior curve is shown in the figure 1c.

The term aboveness refers to whenever the sound source is localized above the listener. The technique is to use a highpass filter that changes the character of the sound when it is localized in high elevations. In this manner, the sound is more acute in high elevations. The behavior curve characterizing this cue is shown in the figure 1b.

The selected sound stimulus was obtained from the sound library Sound Rangers (2009) [16] and its waveform has a clink shape. It is a brief sound with a duration of 0.5 seconds approximately and it is reproduced continuously. The main criteria for selecting this kind of waveform is that brief stimuli have commonly less information to be decoded compared to a voice signals and that it could be more intelligible (easy to recognize), less diffused and probably less annoying than almost any kind of noise signal.

4 Experiment Design

The objective of the experiment is to determine the contribution of the additional cues when added to the HRTF in the localization accuracy and the rate of detected reversal errors. The source localization is made by a pointer. Head movement is not available. The employed HRTF is generalized. The sound display is made by closed headphones.

4.1 Task and Experimental Conditions

The task consists of localizing the provenance of a sound source with a pointer. The subject is asked to indicate the localization of the sound source by Novint Falcon device produced by Novint Technologies [17]. The pointer is restricted to move over the surface of a localization sphere with fixed radius. The figures 2a and 2b show the graphic interface and a photograph of the experiment respectively.

Before executing the experiment, each subject has one minute approximately to get familiarized with the additional audio cues. Then, the mechanism of the additional cues is explained to the subject. The user has the possibility to test freely during one minute his sensation of a sound source which position is attached to the pointer. The table 2 presents the coordinates of the different fixed sound sources used in this experiment. The experimental variables are recorded for all these repetitions. These different localizations are presented to each subject randomly.

The experimental conditions are (1) HRTF and (2) Additional cues + HRTF. The condition (1) corresponds to the use of HRTF only. The condition (2) refers to the sound stimulus enriched by the cues and spatialized with the HRTF. Each

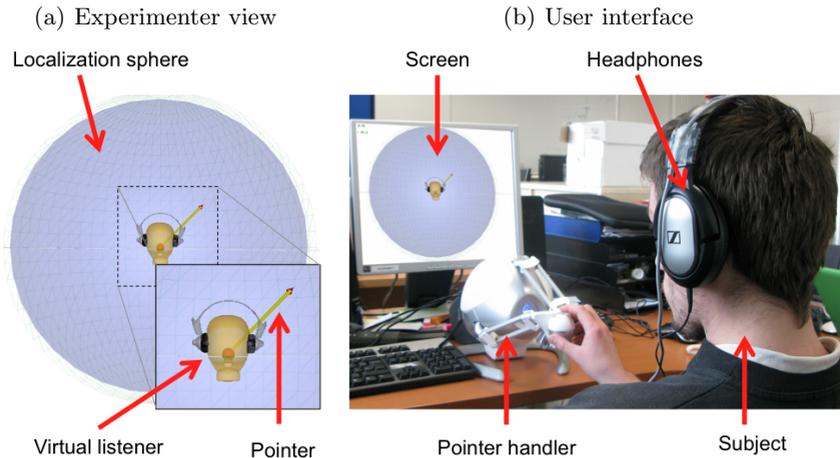


Fig. 2. Experiment setup: (a) screenshot of the graphic interface and (b) photograph of a subject executing the task

Table 2. Different sound source locations

Rep.	Azimuth (φ_0)	Elevation (δ_0)	Rep.	Azimuth (φ_0)	Elevation (δ_0)
1	+0.00	+0.00	10	+180.00	+50.00
2	+60.00	+0.00	11	-60.00	+50.00
3	+120.00	+0.00	12	-120.00	+50.00
4	+180.00	+0.00	13	+0.00	-50.00
5	-60.00	+0.00	14	+60.00	-50.00
6	-120.00	+0.00	15	+120.00	-50.00
7	+0.00	+50.00	16	+180.00	-50.00
8	+60.00	+50.00	17	-60.00	-50.00
9	+120.00	+50.00	18	-120.00	-50.00

user performs the experimental conditions in random order. The random order of repetitions and experimental conditions is included to reduce the carry-over effects.

4.2 Group of Subjects

Fifteen subjects (thirteen male and two female) aged from 20 to 30 years old participated in this experiment. They are right-handed and did not report any visual or audio deficiency. They are university students without prior experience with 3D interaction devices.

Each user had to search for the localization of 18 different sound sources in 2 different experimental conditions, which makes 36 records of measures for each participant. The whole experiment, carried out by each subject, took approximately 25 minutes to be completed. Task execution in both experimental conditions are separated by a break of three minutes approximately.

4.3 Performance Measures

The user performance is measured in terms of the following objective variables: the execution time (in seconds) and the angular errors (in degrees). The angular errors are the azimuth, the elevation and the overall angular error. The overall angular error is the angle between the vectors \vec{U} and \vec{V} that represent the localization of the sound source and the orientation of the listener’s head respectively.

5 Results and Data Analysis

The evaluation of the user performance is based on the measurement of the task execution time, the absolute angular errors (azimuth, elevation and overall angular error) and the percentage of reversal errors (front/back, up/down and left/right). In order to determine the effects of each audio stimulation technique, we performed a Generalized Linear Equations (GEE) model using the collected data.

The table 3 presents the descriptive statistics of the variables of execution time and angular errors for both experimental conditions. Only the valid observations were taken into account. Because the distribution of the data was not assumed to be of normal type, the median and the quartile information were added to the typical statistics measures of mean value and standard deviation. This is carried out in order to provide a more general statistical characterization of the data. The use of the combination of additional cues and HRTF allows reducing the levels of the angular errors. The statistical significance of these difference will be verified later on.

Table 3. Descriptive statistics of the experimental variables for the different conditions (factors) of observations

Condition	Variable	N	Mean	Median	Quartiles			Std. dev.
					Lower	Upper	Range	
1 HRTF	Execution time (s)	252	9.02	7.57	4.94	11.65	6.71	5.49
	Overall error (°)	252	66.64	56.36	37.17	87.97	50.80	41.27
	Azimuth error (°)	252	60.67	39.13	20.41	84.62	64.21	55.21
	Elevation error (°)	252	33.44	31.07	11.50	50.27	38.77	26.40
2 Additional cues & HRTF	Execution time (s)	252	9.87	8.46	5.72	12.43	6.71	5.95
	Overall error (°)	252	39.33	35.75	20.59	47.98	27.39	27.14
	Azimuth error (°)	252	28.50	18.77	4.70	35.32	30.63	35.00
	Elevation error (°)	252	25.79	21.88	9.95	38.70	28.76	19.72

Note: All the statistics measures of the angular errors (azimuth, elevation and overall) were calculated using the absolute values of the corresponding datasets.

The figure 3a shows the boxplots and the mean values of execution time for the experimental conditions. The descriptive values and the dispersion are similar for both conditions. The figure 3b shows the boxplots and the mean values of overall angular error. It is noticed that the descriptive values and the dispersion of data are lower for the condition of combining the additional audio cues and HRFT than for the HRTF only.

The table 4 shows the percentage of the different types of reversal errors for the two experimental conditions. The use of the combination of additional audio cues and HRTF allows reducing the presence of reversal errors. The figures 4a and 4b show the graphical representation of the detected reversals of front/back, and up/down for the experimental conditions respectively. The combined model reduces in approximately 38%, 16% and 10% the rates of front/back, the up/down and the left/right reversals respectively.

The table 5 shows the effects of combining the additional cues with the HRTF on the different experimental measures. The difference in execution time between the experimental conditions is not statistically significant ($p = 0.303$). The effects of the additional cues on the angular errors are statistically significant. The additional cues have statistically significant effects on the rates of detected reversal errors.

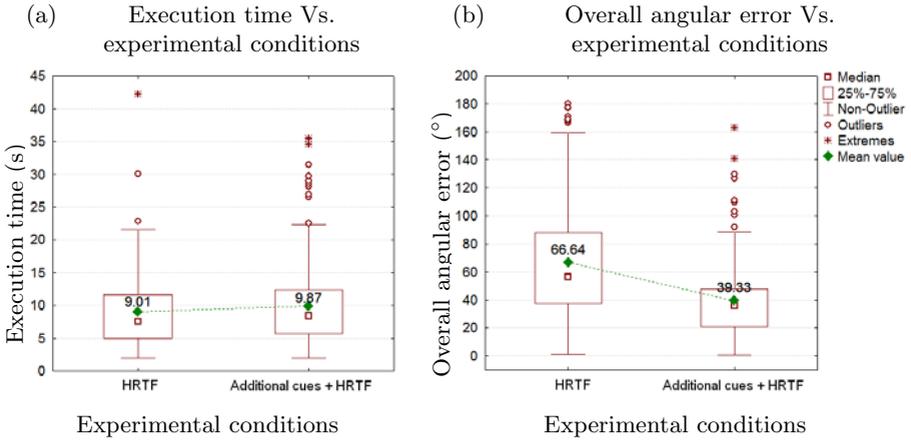


Fig. 3. Box plots and mean values of execution time (a) and overall angular errors (b) for the two experimental conditions

Table 4. Reversal errors for the experimental conditions

Exp. condition	Type of reversal	Observations w/o reversals	Detected reversals	Percent of reversals
1 HRTF	Front/back	129	123	48.81%
	Up/down	177	75	29.76%
	Left/right	216	36	14.29%
2 Additional cues & HRTF	Front/back	225	27	10.71%
	Up/down	218	34	13.49%
	Left/right	240	12	4.76%

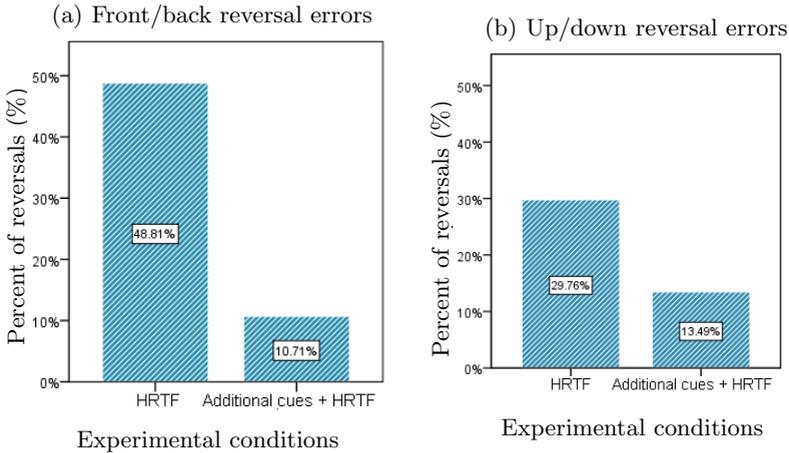


Fig. 4. Percentage of front/back (a) and up/down (b) reversal errors for the experimental conditions

Table 5. Analysis of the effect of factors by the Repeated Measurements Analysis under the Generalized Estimating Equations model

Factor	Variable	Model	Wald χ^2	DF	<i>p</i>	Sig.
Exp. conditions	Execution time (s)	Log-normal	1.061	1	.303	N
	Overall error (°)	Log-normal	34.177	1	.000	Y
	Azimuth error (°)	Exponential	34.478	1	.000	Y
	Elevation error (°)	Exponential	13.320	1	.000	Y
	Front/back reversals (%)	Binomial	94.316	1	.000	Y
	Up/down reversals (%)	Binomial	23.050	1	.000	Y
	Left/right reversals (%)	Binomial	16.525	1	.000	Y

Details of the model: The estimation of the effects is based on the Wald chi-square (χ^2) statistic. The test is of type III for which the order of aggregation of the factors is not relevant. The probability distribution model is indicated for each the variable in the column Model. The confidence level is 95%.

6 Conclusions and Future Work

In this paper we presented an evaluation work of the the approach of combining additional audio cues with the HRTF for sound localization task without the provision of information about head movement. The enriched 3D sound is based on the spatialization technique of HRTF and a group of additional cues associated to sound effects. These cues enrich the sound source with information that intends to assist the user in the localization task. The results show that the addition of audio cues to the HRTF has significant benefits for the improvement of the localization accuracy and the reduction of the rate of detected reversal errors.

Head tracking and customized HRTFs have been typically considered as means to reduce the presence of reversal errors. Nevertheless, head tracking and the use of customized HRTFs are relatively expensive and not suitable for applications of 3D sound of spatial alerting and tracking. The combination of the additional cues and HRTF allows reducing the rates of reversals errors in the absence of head tracking and customized HRTFs.

The future work includes different issues such as the use of head tracking and the comparison of different pointing gestures. The proposed technique is suitable for applications that intend to improve spatial tracking and alerting in simulation such as flight and driving simulation.

By adding information that do not exist in real conditions, the additional audio cues make spatial sound less consistent with the natural phenomena of spatial hearing. Nevertheless, the approach described in this paper and the related applications such as spatial tracking and alerting do not use spatial sound to simulate real acoustics. In these applications, spatial tracking and spatial alerting make use of spatial sound for cueing rather than to provide the user with meaningful spatial information. In this case, the consistency of spatial audio with real conditions is not important.

References

1. Begault, D.: 3D-Sound for Virtual Reality and Multimedia. AP Professional, USA (2005)
2. Begault, D.R., Wenzel, E.M., Anderson, M.R.: Direct Comparison of the Impact of Head Tracking, Reverberation, and Individualized Head-Related Transfer Functions on the Spatial Perception of a Virtual Speech Source. *Journal of the Audio Engineering Society* 49(10), 916 (2001)
3. Bronkhorst, A.W., Veltman, J.A., van Breda, L.: Application of a three-dimensional auditory display in a flight task. *Human factors* 38(1), 23–33 (1996)
4. Durlach, N.I., Shinn-Cunningham, B.G., Held, R.M.: Supernormal auditory localization. I. General background. *Presence* 2(2), 89–103 (1993)
5. Gardner, B., Martin, K.: HRTF Measurements of a KEMAR Dummy-Head Microphone. MIT Media Lab Perceptual Computing, USA (1994)
6. Grohn, M., Lokki, T., Takala, T.: Static and dynamic sound source localization in a virtual room. In: *Proc. AES 22nd Int. Conf. on Virtual, Synthetic and Entertainment Audio*, pp. 15–17 (2002)
7. Grohn, M., Takala, T.: MagicMikes - Multiple Aerial Probes for Sonification of Spatial Databases. In: *Int. Conf. on Auditory Displays (ICAD 1994)*, Santa Fe, USA, p. 108 (1994)
8. Grohn, M., Takala, T.: MagicMikes - Method for Spatial Sonification. In: *IS&T/SPIE Symposium on Electronic Imaging: Science & Technology*, San Jose, USA, p. 108 (1995)
9. Hendrix, C., Barfield, W.: Presence in virtual environments as a function of visual and auditory cues. In: *VRAIS 1995: Proceedings of the Virtual Reality Annual International Symposium (VRAIS 1995)*, p. 74. IEEE Computer Society, Washington (1995)
10. Ho, C., Spence, C.: Assessing the effectiveness of various auditory cues in capturing a driver's visual attention. *Journal of Experimental Psychology: Applied* 11(3), 157–174 (2005)
11. Larsson, P., Vastfjall, D., Kleiner, M.: Effects of auditory information consistency and room acoustic cues on presence in virtual environments. *Acoustical Science and Technology* 29(2), 191–194 (2008)
12. Lokki, T., Grohn, M.: Navigation with Auditory Cues in a Virtual Environment. *IEEE MultiMedia* 12(2), 80–86 (2005)
13. Lumbreras, M., Sánchez, J.: Interactive 3D sound hyperstories for blind children. In: *CHI 1999: Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 318–325. ACM, New York (1999)
14. Marentakis, G., Brewster, S.A.: A comparison of feedback cues for enhancing pointing efficiency in interaction with spatial audio displays. In: *MobileHCI 2005: Proceedings of the 7th international conference on Human computer interaction with mobile devices & services*, pp. 55–62. ACM, New York (2005)
15. McKinley, R., D'Angelo, W.R., Hass, M.W., Perrot, D., Nelson, W., Hettinger, L., Brickman, B.: An initial study of the effects of 3-dimensional auditory cueing on visual target detection. In: *39th Human Factors and Ergonomics Society Annual Meeting*, USA, pp. 119–123 (1995)
16. Music, R.F.: Sound Rangers sound library (2009) (Sound Effect Reference: ding-wip01), <http://www.soundrangers.com/>
17. Novint Technologies Inc.: Novint Falcon User Manual (2007), <http://home.novint.com/>

18. Shinn-Cunningham, B.G., Durlach, N.I., Held, R.M.: Adapting to supernormal auditory localization cues. I. Bias and resolution. *Journal of the Acoustical Society of America* 103(5), 3656–3666 (1998)
19. Shinn-Cunningham, B.G., Durlach, N.I., Held, R.M.: Adapting to supernormal auditory localization cues. II. Constraints on adaptation of mean response. *Journal of the Acoustical Society of America* 103(6), 3667–3676 (1998)
20. Walker, B.N., Lindsay, J.: Navigation Performance With a Virtual Auditory Display: Effects of Beacon Sound, Capture Radius, and Practice. *Hum. Factors* 48(2), 265–278 (2006)
21. Wenzel, E., Arruda, M., Kistler, D., Wightman, F.: Localization using non-individualized head-related transfer functions. *Journal of the Acoustical Society of America* 93(1), 111–123 (1993)