Audio Engineering Society

Convention Paper 7884

Presented at the 127th Convention
2009 October 9–12 New York, NY, USA

An Investigation of Early Reflection's Effect on Front-Back Localization in Spatial Audio

Darrin K. Reed and Robert C. Maher

Department of Electrical and Computer Engineering
Montana State University, Bozeman, MT 59717-3780 USA

ABSTRACT

In a natural sonic environment a listener is accustomed to hearing reflections and reverberation. It is conceived that early reflections could reduce front-back confusion in synthetic 3-D audio. This paper describes an experiment to determine whether or not simulated reflections can reduce front-back confusion for audio presented with non-individualized HRTFs via headphones. Although the simple addition of a single-order reflection is not shown to eliminate all front-back confusions, some cases of lateral reflections from a side boundary can be shown to both assist and inhibit localization ability depending on the relationship of the source, observer and reflective boundary.

1. INTRODUCTION

Accurate localization of a stationary sound source remains as a problem with generalized head-related transfer functions (HRTFs). Since HRTFs account for the reflective paths of sound caused by one's pinna, head, and torso, it seems reasonable that other (non-anthropometric) reflective paths could be important for the accurate localization of virtual sound sources. Many HRTF localization studies have presented stimuli in virtual free-field simulations. However, in natural sonic environments we are generally located near at least a single (ground) boundary.

For this investigation we seek to determine if a single reflection can disambiguate front-back confusions when using generalized HRTFs. A range of reflection amplitudes due to various reflective boundaries will be added to virtually placed speech signals. Listeners will then be presented via headphones with these spatialized stimuli to determine the effect of early-order reflections on front-back localization with generalized HRTFs.

2. BACKGROUND

The investigation of sound and binaural hearing dates to the late eighteenth century from research conducted by Chladni & Venturi [1]. The theories of varying arrival timings and levels were confirmed a century later by Lord Rayleigh which came to be called the Duplex Theory of sound [2]. Rayleigh conceived two binaural cues: inter-aural time difference (ITD) and inter-aural intensity difference, more commonly referred to as inter-aural level difference (ILD). However, this binaural model proved to have limitations, primarily along equidistant curves around the head, termed cones of confusion [3]. This ambiguity of front-back localization has led most modern research to the use of head-related transfer functions for sound spatialization techniques.

It is well know that individualized HRTFs allow for more accurate localization as compared to generic HRTFs [4, 5]. Individualized HRTFs have even succeeded in subjectively presenting sound in comparable localization accuracy to free-field stimuli [6]. However, the difficulty in rapidly obtaining individualized HRTFs has led many to the use of generalized HRTFs, but these non-individualized transfer functions have shown degraded front-back localization performance [4]. Much research has sought to improve the generic HRTFs through many techniques including the anthropometric customization [7-9] and acoustic raytracing of HRTFs [10]. One proven contribution to the reduction of front-back confusions when using generalized HRTFs is allowing listeners to control head movement while keeping the sound stimuli stationary [11-13]. However, the accurate localization of stationary signals is beneficial in other applications [14].

Since listeners are accustomed to localizing sounds in non-anechoic environments, the strict presentation of a monaural signal processed solely by a set of anechoic HRTFs is an unnatural auditory display. [15, 16] have shown that virtual environments help to externalize sounds and create a higher degree of realism. Thus, the use of these room models could help to present a more natural display and aid in sound localization. Although the necessity of room reflection information for localization ability was questioned [17, 18], it has been theorized that the first few reflections could assist in the location of sound within a room environment [19, 20].

Multiple techniques of sound localization were compared by [21], including wavetracing methods to develop a room model for spatial perception analysis. Subjective results reported that head movement was the only factor to show significant improvement in front-back reversals. No specific claim regarding the contribution of early reflections to front-back reversal accuracy was made. The design of a virtual auditory system [22] implemented early reflections as a significant component, however, the analysis of subjective responses focused on sound externalization and no conclusions regarding improved sound localization were made. Experiments in the present investigation seek to provide better certainty on the contribution of early-order reflections to sound localization ability.

3. METHODS

The motivation behind the following experiments lie in the desire to determine if early-order reflections have a role in the localization of virtually spatialized headphone audio. For this investigation three separate experiments were conducted. First, listeners were asked to detect the audibility of various levels and types of reflections. Second, a baseline localization ability was established with the anechoic presentation of speech samples in four eye-level quadrants, and finally, varying degrees and types of reflections were added to the anechoic speech samples for a similar four-quadrant localization experiment.

3.1. Apparatus

All experiments were conducted in an acoustically treated, quiet room (avg. SPL < 23dBA). Stimuli were presented through Denon AH-D2000 headphones which were fed by a Headroom Ultra Micro Amp and Headroom Ultra Micro DAC. Circumaural headphones were used due to the ease of consistent placement on the subjects' head. The SPL at the headphones was calibrated to 75dB at 1kHz using a Type 2236 B&K sound level meter and a flat-plate headphone coupler similar to [23, 24]. At this 75dB reference level, subjects should not experience any temporary threshold shift from the time exposure of these experiments [25].

3.2. Subjects

Thirteen untrained subjects (nine male, four female) between the ages of 20 and 28 participated in the localization experiments. Only five of the thirteen participated in the reflection detection experiment, so that the length of the primary experiments could be kept shorter and attention lapses would not become a factor. Subjects were screened on six octave-band center frequencies at 250, 500, 1k, 2k, 4k, and 8kHz using [26]'s implementation of the classic adaptive maximum-likelihood procedure [27]. No headphone equalization was applied in this entire investigation, so, the SPL of the tones varied on the frequency response of the headphones. Table 1 shows the measured starting SPLs for each frequency.

Freq. (Hz)	250	500	1k	2k	4k	8k
SPL (dBA)	53.8	57.7	63.2	58.1	60.9	52.8

Table 1: Starting SPLs for hearing screening.

Of the thirteen subjects, two of the participants reported clinically-classified hearing problems and two others mentioned declined hearing though not medically defined. It can be seen in Figure 1 that the hearing screening detected a reduced trend in thresholds for these four individuals (Subj. 10-13). The nine remaining subjects (Subj. 1-9) will be referred to as Group A and the other four as Group B.

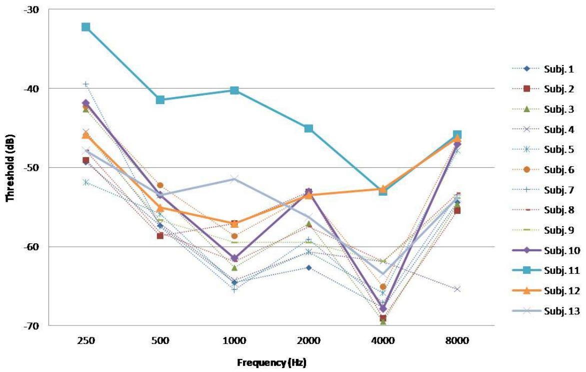


Figure 1: Hearing screening results for the thirteen participants, Group A (1-9) and Group B (10-13).

3.3. Stimuli

Short (1-2 sec.) speech phrases were recorded in a Model 802 Ray Proof Sound Shield rehearsal room using an AKG C414 microphone and Tascam HD-P2 recorder. All samples were recorded at 48kHz, 24bit. Forty phrases were read by a male speaker whose fundamental vocal frequency was slightly below 100Hz. In order to utilize a majority of the audible frequency range for localization, ten of the forty phrases containing extra fricative content were selected.

Spatialization processing of the speech samples was completed using the KEMAR large pinnae HRTF dataset (Subj. 165) from the CIPIC database [28]. Speech samples were virtually placed in four eye-level quadrant locations around the head with locations 45° off the coronal and midsagittal planes.

For these experiments it was desired to have no spectral modification from the raw HRTF measurements. Source and boundary distances were carefully chosen so that no interpolation was necessary. Using the image-source method [29], reflections were created for three boundary cases: right-wall (6ms & 20ms delays) and floor (6ms delay). A 20ms delay for the floor boundary was not tested since it would require an impractical observer-boundary relationship.

A broad range of reflection levels were considered. These simulated reflections were simply variations in gain with no spectral change. Since the goal of this investigation was to determine if any type of first-order reflection would contribute to front-back localization ability, the range of relative levels extended to reflections not physically possible. Table 2 shows a list of reflection factors (ρ) for each of the boundaries investigated, where $\rho = 1$ corresponds to perfect reflection and $\rho = 0$ corresponds to perfect absorption. A $\rho > 1$ would be considered an active boundary, thus not reasonable in real-world environments.

	Floor Boundary	**Right-wall Boundary**			
	6ms delay	6ms delay		20ms delay	
	All	Left	Right	Left	Right
-6dB	0.76	0.85	1.53	2.22	2.13
-9dB	0.54	0.60	1.09	1.57	1.51
-12dB	0.38	0.42	0.77	1.11	1.07
-15dB	0.27	0.30	0.54	0.79	0.76
-18dB	0.19	0.21	0.38	0.56	0.54
-21dB	0.13	0.15	0.27	0.39	0.38
-24dB	0.10	0.11	0.19	0.28	0.27
-27dB	0.07	0.08	0.14	0.20	0.19

Table 2: Reflection factors (ρ) for various levels of reflections in right & left hemispheres.

3.4. Reflection detection

The first experiment was used to determine thresholds of audibility for reflections since the testing of non-audible reflections would provide no insight on their contribution to front-back localization. Five subjects from Group A were presented with two versions of the same virtually placed speech sample, however, one of the two samples potentially contained a reflection. One trial per quadrant for each permutation of boundary case (floor 6ms, wall 6ms & wall 20ms) and reflection amplitude relative to the direct signal (-6dB, -9dB, -12dB, -15dB, -18dB, -21dB, -24dB & -27dB) were tested. Participants were limited to a binary (detect/no-detect) response and indicated their choice via keyboard through a MATLAB interface. This experiment took participants approximately 15 minutes to complete.

3.5. Anechoic front-back localization

To establish the localization ability in an anechoic presentation, all thirteen subjects from Groups A & B were presented with speech samples virtually placed at four eye-level quadrant locations. Each subject was presented with 15 trials in each of the quadrant locations. Participants were limited to a binary (front/back) response and indicated their choice via keyboard through a MATLAB interface. The test generally took less than 10 minutes to complete.

Prior to beginning this experiment, participants listened to instructions and two virtually placed examples per quadrant. Both the instructions and examples were read by the same voice as the test samples.

3.6. Non-anechoic front-back localization

The non-anechoic experiment took results from the reflection detection experiment to determine appropriate relative thresholds. Only the reflection conditions shaded in Table 3 were used for this experiment. Two trials for each permutation of quadrant and boundary condition were tested for the -6dB case; three trials were used for all other relative amplitude cases (-9dB to -18dB). To supplement the anechoic baseline experiment, three trials without reflections were also presented per quadrant. This resulted in a total of 144 trials per experiment and took participants approximately 15 minutes to complete. Again, participants were limited to a binary (front/back) response and indicated their choice via keyboard input.

4. RESULTS

4.1. Reflection Detection

Table 3 shows the results of the reflection detection experiment where the shaded cells correspond to reflection levels used in the non-anechoic experiment. Although the thresholds of detection presented here do not directly correspond with the "rules of thumb" developed by [30], the data does agree that longer delay times have a detection threshold ~9dB below that of a shorter delay. From Table 4 it is apparent that the detection of a reflection is much greater for the left side, i.e. side opposite the wall boundary. Again, this is in agreement with the claim in [30] that larger lateral differences between the direct and reflected signals result in a lower threshold of detection.

Relative dB	Floor 6ms	Wall 6ms	Wall 20ms
-6dB	55%	95%	100%
-9dB	25%	95%	95%
-12dB	0%	50%	100%
-15dB	0%	45%	50%
-18dB	15%	35%	50%
-21dB	0%	20%	15%
-24dB	10%	10%	5%
-27dB	10%	15%	10%

Table 3: Reflection detection accuracy for the three boundary cases at varying relative amplitudes.

Relative dB	Left	Right
-6dB	100%	95%
-9dB	100%	90%
-12dB	90%	60%
-15dB	75%	20%
-18dB	70%	15%
-21dB	25%	10%
-24dB	15%	0%
-27dB	15%	10%

Table 4: Left-Right reflection detection accuracy for both (6ms & 20ms) right boundary reflections.

4.2. Anechoic front-back localization

To test if a listener's front-back localization accuracy changes when early reflections are present, a baseline anechoic localization ability was needed. Table 5 shows the percentages of correct front-back localization for the three groups of subjects in the four eye-level quadrants. For each of the participant groups, localization ability in the front-left and rear-right quadrants reported better localization accuracy than the opposing quadrants.

In general, a fairly random distribution around the mean localization ability was observed as seen in Figure 2, however, Group A subjects appear to have a slightly higher than average front-back accuracy, whereas Group B showed a slightly lower accuracy. Due to a reduced ability to obtain audible information, this result would make sense that listeners with higher detection thresholds have more difficulty localizing than those with normal hearing. However, Subject 11 (included in Group B) was above the average localization ability in two of the quadrants and correctly localized all trials in the front-left quadrant.

	Left			Right		
	All Subj.	Group A	Group B	All Subj.	Group A	Group B
Front	70%	72%	65%	62%	68%	48%
Rear	60%	64%	52%	74%	73%	77%

Table 5: Front-back localization accuracy in anechoic conditions for each quadrant and participant group.

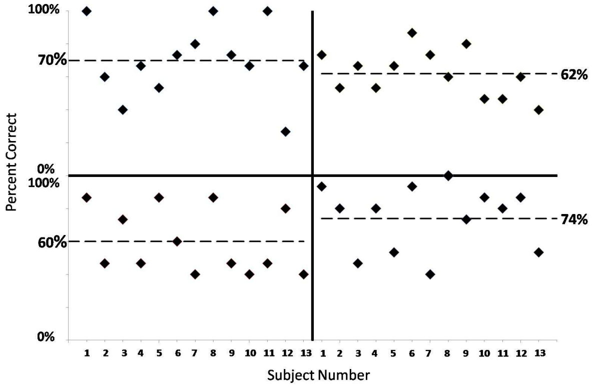


Figure 2: Individual results for localization accuracy in anechoic conditions. Dotted lines designate the average response across all subjects for each quadrant.

4.3. Non-anechoic front-back localization

It is conceived that the addition of an early reflection could provide additional information for the localization of a stationary sound stimuli presented virtually via generalized HRTFs. The final experiment served as the primary motivation behind this entire investigation. Figure 3 shows the difference in localization accuracy between the anechoic and non-anechoic conditions for each individual. A positive percent difference corresponds to improved localization accuracy for the non-anechoic conditions. Again, the localization accuracy is fairly dependent upon the individual, however, most participants improved front-back localization in the left hemisphere when early reflections were presented.

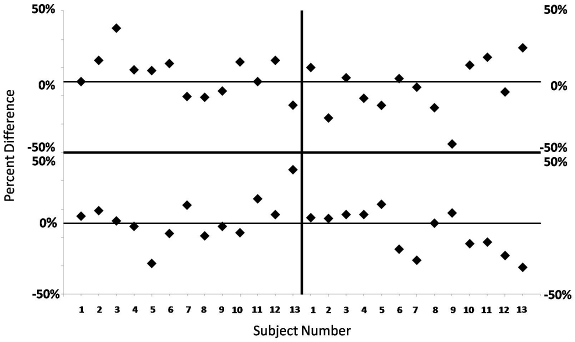


Figure 3: Difference in localization accuracy between anechoic and non-anechoic conditions for each subject.

In addition to comparing localization accuracy purely upon the quadrant responses, the different boundaries and relative amplitudes were investigated. Table 6 shows the difference in localization accuracy from the anechoic baseline provided in the previous section. Both of the wall reflection cases show a general increase in accuracy for the front-left quadrant across each of the participant groups. A more notable result can be seen for stimuli presented in both the front and rear on the right side for the 6ms and 20ms wall reflection cases. With the exception of Group B in the front-right quadrant, localization accuracy decreases on right hemisphere stimuli for all of the groups.

Analyses of the relative amplitudes versus localization accuracy are shown in Figure 4. Although, in general, front-back localization has shown to be higher in the left-front quadrant for these experiments, the accuracy is even higher for all participant groups in the left hemisphere when a 6ms right-wall reflection is added at 6dB below the direct signal. The geometry for this reflection corresponds to a realistic, yet, highly reflective surface (ρ=0.85). It is also seen that the presentation of a left hemisphere direct signal with a 20ms reflection at -12dB & -15db, ρ=1.11 & ρ=0.79 respectively, yield a consistently higher accuracy across all participant groups. Though the geometry for the virtual stimuli would not quite create a reflection only 12dB down from the direct signal, it does appear that a more exaggerated reflection amplitude could potentially help disambiguate front-back localization when the direct signal is presented on the opposite lateral side of a nearby boundary.

	Left			Right		
	All Subj.	Group A	Group B	All Subj.	Group A	Group B
Front	-8%	-12%	0%	4%	-1%	-17%
Rear	2%	3%	2%	0%	11%	27%

(a)

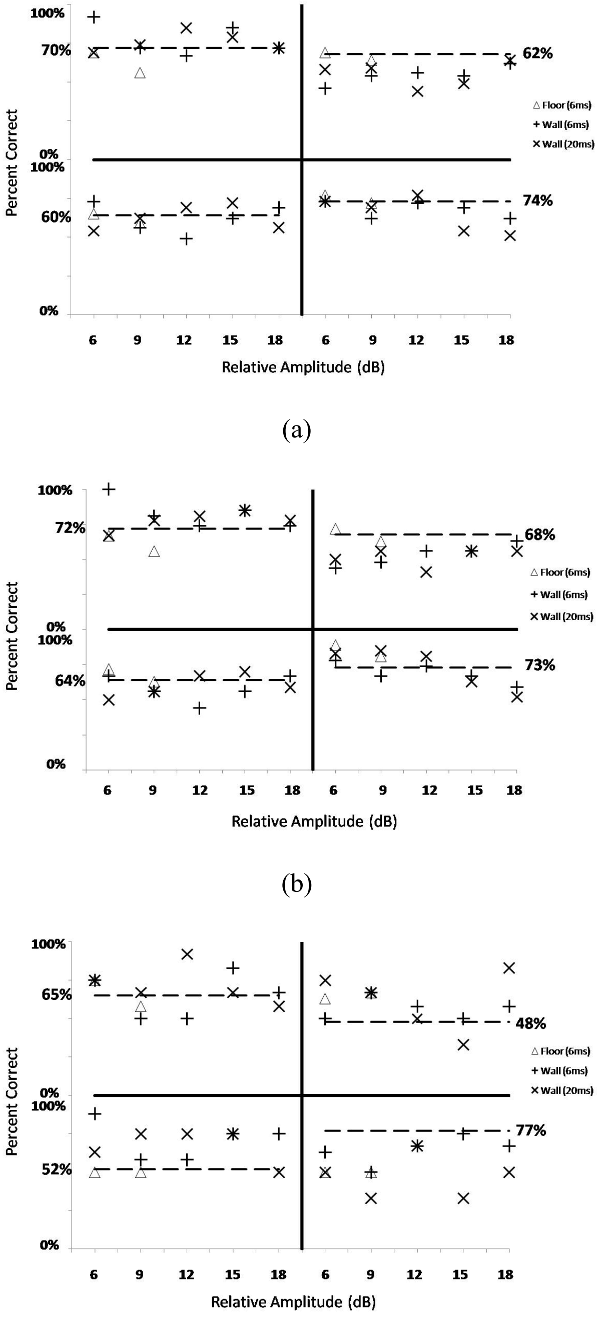
	Left			Right		
	All Subj.	Group A	Group B	All Subj.	Group A	Group B
Front	6%	10%	-1%	-7%	-14%	9%
Rear	1%	-7%	18%	-7%	-5%	-13%

(b)

	Left			Right		
	All Subj.	Group A	Group B	All Subj.	Group A	Group B
Front	6%	7%	6%	-8%	-16%	13%
Rear	3%	-3%	16%	-10%	-1%	-31%

(c)

Table 6: Difference in localization accuracy from anechoic baseline for (a) floor 6ms, (b) wall 6ms, and (c) wall 20ms boundaries.



(a)

(b)

(c)

Figure 4: Comparison of localization accuracy from mean values for different reflection amplitudes in (a) All Subj., (b) Group A, and (c) Group B.

5. CONCLUSION

This investigation tested the effect of a single reflection on listener's ability to localize sound as being in either the front or rear when using generalized HRTFs. Though results varied across all participants, certain trends did appear. Subjects did show a heightened localization ability in the front-left and rear-right quadrants for both anechoic and non-anechoic cases. This is potentially an intrinsic nuance of the HRTF dataset used in this experiment; it would be interesting to try various HRTFs to determine if this actually was the case. Another possible contribution to this variation in quadrant results could be due to the spectral modulation of the speech samples. Variations in word inflection could excite different aspects of the HRTF thus assisting or inhibiting localization performance.

A more important finding was reduced front-back localization ability for stimuli virtually placed on the same side as the reflective boundary. This density of spatial information on one side of the head could be reasoned to produce spectral "smearing", thus resulting in an uncertain source location. In contrast, increased localization ability was found for some cases when stimuli were virtually placed on the opposite side of the reflective boundary. These improvements were shown for both wall-boundary reflection delays, however, this only occurred when the boundary was highly reflective and in one case exaggerated beyond real-world environments. This would follow in agreement with [15, 16] that reflections help to immerse listeners in a virtual environment, thus, creating a better sense of location within the environment.

Variations in hearing levels were also examined for differences in localization abilities. Poorer hearing, though on average decreased front-back localization ability, did not completely inhibit front-back accuracy. Localization remained highly dependent upon the individual. Though almost all subjects reported that they felt an improved ability to localize the stimuli as more trials were presented, no significant trend of improvement was actually found.

In general, first order reflections do not necessarily disambiguate front-back source locations for all cases when using generalized HRTFs, however, results in this investigation show that reflections can play a role in localization for unique reflection geometries.

6. ACKNOWLEDGEMENTS

This work was supported by Montana State University Electrical and Computer Engineering graduate student research support. We would also like to thank Headroom, Dr. Alan Leech, Dr. Kevin Repasky, Diego DeLeon, and all participants for their assistance with this research.

7. REFERENCES

[1] Nicholas J. Wade and Diana Deutsch, "Binaural hearing – Before and after the stethophone," *Acoustics Today*, Vol. 4, No. 3, pp. 16-27, 2008.

[2] J. W. Strutt (Lord Rayleigh), "On our perception of sound direction," Philosophical Magazine 13: 1907.

[3] W. Mills, "Auditory localization," *Foundations of Modern Auditory Theory*, edited by J.V. Tobias, Academic, New York, 1972.

[4] Elizabeth M. Wenzel, Marianne Arruda, Doris J. Kistler, and Frederic L. Wightman, "Localization using nonindividualized head-related transfer functions," *J. Acoust. Soc. Am.*, Vol. 94, No. 1, pp. 111-123, July 1993.

[5] Aleksander Väljamäe, Pontus Larsson, Daniel Västfjäll, and Mendel Kleiner, "Auditory presence, individualized head-related transfer functions, and Illusory Ego-Motion in Virtual Environments," Proceedings of Presence 2004, 7th International Workshop on Presence, pp. 141-147, Valencia, Spain, Oct. 2004.

[6] Frederic L. Wightman and Doris J. Kistler, "Headphone simulation of free-field listening. II: Psychophysical validation," *J. Acoust. Soc. Am.*, Vol. 85, No. 2, pp. 868-878, Feb. 1989.

[7] Patrick Satarzadeh, V. Ralph Algazi, and Richard O. Duda, "Physical and filter pinna models based on anthropometry," 122^{nd} Convention of the Audio Engineering Society, Paper No. 7098, Vienna, Austria, May 2007.

[8] Dimitry N. Zotkin, Jane Hwang, Ramani Duraiswami, and Larry S. Davis, "HRTF personalization using anthropometric measurements," Proc. WASPAA '03 (2003 IEEE

ASSP Workshop on Applications of Signal Processing to Audio and Acoustics), pp. 157-160, New Paltz, NY, 2003.

[9] Navarun Gupta, Armando Barreto, and Carlos Ordonez, "Improving sound spatialization by modifying head-related transfer functions to emulate protruding pinnae," 2002 Proc. IEEE SoutheastCon, pp. 446-450, Columbia, SC, 2002.

[10] Niklas Röber, Sven Andres, and Maic Masuch, "HRTF simulations through acoustic raytracing," Technischer Report Nr.4, Fakultät für Informatik, Otto-von-Guericke Universität Magdeburg, 2006.

[11] Frederic L. Wightman and Doris J. Kistler, "Resolution of front-back ambiguity in spatial hearing by listener and source movement," *J. Acoust. Soc. Am.*, Vol. 105, No. 5, pp. 2841-2853, May 1999.

[12] Elizabeth M. Wenzel, "The relative contribution of interaural time and magnitude cues to dynamic sound localization," Proc. WASPAA '95 (1995 IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics), pp. 80-83, New Paltz, NY, 1995.

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

[14] Durand R. Begault, "Virtual Acoustics, Aeronautics, and Communications," *J. Aud. Eng Soc.,* Vol. 46, No. 6, pp. 520-530, 1998.

[15] Dimitry N. Zotkin, Ramani Duraiswami, and Larry S. Davis, "Creation of virtual auditory spaces," Proc. ICASSP '02 (2002 IEEE International Conference on Acoustics, Speech, and Signal Processing), pp. 2113-2116, Orlando, USA, 2002.

[16] Renato S. Pellegrini, "Quality assessment of auditory virtual environments," Proc. ICAD '01 (2001 International Conference on Auditory Display), pp. 161-168, Espoo, Finland, 2001.

[17] Barbara G. Shinn-Cunningham, "The perceptual consequences of creating a realistic reverberant 3-D audio display," Proceedings of the International Congress on Acoustics, Kyoto, Japan, April 2004.

[18] Pavel Zahorik, Doris J. Kistler, and Frederic L. Wightman, "Sound localization in varying virtual acoustic environments," Proc. ICAD '94 (1994 International Conference on Auditory Display), pp. 179-186, Santa Fe, USA, 1994.

[19] Nail A. Gumerov and Ramani Duraiswami, "Modeling the effect of a nearby boundary on the HRTF," Proc. ICASSP '01 (2001 IEEE International Conference on Acoustics, Speech, and Signal Processing), Vol. 5, pp. 3337-3340, Salt Lake City, USA, 2001.

[20] Brad Rakerd and W.M. Hartmann. "Localization of sound in rooms, II: The effects of a single reflecting surface," *J. Acoust. Soc. Am.*, Vol. 78, No. 2, pp. 524-533, Aug. 1985.

[21] Durand R. Begault, Elizabeth M. Wenzel, and Mark R. 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. Aud. Eng Soc.,* Vol. 49, No. 10, pp. 904-916, 2001.

[22] Dimitry N. Zotkin, Ramani Duraiswami, and Larry S. Davis, "Rendering localized spatial audio in a virtual auditory space," IEEE Transactions on Multimedia, Vol. 6, No. 4, Aug. 2004.

[23] Paul L. Michael and Gordon R. Bienvenue, "Calibration data for a circumaural headset designed for hearing testing," *J. Acoust. Soc. Am.*, Vol. 60, No. 4, pp. 944-950, Oct. 1976.

[24] Arnold F. Heidbreder, "Calibration of earphones on a flat-plate coupler: Exploring two setup procedures," *J. Acoust. Soc. Am.*, Vol. 70, No. 1, pp. 228-230, July 1981.

[25] Shiro Nakamura and Yoshio Katano, "Relation between temporary threshold shift and duration of exposure," International Journal of Audiology, Vol. 5, No. 2, Pages 196-199, 1966.

[26] M. Grassi and A. Soranzo, "MLP: a MATLAB toolbox for rapid and reliable auditory threshold

estimations," *Behavior Research Methods*, Vol. 41, No. 1, pp. 20-28, 2009.

[27] D. M. Green, "A maximum-likelihood method for estimating thresholds in a yes-no task," *J. Acoust. Soc. Am.*, Vol. 93, No. 4, pp. 2096-2105, April 1993.

[28] V. R. Algazi, R. O. Duda, and D. M. Thompson, "The CIPIC HRTF database," Proc. WASPAA '01 (2001 IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics), pp. 99-102, New Paltz, NY, 2001.

[29] Jont B. Allen and David A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Am.*, Vol. 65, No. 4, pp. 943-950, April 1979.

[30] Durand R. Begault, Bryan U. McClain, and Mark R. Anderson, "Early reflection thresholds for anechoic and reverberant stimuli within a 3-D sound display," Proceedings of the 18th International Congress on Acoustics, Kyoto, JP, 2004.