Sound Localization in Varying Virtual Acoustic Environments

Pavel Zahorik, Doris J. Kistler, and Frederic L. Wightman
Waisman Center
University of Wisconsin—Madison
1500 Highland Avenue
Madison, WI 53705
Email:zahorik@waisman.wisc.edu

Abstract

Localization performance was examined in three types of headphone-presented virtual acoustic environments: an anechoic virtual environment, an echoic virtual environment, and an echoic virtual environment for which the directional information conveyed by the reflections was randomized. Virtual acoustic environments were generated utilizing individualized head-related transfer functions and a three-dimensional image model of rectangular room acoustics—a medium sized rectangular room ($8m \times 8m \times 3m$) with moderately reflective boundaries (absorption coefficient, $\alpha = 30.75$) being modeled. Five listeners reported the apparent position of a wide spatial range of virtual sound sources. Judgments of apparent source position were unaffected by acoustic environment manipulation even though sound sources presented in each of the three environments were informally discriminable. These findings question the necessity of spatialized room reflection information for high localization performance in virtual auditory displays, as well as provide further evidence for the robustness of precedence phenomena.

1 Introduction

In standard instantiations of headphone-delivered three-dimensional auditory displays, errors in sound source localization may be roughly assigned to one of three categories:

1. small judgment variation, or "blur," of apparent sound source position about target virtual source position

2. reversal of position judgment about the coronal plane—so called "front to back" or "back to front" reversals

3. judgment errors in degree of cranial externalization

At present, precise explanation for the existence of these localization error types in three-dimensional auditory displays is unavailable. However, it seems clear that such auditory displays are in a number of senses not faithful to the reproduction of auditory stimulation occurring in natural, everyday situations. It therefore appears conceivable that localization errors in 3-D auditory displays are in some fashion a result of non-natural simulation.

One way in which standard 3-D auditory displays may be regarded as non-natural is the lack of reflection and reverberation simulation. Several studies have shown the inclusion of reflection information representative of indoor room environments affects localization errors. Specifically, Begault reports that for listeners localizing sounds in such virtual echoic environments constructed with nonindividualized head-related transfer functions (HRTFs), egocentric distance (or externalization) judgments increased by approximately a factor of three relative to localizing in virtual anechoic environments [1]. Durlach and his colleagues [2] concur with Begault's findings, further claiming that it is most likely a decrease in direct-to-reverberant energy ratio, thought to be an important cue for the perception of auditory distance [3], that accounts for the increase in cranial externalization of auditory images presented with synthetic reflections. Interestingly, these benefits in externalization as a result of reflection simulation have been reported to be at the expense of increases in reversal errors [1]. It should also be noted that these synthetic reflection findings appear to challenge the classical notions of "precedence" as a purely echo suppressive mechanism [4].

Virtual simulation of echoic space involves three principle parameters in addition to those utilized by standard headphone 3-D auditory displays: reflection time delay, reflection attenuation (potentially frequency dependent), and reflection spatial position.

Correct simulation of reflection spatial position is perhaps the most computationally demanding parameter. Hence, the greatest gains in implementational simplicity of virtual echoic space simulations would be realized by constraining this parameter in some sense. As a result of informal listening tests with virtual echoic environments constructed from nonindividualized HRTFs, Begault reports no difference in apparent source position between simulations where reflection spatial information is properly represented and simulations where reflection spatial information is chosen randomly [5]. Such results suggest that it in fact may not be necessary to properly simulate reflection spatial information in virtual echoic displays.

It is the goal of this study to further examine localization performance in virtual echoic environments with two principal additions. First, displays will be constructed with individualized HRTFs. Second, the echoic environment will be manipulated by varying the spatial information contained in the reflections. The latter addition will seek to formally determine the necessity of spatially correct reflection information for successful localization performance.

2 Method

2.1 Listeners

Three male and two female paid volunteers served as listeners. All had audiometrically verified normal hearing, as well as previous experience with localization judgment tasks.

2.2 Stimuli

Three classes of stimuli were used: virtual anechoic stimuli, virtual echoic stimuli, and virtual perturbed-echoic stimuli. The virtual anechoic stimuli were produced by filtering 250ms gaussian noise-bursts (chosen at random from a sample of 50 pre-computed noise bursts, then bandpass filtered from 200–14000 kHz and windowed with a 10ms ramp up/down raised cosine function) with left/right pairs of HRTFs corresponding to an array of 144 source positions. HRTFs were derived from individual listener probe-tube microphone measurements taken from 450 source positions in anechoic space (Wightman and Kistler provide a detailed description of this HRTF measurement procedure [6]).

Virtual echoic stimuli were constructed using a three dimensional image-source room acoustics model [7]. Such a model provides information as to the spatial position of each reflection (i.e., the incident angle of the reflection on the listener), as well as time delay and attenuation information. In this study, an $8m$ x $8m$ x $3m$ rectangularly shaped room with a centrally located listener was modeled. Each of the six reflecting surfaces were defined to have uniform frequency 0.75 absorption coefficients, α, which were independent of incidence angle. Loss of intensity due to distance of sound travel obeys the inverse square law in the acoustic free-field and was computed as such in this setting. Therefore, total attenuation of each reflection is a function of the number of reflector contacts and the total distance of sound wave travel. It should be noted that a variety of other room acoustic models exist. The image-model was chosen in this rectangular room setting for its simplicity and computational efficiency [8].

Binaural room impulse responses (BRIRs) were constructed from the information provided by the image-model. Specifically, right/left pairs of HRTFs (the time-domain equivalents thereof) corresponding to the appropriate spatial positions of the direct sound source path and each of its reflections were individually scaled and time-shifted the appropriate amounts, and then summed together. An interpolation algorithm was implemented when the spatial positions of reflections were disparate from measured HRTF positions. The resulting BRIRs were then convolved with the same type of noise-burst as described previously. In this study, the BRIRs were limited to include only the first 20 reflections occurring in time after the direct source path.

The third type of stimuli, the virtual perturbed-echoic stimuli, were constructed in a fashion analogous to the construction of the virtual echoic stimuli, but with one crucial difference. In this case the spatial positions of the reflections were chosen at random, rather than as prescribed by the image-model. All other stimulus parameters (including attenuation values and time delay) remained the same as for the virtual echoic stimuli.

All stimuli were pre-computed and stored for subsequent experimental presentation over headphones.

2.3 Procedure

Three virtual acoustic conditions were presented: A baseline condition with the virtual anechoic stimuli described above, a "correct" reflection condition with the virtual echoic stimuli, and a random reflection condition with the virtual perturbed-echoic stimuli. Listeners verbally reported apparent sound source position in terms of azimuth, elevation and distance via a polar coordinate system. The three virtual acoustic conditions were presented in successive blocks of the same 144 virtual source positions. Order of presentation within a block was randomized. The 144 source positions were chosen at random from the possible 450 positions at which HRTF measurements were performed. Four replications in each of the virtual acoustic conditions yielded 576 judgments per condition for each listener.

3 Results

The three virtual acoustic environments examined here were found to have little effect on localization performance. Figures 1–3 display localization data from three representative listeners. Virtual source position is plotted as a function of apparent source position (for each of the experimental conditions) in three different transformed coordinated systems: right/left, front/back, up/down. The right/left dimension is determined by collapsing sources and judgments across both the front/back and up/down dimensions, such that a $-90°$ angle is directly to the listener's left, a $0°$ angle directly in front of the listener, and a $90°$ angle to the listener's right. Front/back and up/down dimensions are determined analogously, by collapsing across the remaining two dimensions.

Symbol shading is proportional to the number of judgments at a given position. Visual examination of Figures 1–3 suggests the existence of little within-subject difference across experimental

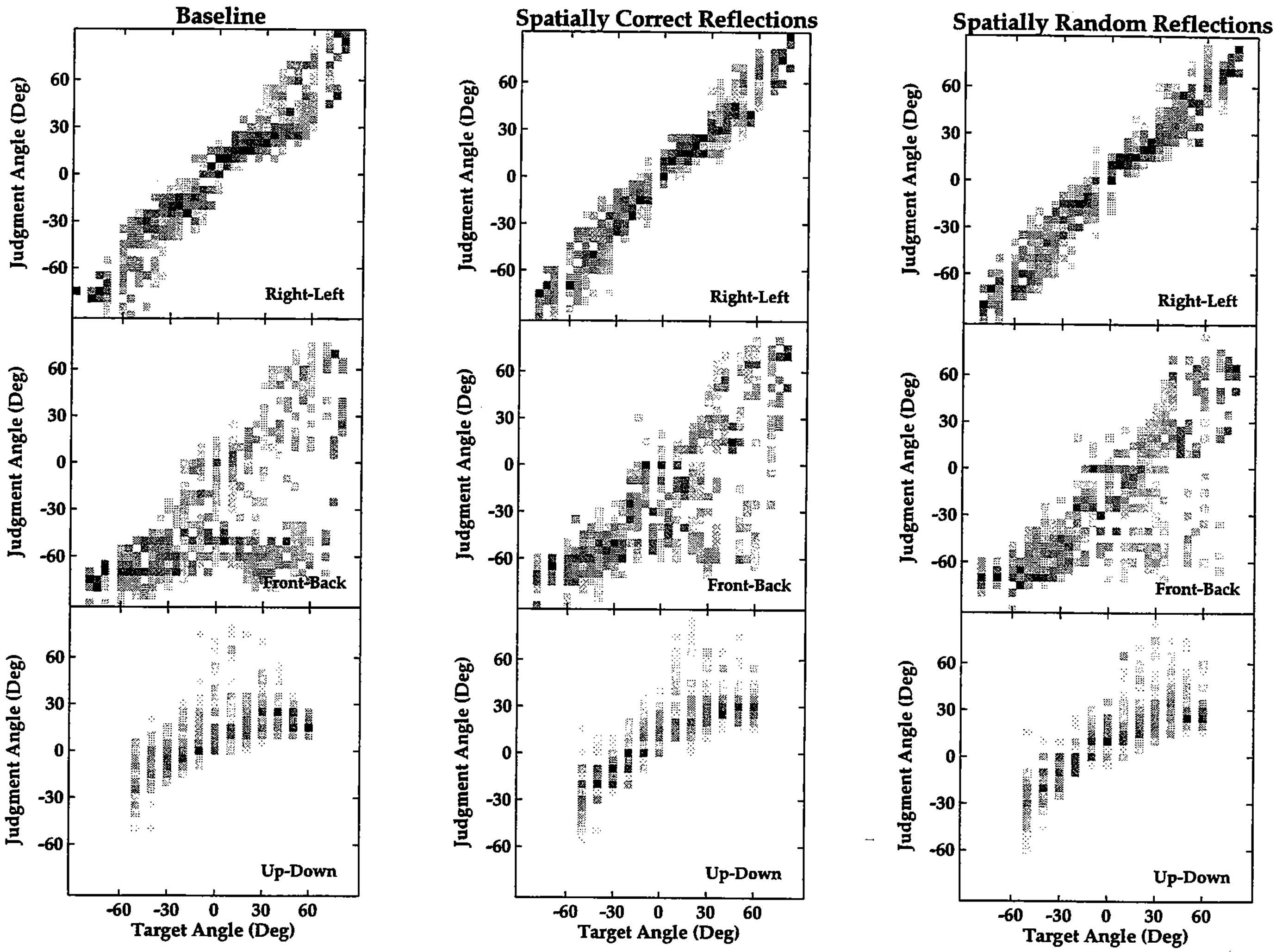

Figure 1: Localization data for listener SMQ.

Listener	Baseline	Spatially Correct Reflections	Spatially Random Reflections
SMQ	0.2083	0.1424	0.1441
SNF	0.1892	0.1563	0.1319
SNJ	0.0677	0.0434	0.0522
SNX	0.1267	0.1094	0.0922
SNY	0.0838	0.0991	0.1270

Table 1: Reversal porportions.

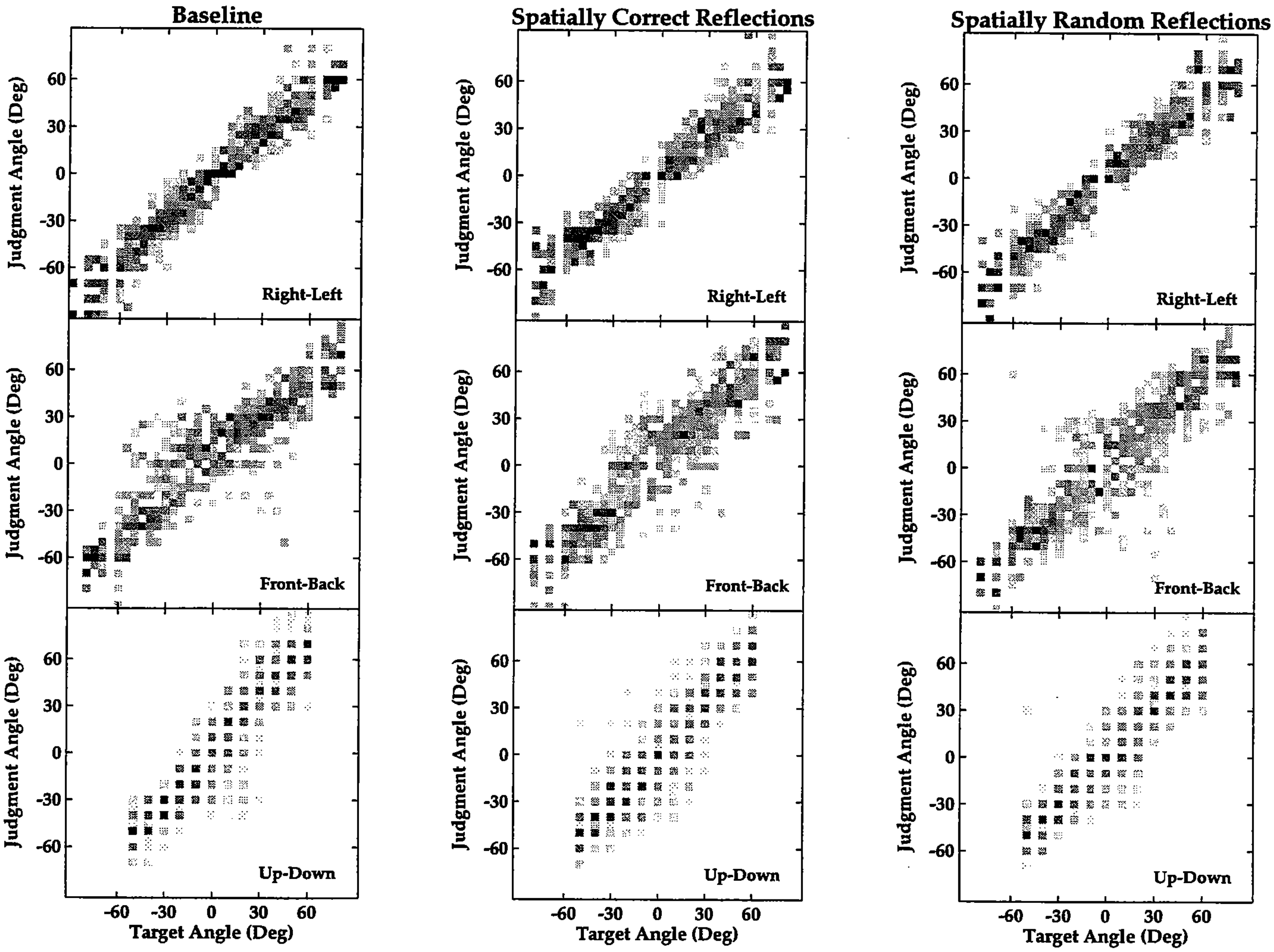

Figure 2: Localization data for listener SNJ.

Listener	Baseline	Spatially Correct Reflections	Spatially Random Reflections
SMQ	5.20	5.03	5.03
SNF	3.58	3.73	3.72
SNJ	3.06	3.00	3.00
SNX	6.08	6.08	6.03
SNY	2.65	2.90	2.90

Table 2: Distance judgments (ft).

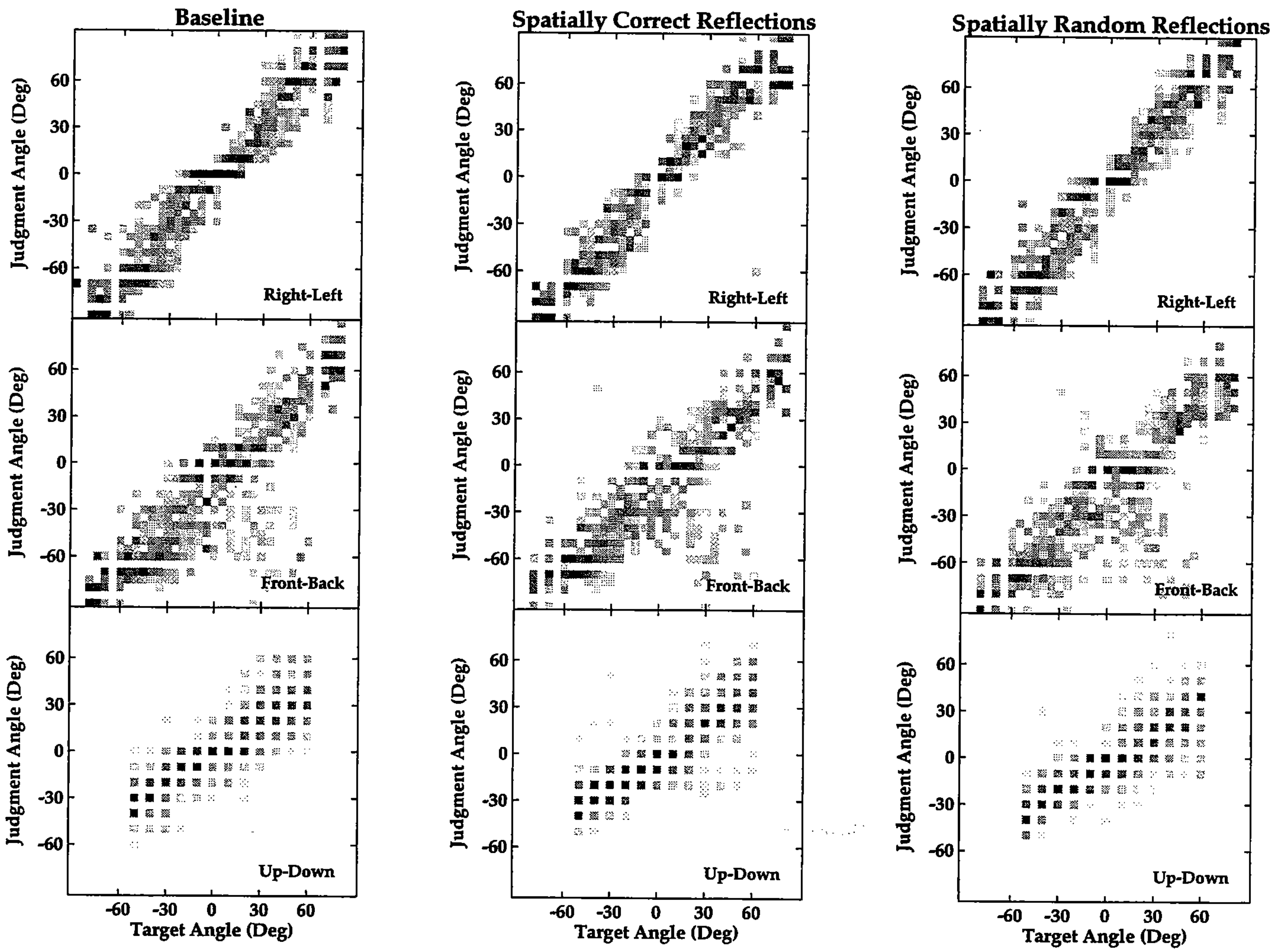

Figure 3: Localization data for listener SNY.

condition. Front-Back and Back-Front reversals may be seen on Figures 1–3 as judgments lying on or near the negative diagonal (i.e., $y = 3 - x$) of the Front-Back dimension panel.

Table 1 displays combined Front-Back and Back-Front reversal rates for each listener in each acoustic condition. A within-subjects ANOVA on the arcsine transformed reversal rates (a recommended transformation for small proportional scores [9]) revealed no significant differences across experimental conditions, $F(2,4) = 1.95$, $p = .204$.

Table 2 shows listener's distance judgments for each condition. Results of a within-subject ANOVA suggest that distance judgments were also unaffected by experimental condition, $F(2,4) = 0.17$, $p = .844$.

4 Conclusion

These null results are perhaps somewhat surprising, given both the findings of Begault and others, and the fact that the stimuli presented in these three virtual acoustic conditions, upon subjective evaluation, were markedly different. It is not inconceivable to attribute these differences to, at least in part, the use of individualized HRTFs, since it has been shown that the use of nonindividualized HRTFs (such as Begault [1] and Wenzel et al. cite5) suffers from both a degradation in externalization and an increase in reversal error rates [10]. Therefore, it is quite possible that the lack of increase in distance judgments, as well as reversal error rates, for echoic conditions as compared to anechoic conditions is a result of the use of individualized HRTFs. It should be noted that the constancy of reversal error rates across experimental conditions is in fact an encouraging result when compared to the results of previous studies.

Regardless of cause, a clear difference in results between this study and previous studies exists. Localization performance, in terms of apparent sound source position, has been shown to be quite robust with respect to the varied virtual acoustic environments examined. Therefore, if particular applications of 3-D auditory displays are concerned only with localization performance, and individualized HRTFs are available, two conclusions exist: Reflection spatial information need not necessarily be realistic, and further, such reflection information is perhaps wholly unnecessary.

Acknowledgments

Financial support provided in part by grants from NIH, NASA, and ONR.

References

[1] Begault, D. R. "Perceptual Effects of Synthetic Reverberation on Three-Dimensional Audio Systems." *J. Audio Eng. Soc.* **40** (1992): 895–904.

[2] Durlach, N. I., W. S. Woods, A. Kulkarni, H. S. Colburn, and E. M. Wenzel. "On the Externalization of Auditory Images." *Presence* **1** (1992): 251–257.

[3] Mershon, D. H. and L. E. King. "Intensity and Reverberation as Factors in the Auditory Perception of Egocentric Distance." *Perception & Psychophysics* **18** (1975): 409–415.

[4] Blauert, J. *Spatial Hearing.* Cambridge, MA: MIT Press, 1983.

[5] Begault, D. R. "Binaural Auralization and Perceptual Veridicality." *J. Audio Eng. Soc.* (1992): Preprint 3421 (M-3).

[6] Wightman, F. L. and D. J. Kistler. "Headphone Simulation of Free-Field Listening. I: Stimulus Synthesis." *J. Acoust. Soc. Am.* **85** (1989): 858–867.

[7] Allen, J. B. and D. A. Berkley. "Image Method for Efficiently Simulating Small-Room Acoustics." *J. Acoust. Soc. Am.* **65** (1979): 943–950.

[8] Vorlander, M. "Simulation of the Transient and Steady-State Sound Propagation in Rooms Using a New Combined Ray- Tracing/Image-Source Algorithm." *J. Acoust. Soc. Am.* **86** (1989): 1720–178.

[9] Kirk, R. E. *Experimental Design*, 2nd edition. Monterey, CA: Brooks/Cole, 1982.

[10] Wenzel, E. M., M. Arruda, D. J. Kistler, and F. L. Wightman. "Localization Using Nonindividualized Head-Related Transfer Functions". *J. Acoust. Soc. Am.* **94** (1993): 111–123.