DECOUPLED LOUDNESS AND RANGE CONTROL FOR A SOURCE LOCATED WITHIN A SMALL VIRTUAL ACOUSTIC ENVIRONMENT

William L. Martens
McGill University, Faculty of Music
Montreal, QC, H3A 1B4 Canada
wlm@music.mcgill.ca

ABSTRACT

For headphone-based spatial auditory display systems, binaural synthesis of sound localization cues typically use source reproduction level as the primary control for source range. This approach can be quite effective when indirect sound is simulated in order to externalize virtual sources within a small virtual acoustic environment. A computationally efficient simulation solution is described here that does not rely solely upon the sound reproduction level of the source to control source range (i.e., perceived egocentric distance), and provides an extremely economical synthesis of the indirect sound component that is effective in creating externalized spatial auditory images. The performance of the solution has been psychophysically validated using indirect scaling methods that required experimental listeners to compare two displayed sound stimuli and report which of the two was the louder or the closer. In particular, it was shown that the simulation allows for decoupled loudness and range control for a source located near the listener's head, so that equally loud sources can be positioned at varying source range. Likewise, within certain limits, source loudness may be varied while holding source range constant. This performance feature has benefits for auditory display applications for which selective attention should be supported for a spatially distributed set of virtual sound sources.

1. INTRODUCTION

The spatial sound processing technology that is commonly used in creating immersive virtual acoustic environments has matured over the past 20 years, but there is still a real need for computationally efficient implementations of this technology that are carefully designed to provide adequate results for specific applications. For headphone-based spatial auditory display systems, binaural synthesis using generalized or customized Head-Related Impulse Responses (HRIRs) have proven adequate for controlling the apparent direction of virtual sources, even without the inclusion of simulated indirect sound (i.e., early discrete reflections and later diffuse reverberation). The author has recently reviewed the success of such systems in positioning such dry virtual sources [1], and no systematic review is included in this paper. Suffice it to say that perceptual evaluation studies have confirmed that dry, HRIR-based processing provides useful directional cues, but is not typically employed in such a way as to provide useful cues to the egocentric distance of a virtual source. This is despite the fact that simple manipulation of interaural level difference (ILD) can modulate perceived source range even when no indirect sound simulation provides a virtual acoustic environment within which range might be disambiguated via the powerful cue available in the ratio of direct to indirect sound levels [2].

For the research reported in this paper, a computationally efficient reverberant sound field simulation is added to an ILD-based range cueing solution to support multiple source range cues that extend beyond those typical of less-expensive solutions, which most often use source reproduction level as the primary control for source range. Of course, the range control attempted here was quite limited, since the targeted applications are those that need to position a virtual source quite near the listener's head; in fact, within arm's reach (see [3] for a discussion of how the interaural time and level differences present at such close range create "Tori of Confusion" with regard to 3D sound localization). Furthermore, the indirect sound simulation was designed in order to externalize virtual sources within a small virtual acoustic environment; in fact, just large enough to allow the source to be positioned within the very personal space, the walls of the simulated enclosure also nearly within arm's reach (such as those of an idealized aircraft cockpit or automobile environment). This focus was chosen for two reasons. First, it is frequently reported in informal binaural listening tests that virtual sources located within this personal space have a tendency to be easily noticed, less likely to be ignored. Hence, for many spatial auditory display applications, this is where important sound messages will be positioned. Second, it is within this personal space that spatial sound processing technology is commonly available for simple but effective manipulation of source range, based upon well documented acoustical phenomena, such as the growth in the ILD exhibited in HRTFs measured for sources approaching the listener's head at close range [4], which is also well understood theoretically [5].

Such computationally efficient reverberant sound field simulations have often been presented, but many suffer from the attempt to cover too wide an area of application, and therefore perform poorly in many of the possible uses. One such attempt was reported by Robinson, et al [6]. They report poor results with respect to control over source range. In the current research, the performance of an efficient headphone-based spatial auditory display system was evaluated only within a narrowly constrained region of space, source localization within which has been shown to be successful [7]. This paper begins with an overview of the spatial sound processing solution employed in the research reported here, along with a presentation of the rationale upon which simplifying choices were made. Then the paper describes the results of experimental listening tests designed to determine the success of the binaurally synthesized sound localization cues used to provide independent control over source range and source loudness. In contrast to headphone-based spatial auditory display systems that use source reproduction level as the primary control over source range, the performance of the current system is expected to allow equally loud sources to be positioned with varying source range. Also, conversely, source loudness manipulations should be possible while holding source range constant. The performance of the solution in this regard was psychophysically examined using indirect scaling methods to determine the validity of the range control methods.

2. METHODS

This section describes both the stimulus generation methods and the experimental methods used in the listening tests. First, an overview of the signal processing underlying the employed spatial auditory display system is presented.

2.1 Stimulus Generation: Signal Processing Overview

The binaural signal processing required to move a single virtual sound source through a virtual acoustic environment of fixed geometry is described here for a single channel of audio input, though multiple channels of such software-based processing can be implemented with significant savings since there is no great loss in effectiveness if a single reverberant "tank" [8] is used for all sources needing indirect sound simulation. In the case that multiple channels of audio inputs are to be processed in realtime, it would typically be decided to process each separately for both direction and range cueing, while sending the unprocessed inputs to a single separate reverberation generation system. For sake of clarity, the processing of the dry sources is illustrated in a first diagram (Figure 1), and the reverberation generation system is shown in a second diagram (Figure 2).

Figure 1 gives a block diagram illustrating the binaural signal processing for a dry virtual sound source, and includes a send to and return from the reverb generation processes that is defined in a separate diagram. Note that this diagram shows the delay buffer required for implementation of realtime update of both interaural time difference (ITD) and Reverb Pre-Delay. The three taps into the Audio Input Delay Buffer are smoothly interpolated at runtime to implement Doppler shift for both the dry source and the first simulated reflection, providing powerful cues to changing source range. Besides the delay-based distance cues, the proposed solution separately manipulates gain on both ipsilateral and contralateral signals. This allows control of source loudness and source range to be decoupled, especially for nearby sources (within arms reach), where increasing contralateral attenuation makes sources sound closer to the listener [1].

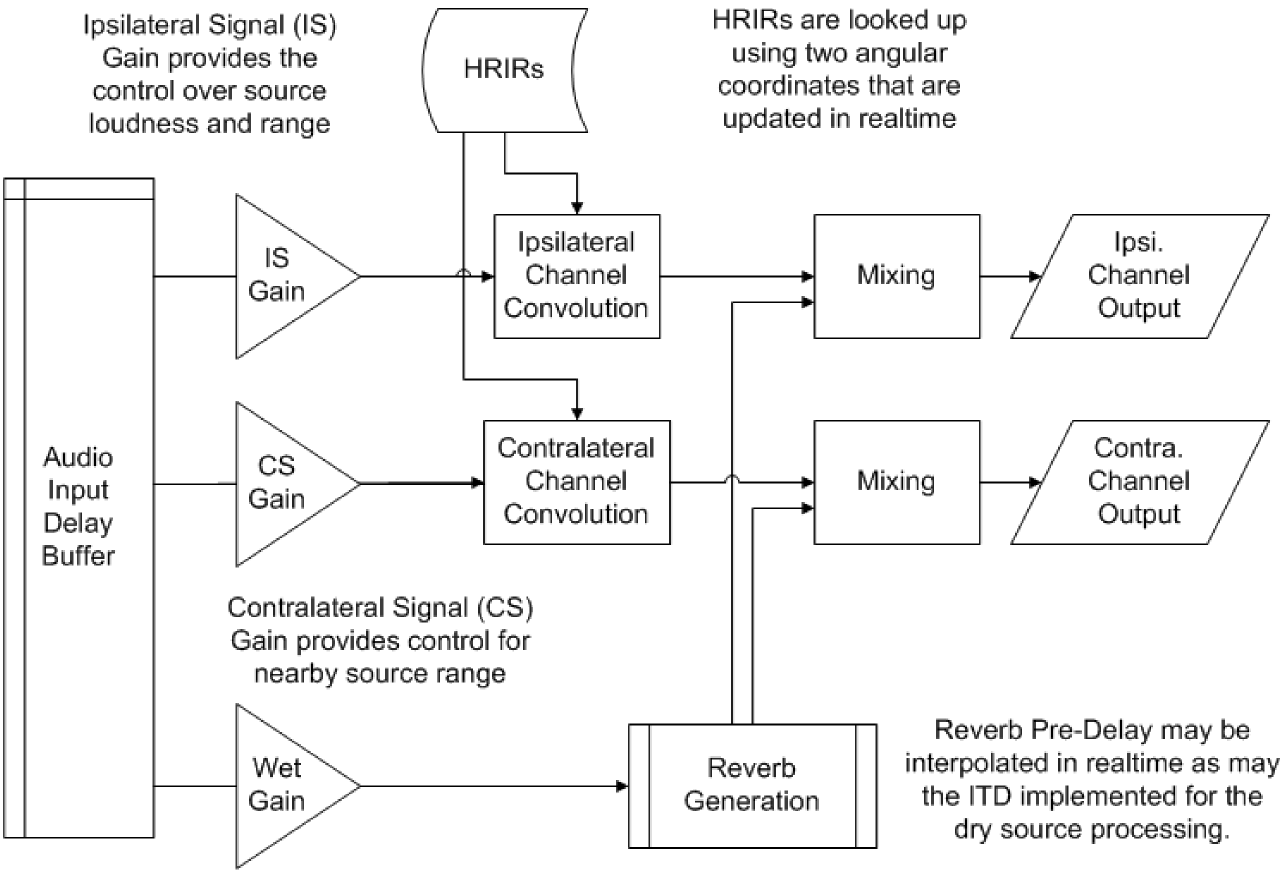

Figure 1. *Block diagram providing details regarding the employed binaural signal processing of a dry source that includes parameters for control of distance. Note that the Reverb Generation is described in a separate diagram.*

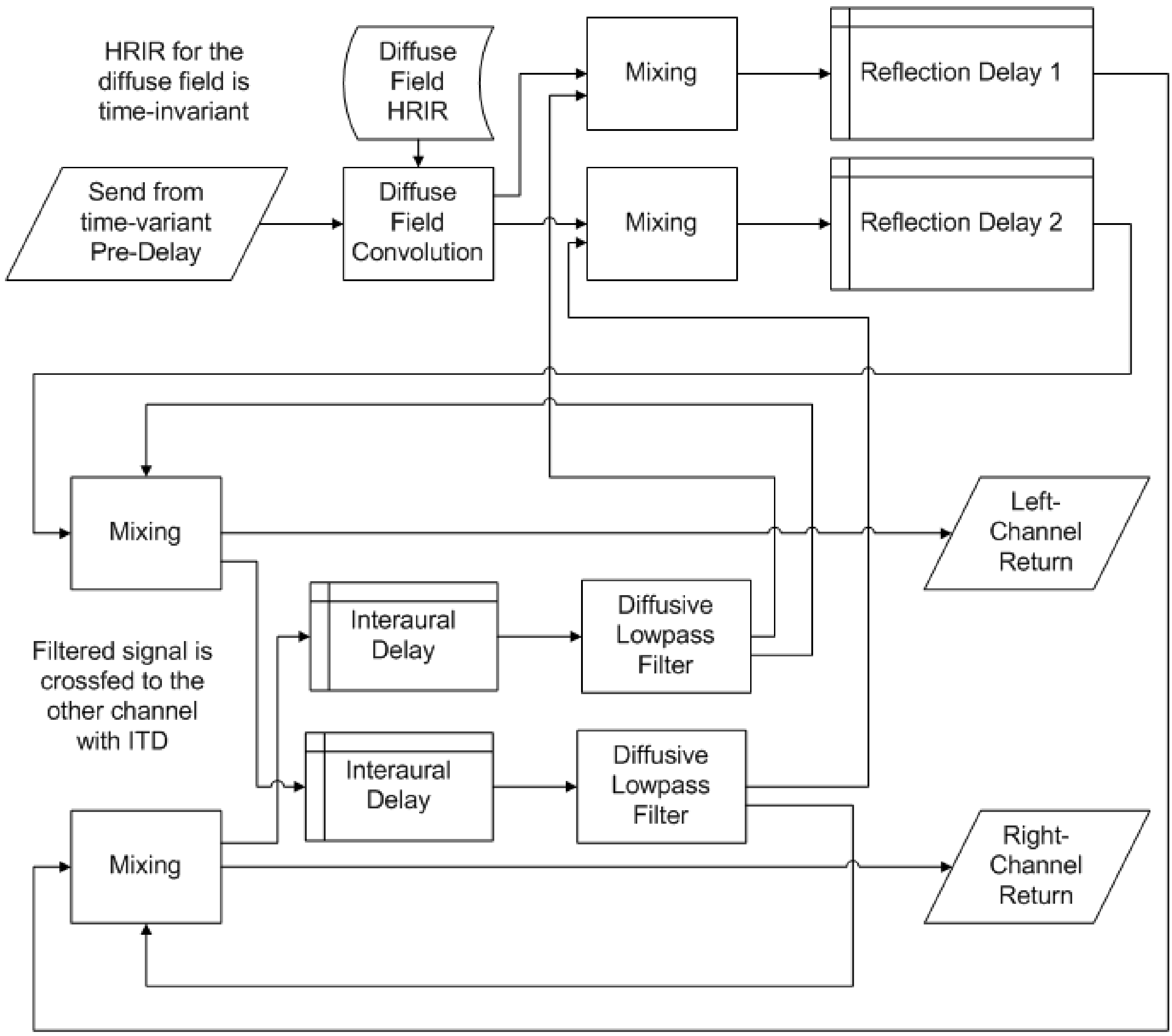

Figure 2. *Block diagram providing details regarding the Reverb Generation signal processing providing environmental simulation.*

Figure 2 shows a block diagram illustrating the signal processing responsible for the reverb generation system. Prior to reverberation processing, the single-channel send from the audio input delay buffer is convolved with the HRIR simulating diffuse field response for headphone reproduction. The proposed simulation of indirect sound includes both discrete early reflections and later diffuse reverberation, with a smooth transition between these two provided by progressive diffusion of recirculating delayed signals. In addition to creating a natural reverberant decay that attenuates high-frequency energy at a faster rate than low-frequency energy, the proposed signal processing creates natural interaural differences for each simulated reflection (by simulating both ITD and interaural spectral differences). This is accomplished with unprecedented computational efficiency by requiring the implementation of only four delay buffers of "relatively" fixed duration and two recursive lowpass filters. The four delay buffers are said to be of "relatively" fixed duration as all are in actuality undergoing a constant slow modulation between two delay lengths, via a crossfade between taps taken at two adjacent samples in the buffer, to create a subtle shifting that has been found to reduce the audibility of unwanted comb-filtering results (cf. [8] for similar reasoning applied to commercial reverberation units popular for music production). These delay length modulations are not dependent on source location parameters or reflection paths within the simulated environment, but rather are designed for optimal reverberation sound quality.

Three source positioning parameters may be updated in realtime, those being the azimuth and elevation angle of the source, and the egocentric range of the virtual source (its distance from the listener). Two environmental parameters may be updated in realtime, those being the initial time gap between the arrival of the source and the arrival of the first simulated reflection (termed Reverb Pre-Delay), and the relative level of the first refection and subsequent reverberation (termed the Wet Gain). These two parameters characterize the perceptually salient features of the relative size and relative liveness of the space within which the virtual source is positioned (see [2] for more in depth discussion of the interaction of direct sound level, indirect sound level, and initial time gap in determining the perceived range of a sound source when "liveness" is varied).

2.2 Listening Experiment

In a previous direct scaling study [1], the author confirmed that the auditory range of well-lateralized virtual sound sources could be manipulated by a frequency-independent ILD adjustment, at least for sources within close range of the listener (defined here as roughly within the listener's reach). The standard deviation of ratings for the experimental stimuli employed in that study were rather large, however, which raised the question of how robust this effect might be. In particular, this would be a concern for those seeking to utilize ILD for control of virtual source range in auditory display applications. A number of questions can be asked about the relative range of standard and comparison stimuli that don't require magnitude estimation as in the previous study. Instead of producing a number indicating how much closer the comparison stimulus was relative to a standard stimulus at a fixed and greater distance, the current study required only a choice of which stimulus seemed to be located at closer range. Here, the standard stimulus was not always placed at an extreme distance, as in the previous study [1], but rather could produce an auditory image that was either closer or farther than that associated with the comparison stimulus. In all the experiments reported here, the standard stimulus was held constant in every respect. Within a given block of trials, the ILD of the comparison stimulus was also fixed, but the overall gain on the comparison stimulus was varied over trials, sometimes making it louder than the standard, and sometimes making it softer. So on roughly half of the trials, the standard might be chosen as closer than the comparison, and the decision of which seemed closer was what the task required of the experimental listeners.

The method of constant response was utilized to estimate the point of subjective equality with regard to the range of the comparison and the standard stimuli. This method required the listener to complete a two-alternative, forced choice staircase tracking the level of the comparison stimulus that produced a range percept matching that of the standard stimulus. The question asked on each trial was, "Which of the two displayed stimuli seemed closer?" Five staircase turnarounds were completed before the set of trials was terminated. This procedure was completed for each of three ILD manipulations of comparison stimulus. These ILD values were superimposed upon the naturally measured ILD values present in the HRIRs of the listener. In one extreme case, the natural ILD was shifted by 9 dB through an attenuation of the contra lateral ear signal for the comparison stimulus. In the other extreme case, the natural ILD was not modified at all, and so the ILD for the comparison stimulus matched that of the standard stimulus. Of course, no shift in perceived range would be expected in the case with an ILD boost of 0 dB.

3. RESULTS

As explained above, the ILD of the comparison stimulus was manipulated by varying only the level of the contralateral ear signal, which was held in a fixed proportion to the ipsilateral ear signal to maintain constant ILD as the level of the comparison stimulus was varied relative to that of the standard stimulus. The upper panel of figure 3 plots as a function of the relative gain of the comparison stimulus the proportion of trials on which the comparison stimulus was chosen as "Closer" than the standard stimulus. When the comparison ILD was increased by attenuating the contralateral ear signal by 9 dB, the comparison stimulus had to be attenuated by more than 3 dB to produce an auditory image that matched the range of the standard stimulus. It is therefore not only the relative loudness of the stimuli that determined their relative range, since the comparison was certainly less loud than the standard stimulus; rather, the louder standard and the softer comparison produced auditory images that were judged to be at the same range only by this additional attenuation of the comparison stimulus. Without this additional attenuation, the comparison stimulus with greater ILD would produce an auditory image judged to be closer than that produced by the standard stimulus, around 90 percent of the time according to the results shown in the upper panel of Figure 3. The lower panel of Figure 3 allows these results to be compared against the zero dB ILD-boost control condition in which no shift in perceived range of the comparison stimulus would be expected. Though the staircase tracking of the point of subjective equality for range was completed at 4 levels of ILD boost (namely 0, 3, 6, and 9 dB), the results for only the two extreme ILD boost values are shown in Figure 3.

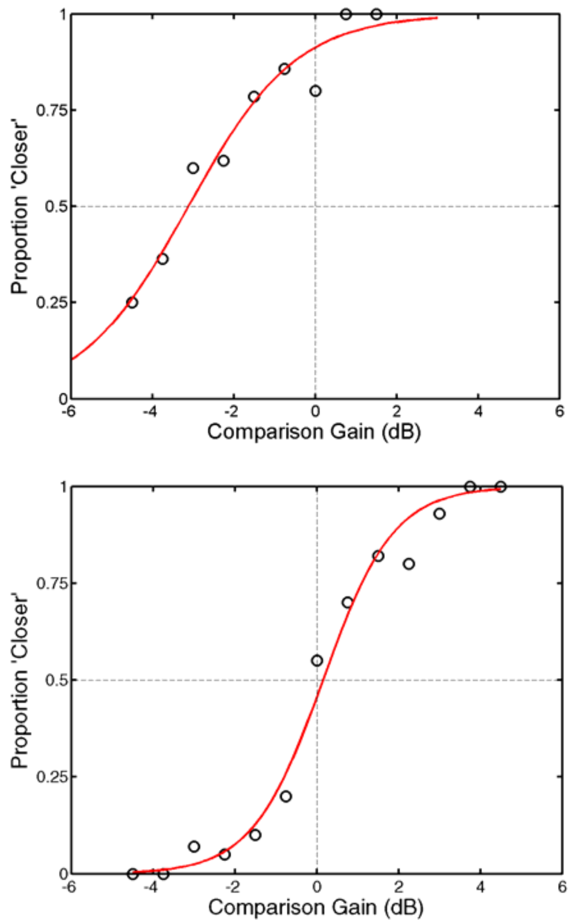

Figure 3, upper panel:. *Proportion of trials on which the comparison stimulus was chosen as "Closer" than the standard stimulus, when the ILD of the comparison stimulus had been boosted by 9 dB. Note that the comparison gain of zero dB is the same as that of the standard stimulus.*

Figure 3, lower panel: *Proportion "Closer" responses when the ILD of the comparison stimulus matched that of the standard stimulus (therefore identical at zero dB).*

Just as in the upper panel of Figure 3, the lower panel plots the proportion of trials on which the comparison stimulus was chosen as “Closer” than the standard stimulus, but this time in the case in which the comparison stimulus had the same ILD as the standard stimulus.

In effect, when the comparison gain value shown on the x-axis in the lower panel is zero dB, this indicates the situation in which the comparison stimulus and the standard stimulus were in fact identical. Therefore, the expected proportion of “Comparison Stimulus Closer” responses should be near 0.5 in this case. It should be noticed that the slope of the psychometric function fit to the proportions in the lower panel is steeper than the slope of the curve fit to the data shown in the upper panel of Figure 3 (curves derived here using logistic regression analysis). This suggests that the listener was more certain in the 0 dB ILD-boost control condition that the comparison stimulus was closer when the difference between comparison stimulus gain and standard stimulus gain was very small, and that a larger difference was required in the 9 dB ILD-boost control for the same choice probability for a source range difference.

4. DISCUSSION

The obtained range matches for virtual sources processed using exaggerated ILD values show that roughly equal loudness does not make auditory spatial images at the same perceived range. Likewise, they show that a softer virtual source might actually be shifted to a position closer to the listener’s ear than a louder virtual source using such processing that mimics the variation in HRTFs at close range [4]. This should be no surprise, since several previous studies using alternative psychophysical methods have shown similar results [e.g., 1, 7]. What is novel here is that the employed spatial sound processing technology included an indirect sound simulation that placed the virtual sources in a small virtual acoustic space. This indirect sound simulation, while providing externalization superior to that associated with dry HRTF-based processing, may well have provided a context that made judgments of range less difficult to make (cf. [9]). It may also be that the perceptual consequences of including realistic reverberation in spatial auditory display are not as objectionable as been conventionally held, a point well made in recent papers by Shinn-Cunningham [e.g., 10, 11]. Of course, the generality of the current results is limited in a number of important ways, which may be summarized as follows: First, the results apply only to virtual sources that are well lateralized (as was also pointed out in [7]). Second, the employed indirect sound simulation placing the virtual source in a small virtual acoustic space would not be appropriate for larger spaces where the reverberation could be “heard out” as separate from the “precedent” sound, since the short duration of the indirect sound included here fused perceptually with the direct sound and gave no real sense of “room character.” It should also be pointed out that varying the initial time gap will affect on the resulting range percepts associated with this sort of simulation, and variation in this parameter was not investigated in the current study. Of course, dynamic modulation of the initial time gap may also be used as a cue to variation in source range, as taught in [12].

5. CONCLUSIONS

For headphone-based spatial auditory display systems, binaural processing solutions typically use source reproduction level as the primary control for source range. This approach can be quite effective when indirect sound is simulated in order to externalize virtual sources. A computationally efficient simulation solution for binaural synthesis of sound localization cues was presented which does not rely solely upon the sound reproduction level of the source to control the apparent range of a virtual sound source. The solution includes an economical synthesis of the indirect sound associated with a small virtual acoustic environment, which is effective in creating externalized spatial auditory images. Experimental listening tests confirmed that the simulation allows for decoupled loudness and range control for a source located near the listener’s head, so that within certain limits, source loudness and source range may be independently varied. This simulation solution should be especially useful in auditory display applications that require selective attention to multiple sound messages delivered from different spatial positions, but without the typical coupling of variation in source loudness and source range.

6. REFERENCES

[1] W. L. Martens, “Perceptual Evaluation of Filters Controlling Source Direction: Customized and Generalized HRTFs for Binaural Synthesis,” Acoustical Science and Technology, 24 (5), 220-232, Sept., 2003.

[2] W. L. Martens, “Psychophysical calibration for controlling the range of a virtual sound source: Multidimensional complexity in spatial auditory display,” In: Proceedings of the 7th International Conference on Auditory Display, Espoo, Finland, July, 2001, pp. 197-207.

[3] B. G. Shinn-Cunningham, S. G. Santarelli, and N. Kopco, “Tori of confusion: Binaural cues for sources within reach of a listener,” Journal of the Acoustical Society of America, 107(3), 1627-1636, 2000.

[4] D. S. Brungart and W. M. Rabinowitz, “Auditory localization of nearby sources: Head-related transfer functions”, Journal of the Acoustical Society of America, 106(3), pp. 1465-1479, 1999.

[5] R. O. Duda and W. L. Martens “Range dependence of the response of a spherical head model,” Journal of the Acoustical Society of America, 104 (5), 3048-3058, 1998.

[6] D. J. M. Robinson and R. G. Greenfield, “A Binaural Simulation which Renders Out of Head Localization with Low Cost Digital Signal Processing of Head Related Transfer Functions and Pseudo Reverberation,” Proceedings of the 104th Audio Engineering Society Convention, Preprint 4723, May, 1998.

[7] D. S. Brungart, “Near-field auditory localization,” Proceedings of the 3rd International Conference on Auditory Display, Palo Alto, CA, Nov., 1996.

[8] J. Dattorro, “Effect Design, Part 1: Reverberator and Other Filters”, Journal of the Audio Engineering Society, 45(9), 660-684, 1997.

[9] W. L. Martens, “Discrimination of Spatial Offsets on the Cone of Confusion: Closer/Farther, Upward/Downward, and Frontward/Rearward.,” Proceedings of the International Congress on Acoustics, Kyoto, Japan, April 2004.

[10] B. 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.

[11] B. G. Shinn-Cunningham, “Speech intelligibility, spatial unmasking, and realism in reverberant spatial auditory displays,” in Proceedings of the International Conference on Auditory Display, July 2002, 183-186.

[12] G. S. Kendall, W. L. Martens, & M. D. Wilde, “A spatial sound processor for headphone and loudspeaker reproduction,” Proceedings of the 8th International Conference of the Audio Engineering Society, Washington D.C., May, 1990.