ICAT 2001
December 5-7, Tokyo, Japan

Subjective Evaluation of Room Geometry in Multichannel Spatial Sound Reproduction: Hearing Missing Walls in Simulated Reverberation

Kenji Suzuki and William L. Martens

Multimedia Systems Lab, University of Aizu
Aizu-Wakamatsu 965-8580, Japan
voice: [+81](242)37-2791; fax: [+81](242)37-2731
e·mail: m5041116, wlm@u-aizu.ac.jp

Abstract

The goal of this research project was to determine whether a relatively simple 3D model for multichannel simulation of room reverberation could produce identifiable differences in room geometry. This simple, image-model-based simulation was designed to produce distinctive-sounding results as the material is varied on each of the six walls of a modeled rectangular room. In particular, a realistic-sounding wall-reflection simulation was employed that captured differences between wall materials via a **R**eflection **T**ransfer **F**unction (RTF). By varying this filter for each wall individually, rooms with six walls of different material could be simulated. In the extreme cases investigated in this study, one wall at a time was made completely absorbent, and it was a primary focus of this research project to determine whether listeners could hear which of six walls had been selectively eliminated from the simulated room reverberation. A secondary focus was to determine the means by which listeners might able to make this identification. The results of blind listening experiments showed that listeners were not particularly good at determining which one of five walls had been eliminated (when comparing between five simple cases of missing walls, within which cases only the floor of the rectangular room model was never removed). On the other hand, listeners were able to consistently distinguish between the spatial images associated with these five cases (five room geometries) in terms of subjective attributes such as perceived **L**istener **E**n**V**elopment (LEV). The results were quantified via **M**ulti**D**imensional **S**caling (MDS) analysis of the perceived similarities between the spatial images, and interpreted via direct ratings of their perceived spatial attributes.

Keywords: multichannel sound reproduction, virtual acoustic rendering, binaural hearing, psychological acoustics, subjective evaluation

1. Introduction

The relatively young science of virtual acoustics needs answers to basic questions such as how to best simulate complex virtual acoustical environments. Models simulating rectangular rooms having walls with unequal absorption properties may be computationally intense if physically accurate results are required. In many applications, however, it is sufficient to create a spatial image that is recognizable by human listeners. Indeed, it may be that physical accuracy in simulated reverberation does not guarantee accurate perception of room geometry, since some changes of room geometry may not be identifiable from changes in spatial imagery. In the interest of improving audio rendering efficiency, simplified filtering models have been developed and their audio output has being submitted to perceptual evaluation. This approach directly addresses a technological need that has not been adequately addressed elsewhere. Existing models of virtual acoustical events admit of many levels of detail, ranging from simple parametric models to full-blown, three-dimensional (3D) solutions aimed at accurate "auralization" [1]. There is typically no indication given in prior work regarding the ability of human listeners to hear changes in the acoustical details being simulated, such as room geometry, differential absorption due to changes in wall materials, etc. The research described in this paper is part of an ongoing project to determine what acoustical features are necessary and sufficient to provide a virtual acoustic simulation result that is satisfying to the human listener. In 3D graphic rendering for virtual environments, the material properties of each wall are rather well described by a few attributes such as specular and diffuse reflection coefficients; however, in 3D audio rendering for the same virtual environments, the material properties of each wall must be inferred from changes in the overall spatial auditory imagery in a manner that depends as much upon the sound source as the environment. This research project began with a single simple question about human spatial hearing in virtual acoustic environments: It was questioned whether listeners can judge which of six walls seems to be missing from a multichannel reproduction of simulated room reverberation based upon the sound of a single, spatially-stationary sound source. It was also of interest to discover the means by which listeners are able to make this determination, since it is questionable that a missing wall produces a sensation that could be called the "missing-wall" sensation. Rather, the identification of which wall has been removed might be more likely made by virtue of that the spatial imagery associated with an imbalance in the reflections that the listener hears, rather than those that are not heard. Therefore, this investigation included some perceptual evaluation of the overall perceptual similarities and dissimilarities between the spatial images of rooms with various walls missing.

The studies reported here used a multichannel-loudspeaker reproduction of simulated room reverberation based upon a 3D model of the simulated enclosure. The simulated reverberation was presented in an anechoic chamber via a 4.2 channel reproduction system (where the ".2" means that two subwoofers were used). The four satellites in this system act as "acoustic windows" for mid- and high-frequency stimulation, and the two subwoofers are placed on either side of the listeners to allow for low-frequency support for LEV [2]. As walls were selectively removed, changes in spatial attributes, such as LEV, potentially enable listeners to identify differences between simulated reverberant spaces. Using a realistic-sounding wall-reflection model (captured as a frequency-dependent RTF), various rooms were simulated. In preliminary experiments, many paired comparisons were presented, the first stimulus of the pair being the standard (having all walls present) and the second one having identical values on all parameters except that one of the walls was missing from the simulation. The results of these preliminary listening experiments generally showed that the location of missing walls is more easily identified in simulations containing higher numbers of reflections, and so the experiments reported in this paper were based upon simulations that included reflections up to 5^{th} order (i.e., those that have reflected from walls five times).

2. Methods

2.1 General Context

Our multichannel reverberation simulation model was based on the idea of a room within a room, probably first presented in Moore's 1983 paper [3] entitled "A general model for spatial processing of sounds." The concept here is contrasted with the more common head-related auditory spatial sound reproduction, described first in Bauer's seminal 1961 paper [4] entitled "Stereophonic earphones and binaural loudspeakers." Moore's general model has been termed room-related reproduction, since the indirect sound simulation was explicitly intended for multiple listeners located at arbitrary locations within the reproduction space. In such room-related spatial sound reproduction, the loudspeaker locations act as windows for sound to enter an imaginary box defined by connecting those loudspeaker locations, and virtual sources are typically localized only outside the borders of this box (localized, that is, outside the boundaries defined by the perceived loudspeaker locations within the listener's auditory space). In contrast, the loudspeakers in head-related spatial sound reproduction should disappear to enable virtual sources to be localized anywhere in the listener's auditory space.

In multichannel simulation, how many loudspeakers are used is an important point for implementation (see [5]). Instead of the more conventional 5.1 channel system (meaning five satellites and one subwoofer), the present system is best termed a 4.2 channel reproduction system (where the ".2" means that two subwoofers were used). The four satellites in this system act as "acoustic windows" for mid and high frequency stimulation, and the two subwoofers are placed on either side of the listeners to allow low frequency listener envelopment (see [6]). Whereas the surround sound loudspeaker angles in the conventional 5.1 channel system are optimized for creating spacious auditory imagery, the four satellites in the current implementation were spaced in equal 90 degree intervals, to maximize coverage of space for placing virtual sound sources.

The system was developed on an SGI workstation with four channel sound output. For simulating indirect sound an image model was used to determine the delay and level of early reflections (see [7] for a general introduction on the image model, first reported by [8]). Each of the walls simulated (six in the case of a rectangular room that included floor and ceiling) had its own absorption coefficient and filter for simulating wall material. The material property simulation enabled the creation of virtual acoustic environments with walls missing, as well as walls of different material. In this study, a single, percussive sound source was presented at a fixed azimuth angle of 30^o, providing a more difficult test case for reflection detection than do moving, continuous sources.

2.2 Sound Simulation Model

Sound simulation model in our system has two rooms; one is outer virtual acoustic room and another is inner actual fitting reproduction room (space). Shape of outer virtual acoustic room is just a simple cube and all six walls have their own material simulation filter and reflection coefficient. The sound emanates in the outer room and go into the inner room through "acoustic windows" (loudspeaker position) after reflection(s). We uses the image model to simulate accurate early reflections and the distance is calculated between each image source and corresponding loudspeaker which play that image source. Frequency dependent reflection is simulated by using 2^{nd} order IIR filter. Simulated reflections are passed to the block of gain control for two loudspeakers to reproduce four channel sound output (details will be described in Section 2.6).

2.3 Sound Reproduction System

In our sound reproduction system, an SGI Indigo2 workstation supporting four channels sound output, SONY AV receiver (designed for 5.1 channels) to handle 4-channel amplification, four broadband loudspeakers and two subwoofers are used (Fig. 1).[1] A pair of left and right output signals were amplified by the AV receiver and these stereo pairs of outputs unit are then fed into the left and right subwoofer respectively, such as both front and rear left-channel signals are summed at the input to the left subwoofer, and the front and rear right-channel signals are summed at the input to the right subwoofer. The two subwoofers provide not only surrounding low frequency energy, but also extended control over auditory spatial imagery [6]. Separate L/R input to the two subwoofers enables virtual sources to move between extreme left and right locations, well beyond the 45^o angle of the satellite loudspeakers, reaching the 90^o angle of the subwoofers. Without the subwoofers, such lateralized imagery is only possible with cross-talk cancellation (*a.k.a.* transaural stereo [9], the inclusion of which was contrary to the design goals of our system). As our reverberation simulation is "room-related" rather than "head-related" [10], our system exhibits none of the advantages associated with cross-talk cancellation (e.g., precise control over interaural cross-correlation at the listener's ear is possible only when the system is head-related).

Low frequency panning between two subwoofers provides better spatial coverage, that is, a wider listening area than sound reproduction using just one. Four loudspeakers are positioned at each corner of the square which fits the simulated inner room in sound simulation model. Within this space (inner room), listeners can be located at arbitrary position. Even though the center of inner room provides the best spatial sound reproduction, listeners near the boundary of inner room still hear imagery that is appropriate to their listening position, since the model treats the loudspeakers as if they were "acoustic windows" into the larger simulated room.

Even though listeners can be located at arbitrary position within reproduction space, in our listening experiments described below, they were always seated in precisely the optimal position, as illustrated in Fig. 2. Despite the advantages of spatial sound reproduction using four rather than two loudspeakers, there still many limitations. Sitting on the center line (equal distance between both left and right loudspeakers) yields better spatial imagery for sources that are intended to arrive from a direction midway between left and right loudspeakers. The same holds true for front/back localization. In spite of these limitations listeners can still get almost the same information about reflections, including a global impression of the overall spatiotemporal distribution of reflections. Since our system supports four-channel sound output, we can control front/back distinctions difficult to support using only two loudspeakers. Also, sound image control between (two) loudspeakers is easier than other multichannel reproduction system (6-channel, 8-channel, etc) and it is good and important point for multichannel reproduction system.

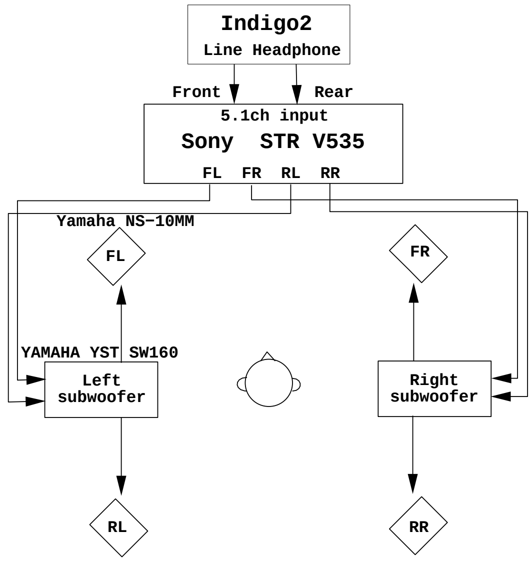

Fig. 1: Sound Reproduction Flowchart

2.4 Control of Source Direction

Sound source direction was controlled by adjusting the output gain between two loudspeakers at a time, mimicking Interaural Level Difference (ILD) for speaker pairs in left $\Longleftrightarrow$ right opposition. Fig. 3 shows the gain control between two loudspeakers. Of course, panning amplitude between loudspeakers in front $\Longleftrightarrow$ rear opposition does not match the stimulus variation associated with actual motion from front to rear, at least not in the way that left $\Longleftrightarrow$ right panning does. Nonetheless, informal listening test confirmed that front $\Longleftrightarrow$ rear distinctions were well supported by the current gain control scheme. Four loudspeakers were placed at the each corner of the square within

[1] Users of the Indigo2 and Indy workstations should know that these SGI computers reproduce four channels of audio with no special hardware. By selecting the "four-channel output" mode in the audio control panel, the four channels of sound are routed to the audio connections as follows: front left/right signals are sent to "Line Out" and rear left/right signals are sent to "Headphone Out".

Fig. 2: The 4-channel loudspeaker setup in the University of Aizu's large anechoic chamber.

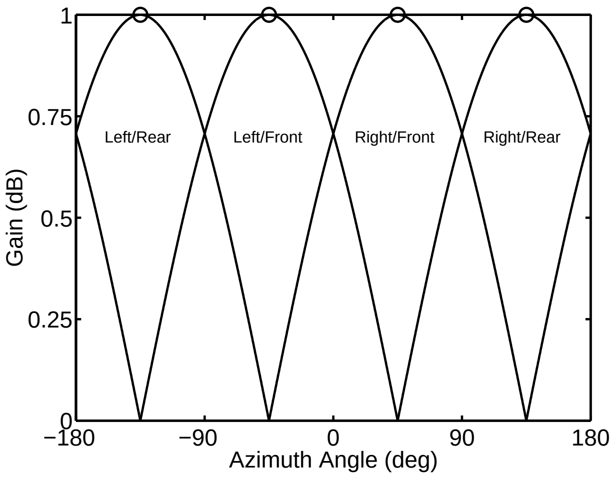

Fig. 3: Four-channel gain control scheme.

which the listener was positioned. When the sound source was positioned at azimuth angle of 45^o, it was reproduced only by the **F**ront/**R**ight speaker (FR in Fig. 1). A sinusoidal curve was used for the gain control function so that the total energy would be constant in any direction as the source was panned between loudspeakers. In addition to gain control between each speaker pair, Interaural Time Difference (ITD) was simulated by applying delay to the two speaker signals for direction of sound in addition to the propagation time from source location to speaker location.

2.5 Wall Material Simulation

Most wall has generally frequency dependent magnitude response. This effect is affected by the surfaces and materials of wall. Simulating wall materials is important to render natural-sounding reflections. To simulate wall materials, we use 2^{nd} order IIR filter implemented as a direct form II transposed structure (see Eq. 1). $x[n]$ is the input, $y[n]$ is the output, and a and b are filter coefficients for feed-back and feed-forward part respectively. This filter is a combination of two 1^{st} order IIR filters; one simulates low frequency energy absorption and another simulates high frequency energy absorption. Each 1^{st} order IIR filter has a pair of filter coefficients for feed-forward and feed-back, and a 2^{nd} order IIR filter is constructed from another pair of coefficients from a second 1^{st} order filter. The linear difference equation for this filter (with $a_0 = 1.0$) is thus:

$$\begin{aligned} y[n] &= b_0 * x[n] + b_1 * x[n-1] + b_2 * x[n-2] \\ &\quad -a_1 * y[n-1] - a_2 * y[n-2] \end{aligned} \qquad (1)$$

With this filter implementation, we can control the attenuation of low and high frequency independently (Fig. 4). In these figures, filter A attenuates both low and high frequency, with more absorption in the high frequency portion. Filter B has the same low frequency attenuation, but more high frequency attenuation of filter A. This filter renders the most natural-sounding reflections. Filter C has no low frequency attenuation and the same high frequency attenuation as filter B. In the implementation of these filtering simulations, we normalize frequency dependent amplitude attenuation to maintain a maximum of unity gain.

2.6 Early reflections

Early reflections are simulated by using the image source method, because we need accurate early reflections which tell propagation delay, distance atten-

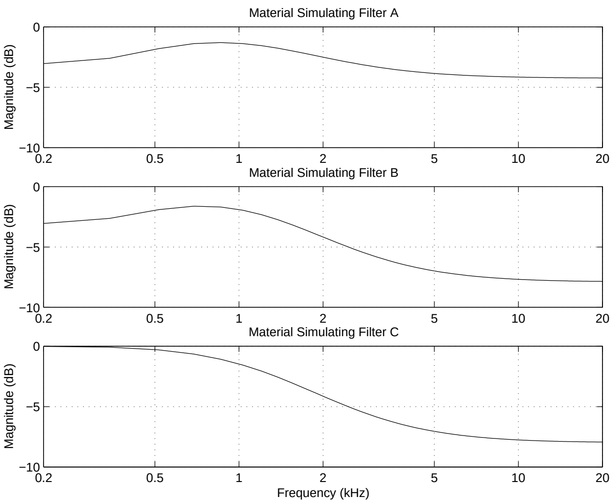

Fig. 4: Magnitude response of three material simulating filters. Filters A and B have the same low frequency attenuation, while filters B and C have the same high frequency attenuation.

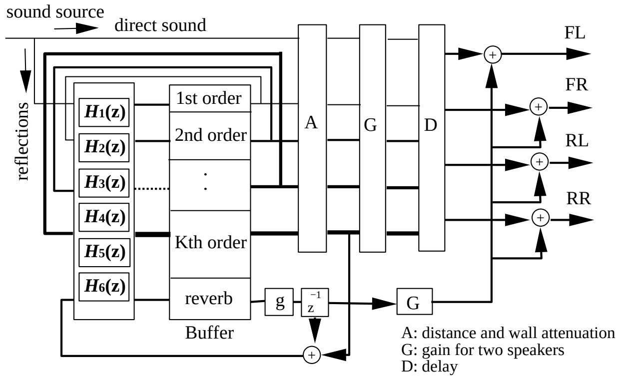

Fig. 5: Block diagram of our implementation. The blocks A, G, and D use look-up-tables (LUTs) for all indexed image sources. Filtered signals are stored in a buffer and then processed appropriately.

uation, from where the sound comes, and what wall the sound reflected off to examine which cues are most salient in allowing the listener to perceive which wall is missing. The number of image sources is determined by max order of reflection and it increase exponentially as max order increase. Fig. 5 shows the block diagram of our system. In this figure, $H_1(z)$, $H_2(z)$,..$H_6(z)$ represent the 2^{nd} order IIR filter described in Section 2.5 for front wall, left wall, and so on. We prepare simple drawing for this diagram; for example, output line from 1^{st} order reflection buffer is one. But actually, the number of buffer for 1^{st} order reflection is six, this means, output line from 1^{st} order reflection buffer in this figure consists of six lines. The line connected to buffer for 2^{nd} order reflection consists of the number of 2^{nd} order reflections (= 18) as well. The way to render 1^{st} order reflection for one image source is, first, the sound source is filtered by one of six wall material filter (this is determined by the image source position), and then, the filtered signal is stored at buffer for getting 2^{nd} order reflection. We use look-up-table for which image source should be filtered by what filter. Also, this table has information for distance attenuation, wall absorption, and delay for all image source. Next, filtered signal is applied distance attenuation and wall absorption (Block A). At last, signal is applied gain control for two loudspeakers (Block B), and then, applied delay (Block C). The way to render 2^{nd} order reflection is same as 1^{st} order reflection, and the input signal is the filtered signal stored in the buffer. This process is ended up K-th order. In the rendering process of 2^{nd}- and higher-order reflections, some image sources have almost the same filtering; for example, assuming that image source A is first reflected off the front wall, and then, reflected off the rear wall, image source B is reflected off walls in reverse order of A. In this case, the difference between filtered signals of A and B is quite small. Therefore, for computation efficiency, the filtering is shared in the rendering process of image source B (if A is rendered before B). We also use a look-up-table to find which image sources have the same filtered signal for such simplification.

Levels and delay of early reflections are calculated in above way. We can get accurate response of virtual acoustic room we simulate with varied order of reflection. Fig. 6 and Fig. 7 represent the temporal and spatial distribution of direct sound and all image sources of 5^{th} order reflection respectively. These figures were drawn from the view point of the listener. The listener is assumed to sit on the chair at the center of inner room and ear level is 1.2 m from the floor. The sound source is positioned at 30° right to the front of listener (azimuth angle of 30°), height is 1.5 m (slightly higher), and about 3.4 m away from the listener. Geometry of virtual acoustic room is 20x28x34 (medium sized and long room) so that reflections from walls reach to the listener at different time and come from different angle. All walls have the same reflection coefficient $g = 0.8$ (-1.94 dB down per each reflection). The direct sound is the 0 dB reference level for reflections in both figures. In Fig. 7, the circles represent all image sources; the radius of circle represents the volume of sound after applying wall absorption and distance attenuation. This figure tells us from where the sound comes (azimuth and elevation), their volume after reflected off one or several walls, and the number of reflections from the walls. The direct sound is represented as the bold circle. The arrival time of the earliest 1^{st} order reflection comes from right and left wall is near 50 ms with this long room. In case of wide room, these reflections comes after reflections from front and rear wall. This should affect the perceived information of geometry of the room.

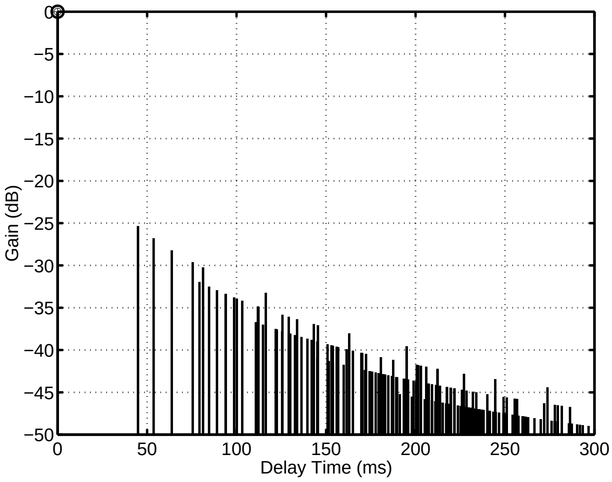

Fig. 6: Temporal distribution of 5^{th} order reflection with all walls by the image source model

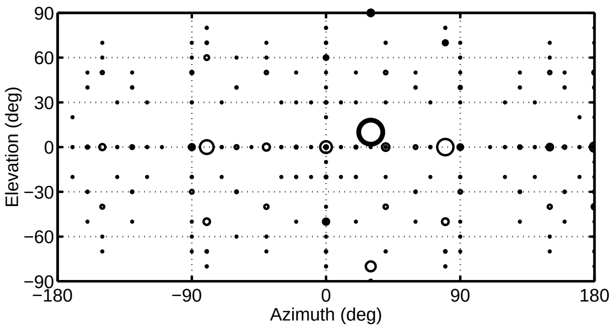

Fig. 7: Spatial distribution of 5^{th} order reflection with all walls

Fig. 8 represents the spatial distribution of image sources of 5^{th} order reflection and only front wall is missing. Actually, the magnitude of the missing image sources is zero, we still represent such image sources as 'x' for easier understanding of missing image sources' angle in this figure. Virtual acoustic environment is the same one as simulated for Fig. 7 except reflection coefficient of front wall was changed to zero. This figure indicates that what image sources reflected off the front wall at least one time and spatial distribution of such image sources. This figure also indicates the fact that not only reflections from front wall were missing, but also some reflections from other three directions were missing. Those missing image sources are positioned at, for example, about $\pm160°$ (rear left/right) and $\pm60°$ (front left/right) in azimuth.

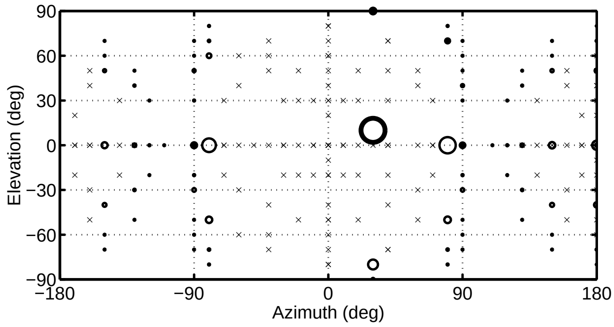

Fig. 8: 5^{th} order reflections when front wall is missing. Missing image sources are represented using the "x" symbol.

2.7 Late Reverberation

Late reverberation can be based upon the image model, just as the early reflections were, though it should be generated recursively rather than discretely [11]. The input for late reverberation synthesis is the K^{th} order filtered signal (K is the max order of reflection for simulated early reflections). The input is then filtered by filter B in Fig. 4 which makes natural-sounding reflections for recursive network. Then, the filtered signal is multiplied by gain g and delay z^{-n} which are determined in terms of the mean free path of the simulated room. At last, signal is passed to both feed-back loop and gain for two loudspeakers block G. In the rendering process of reverberation, if the image sources of K^{th} order reflection do not exist because one or more walls are missing, filtering for such signals are eliminated.

3. Listening Experiments

3.1 Preliminary Experiments

To examine which cues were most salient in allowing the listener to perceive which wall was missing, forty different room conditions were prepared (four room shapes, two material filters, and five possible missing walls). In a pilot study, the maximum order of reflections included in the simulation was varied from 2^{nd} to 5^{th} order. The two absorption material filters were designed to produce natural-sounding reflections (meaning no fluttering sound but not so much absorption as to make the spatial imagery sound dry). The difference between the two filters was the amount of low frequency energy absorbed. Magnitude response of one filter was presented in Fig. 4 of Filter B. For the preliminary listening experiment then, 160 stimulus pairs were presented; the first stimulus of the pair was the standard (having all walls present) and the second one had identical values on all parameters except that one of the walls was missing from the simulation. The task was for listeners to identify which of five walls was missing, the alternatives being left, right, front, rear, or ceiling. Although informal listening would seem to suggest that it is quite possible to perform this task successfully, in a blind listening test with a spatially-stationary sound source, performance was almost always near chance levels (i.e., at 20%, since there is one chance in five of guessing correctly between the five walls selectively removed from the simulation). Nonetheless, the following tentative conclusions seem warranted: First, the location of missing walls was more easily identified in simulations containing higher order reflections, and the reduction of low frequency content in the reflections also made identification less difficult. On the other hand, room shape seemed to play the role of a random variable in this particular experiment, as no systematic effects of this variable were observed. Of course, this failure to find effects of room shape on identification performance may be due to a 'floor effect," since performance was generally quite poor.

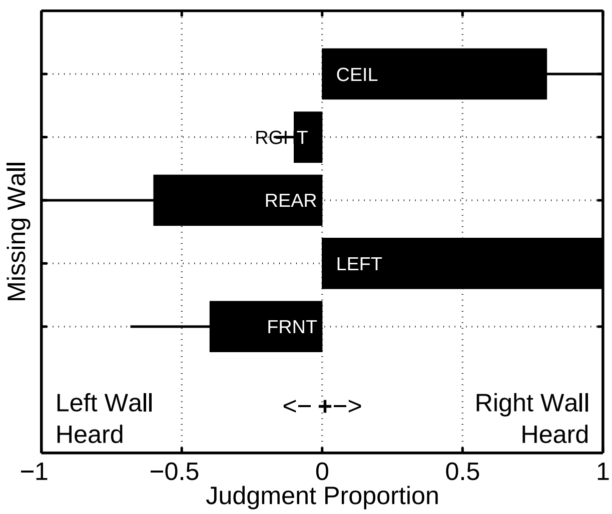

Fig. 9: Averaged results of four subjects in the 3AFC identification task. Standard error is indicated by the line segment extending horizontally from each bar. See text for details.

3.2 "Heard-Reflection" Identification

As the identification of which wall was missing required listeners to report on the direction from which reflected sound seemed **NOT** to come, the task was modified to one in which the listener reported the direction from which they thought reflected sound **DID** seem to come. Furthermore, the task was simplified to an indication of whether the "Heard-Reflection" arrived from the right or left. Also, in contrast to the 5-alternative forced choice (5AFC) task performed in preliminary listening experiments, the current task required the selection of a response from only three alternatives: from the left, from the right, or from neither side (meaning simply balanced in lateralization, whether due to clear centering, or unclear localization). Also, to simplify the listening task even further, only one room shape was presented, since there is a chance that varying room shape was a source of confusion to the listeners. Under these conditions, the 3-alternative forced choice (3AFC) task yielded much improved performance relative to the 5AFC, though some further exploration of the results may still be needed for the results to be properly interpreted.

Figure 9 shows judgment proportion averaged over four listeners for five stimuli. The judgment proportion totaled -1 if a given stimulus was given the response "Left Wall Heard" on all trials for a given stimulus, and the judgment proportion totaled $+1$ if the response "Right Wall Heard" was given. The horizontal bar plot in the figure shows that the room simulation with the left wall missing (labeled "LEFT" inside the plotted bar) was virtually always identified as a spatial image with a strong reflection arriving from the right (therefore, a positive judgment proportion). If, however, the right wall missing from the room simulation (labeled "RGHT" inside the plotted bar), no strong dominance of "Left Wall Heard" was observed. Thus, when the right wall was missing, the listener was not provided with a strong cue to the location of the remaining (left) lateral wall within the presented virtual acoustic environment. On the other hand, removing the rear wall from the room simulation did, in fact, increase the proportion of judgments of "Right Wall Heard." Of course, some strong reflections from the rear wall do come from the rear-right in the simulation, and this explains why a rear reflection might be judged as coming from the right. Likewise, even though many left-wall reflections are present in the simulation with left wall removed, a different pattern of left wall reflections is produced when the right wall is not removed (e.g., when only the rear wall is removed), and therefore the left-wall reflections may be more noticeable to the listener under these conditions. The problem with a simplistic interpretation of the effects of removing virtual walls is that the listener never hears the walls themselves, but only the pattern of reflections arrival from all angles in the simulation (and this is true in actual environments as well). Therefore, a means of investigating what **IS** audible in the reflection pattern seems to be required. This need is addressed in the following exploratory study of the differences between the spatial imagery associated with the five room simulations presented here.

3.3 Dissimilarity Judgments

As this stage of the study was primarily exploratory in nature, and as an adequate model for predicting perceptual differences between stimuli is not available for such spatial sound reproduction, a perceptual scaling study was designed. In this stage of our investigation we attempted to uncover the perceptual structure underlying judgments of inter-stimulus similarity. The most common analytic tool used for such exploratory investigation is MultiDimensional Scaling (MDS), in one of its many implementations that have evolved over several decades. It is instructive to read an early explanation of the role of MDS in this context from Torgerson's 1952 book [12]:

> The traditional methods of psychophysical scaling presuppose knowledge of the dimensions of the area being investigated. The methods require judgments along a particular defined dimension ... In many stimulus domains, however, the dimensions themselves, or even the number of relevant dimensions, are not known. What might appear intuitively to be a single dimension may in fact be a complex of several ... Other dimensions of importance may be completely overlooked. In such areas the traditional approach is inadequate ... This model differs from the traditional scaling methods in two important respects. First, it does not require judgments along a given dimension, but utilizes, instead, judgments of similarity between the stimuli. Second, the dimensionality, as well as the scale values, of the stimuli is determined from the data themselves.

Of course, individual subjects may differ in how they form judgments of global dissimilarity, and so a refined method for doing a *weighted* MDS analysis [13] that takes such individual differences into account is to be recommended. This paper teaches the use of INDSCAL (INdividual Differences SCALing) [14] analysis as a powerful means for deriving an interpretable representation of the dimensions underlying reported inter-stimulus dissimilarities obtained from a potentially inhomogeneous group of subjects, each of whom may place different *weights* upon each of the perceptual dimensions. While sets of dissimilarity data can be averaged across subjects to obtain one aggregated dataset for submission to classical MDS analysis [15], this paper shows the advantages provided by the INDSCAL model for the analysis of multiple sets of dissimilarity data, without requiring the assumption of a homogeneous group of subjects who share an identical perceptual structure for the stimuli. Beyond this, the two primary advantages of INDSCAL are as follows:

1. INDSCAL provides a quantitative characterization of the individual differences that exist within a group of experimental subjects, based upon dissimilarity judgments obtained from each subject. The individual differences are captured in a set of *weights* placed upon each of the stimulus dimensions by each subject.
2. INDSCAL provides an inherently unique configuration solution that requires no further analysis to find a meaningful rotation, in contrast to the orientational ambiguity inherent to classical MDS[2] .

Following the successful derivation of a two-dimensional (2D) perceptual space for the 5 different room geometries, there remains to be solved the important problem of interpreting the identity of the perceptual dimensions. For this reason, was decided to obtain ratings of the same stimuli on a particularly salient attribute of the spatial auditory image, as an aid in interpreting the meaning of the dimensions of the INDSCAL configuration solution. It is not assumed, of course, that the ratings on the single attribute selected will necessarily capture differences between stimuli with regard to the fundamental perceptual differences between the stimuli. Nonetheless, our informal listening indicated that the variation in perceived **L**istener **E**n**V**elopment (LEV) was particularly important for the stimuli currently under investigation. Four listeners with no reported hearing loss served as subject in both of the experimental sessions reported here. In this first experimental session, each listener gave dissimilarity ratings on a 5-point scale for all pairwise comparisons of the 5 sound stimuli. The obtained dissimilarity ratings were submitted to INdividual Differences SCALing (INDSCAL) analysis using the ALSCAL procedure of the SPSS software package [16].

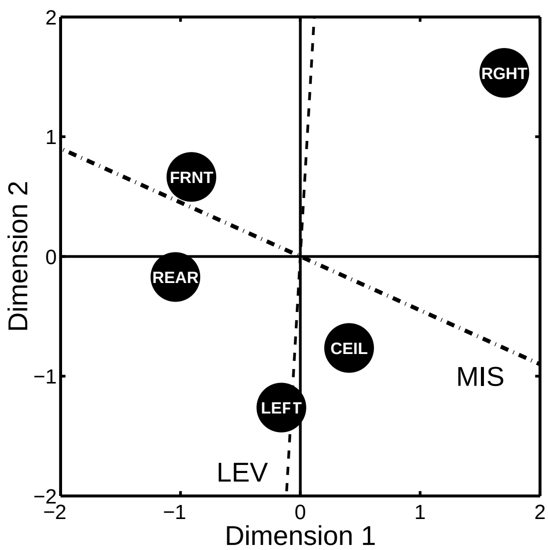

Fig. 10: INDSCAL-derived *Subject Space* with projections of attributes LEV and LEV indicated by the dashed line segment and the dot-dashed line segment, respectively. See text for details.

[2] As noted by Shiffman et al [13], the non-rotatability of the INDSCAL solution assumes error free data. Some rotation may be justified in the presence of error.

Fig. 10 presents a summary of the 2D INDSCAL analysis results. The plot shows the configuration of the stimuli in what is termed the *Group Stimulus Space* that results from transforming obtained inter-stimulus dissimilarity into estimated inter-stimulus distance in a 2D Euclidean space. The goodness of fit of the INDSCAL result is typically reported as a single value of *stress* between the dissimilarities from individual subjects and the distances in the single, common *Group Stimulus Space*. The results of the 2D INDSCAL analysis yielded a *stress* value of 0.214 ($RSQ = .863$), and the individual differences in the weights put on each of these INDSCAL-derived dimensions were well distributed within the INDSCAL-derived *Subject Space* (i.e., no subject emphasized one dimension over the other in an extreme manner). Two of the subjects emphasized Dimension 1 over Dimension 2, while the other two subjects emphasized Dimension 2 over Dimension 1.

Interpreting the results of classical MultiDimensional Scaling (MDS) is problematic because the resulting *Stimulus Space* can be rotated through an arbitrary angle without violating the structure of the solution. Of course, inter-stimulus distances remain invariant under rotation of both classical MDS and INDSCAL solutions alike; but the orientation of the INDSCAL solution is determined by modeling agreement between subjects. INDSCAL is designed to separate those factors that are common to a group of subjects from the ways in which subjects differ. Although the subjects in the current study seem to generally agree, the fact that there is some variation in the *Subject Space* vectors is advantageous, for it is according to these differences between subjects that the orientation of the obtained *Group Stimulus Space* is uniquely determined.

In a subsequent experimental listening session, subjects rated perceived LEV of the stimuli (five repetitions each for each of the five stimuli). The correlation between these ratings and the coordinates of the stimuli on the two INDSCAL-derived dimensions allow us to make the following interpretation: Dimension 2 corresponds to LEV, the greatest perceived LEV being associated with the room simulation having the left wall missing. The lowest ratings of perceived LEV were associated with the room with the right wall missing. The correlation between LEV ratings and coordinates of the stimuli on INDSCAL Dimension 1 was not significant. This is made clear by the angle of the dashed line segment in Fig. 10 that is labeled LEV. The dot-dashed line segment labeled MIS shows the correlation between the two INDSCAL-derived dimensions and the judgment proportions of the four listeners from the "Heard-Reflection" identification task. If a given stimulus was often given the response "Left Wall Heard" on that task, then the stimulus coordinate on INDSCAL Dimension 1 for that stimulus would likely be high; conversely, a stimulus for which "Right Wall Heard" was the dominant response then its stimulus coordinate would likely be high on INDSCAL Dimension 1. Taken together, these results reveal the underlying perceptual structure that makes it possible for listeners to distinguish between room reverberation simulations with various single walls removed from each simulation. The results also make it clear that simply removing a wall from the simulation does not create the direct perception of a "wall-missing" attribute; but rather a change in spatial imagery that is better understood in terms of perceptual attributes such as LEV.

4. Conclusion

The specific achievements of this study were the following: First, a software spatial sound solution for multichannel reverberation simulation was developed using simple filtering models designed to capture acoustical features of walls very efficiently. Second, a representative set of sound sources were processed by this efficient method, and several of the model's parameters were systematically varied (particularly, room shape, wall absorption characteristics, the location of a missing wall, and maximum order of simulated early reflections). Finally, a controlled listening experiment was executed in which listeners were asked to identify which wall was missing out of the six possible choices in a rectangular room. The results of a preliminary identification task reveal which conditions make it relatively more or less easy for listeners to determine which wall is missing simply from hearing the multichannel reverberation simulation. The results of a second identification task and a subsequent exploratory study of the differences between the spatial imagery associated with five different room simulations reveal something of the perceptual structure underlying human abilities to distinguish between rooms with various walls missing simply from hearing a multichannel reverberation simulation.

The sound stimuli used in these listening experiments were produced off-line using the Matlab software, and did not support user interaction within the virtual environment. Currently, we are developing our system in C++ and studying the real-time performance of our algorithm for workstations. The results of these listening experiment have contributed to our developing software spatial sound solution for multichannel reverberation reproduction. Of course, with respect to the system's ability to render distinctive spatial imagery for virtual acoustic environment with one or more walls are missing, the results have shown that listeners are not particularly good at determining which wall is missing. In such virtual acoustic environment, although our simulation included fewer image sources than will be encountered in actual enclosed rooms, it is unlikely that higher order reflection simulation would disambiguate the situation; rather, more reverberation might hurt the identification more than it might help.

Future research will include the following; using a software spatial sound solution in real-time, investigating imagery resulting when one wall is missing and the sound source is moving. Since some early reflection patterns associated with a spatially-stationary sound source are rather idiosyncratic, there may be more reliable information available given a virtual source moving through a virtual environment. Especially if the listener is allowed to interactively explore the virtual environment, the real-time simulation might enable determination of the location of missing walls.

Acknowledgment

The first author was visiting the Duesseldorf University of Applied Sciences while writing this article. The authors thanks Jens Herder of the Virtual Sets and Virtual Environments Laboratory for his support and fruitful discussions.

References

1. M. Kleiner "Auralization – an overview," **J.** **A***udio* **E***ng.* **S***oc.*, vol. 41, no. 11, pp. 861–874, 1993.

2. D. Griesinger, "The psychoacoustics of apparent source width, spaciousness and envelopment in performance spaces," *Acta Acustica*, vol. 83, pp. 721–731, 1997.

3. F. R. Moore, "A general model for spatial processing of sounds," *Computer Music Journal*, vol. 7, no. 3, pp. 6–15, 1983.

4. B. B. Bauer, "Stereophonic earphones and binaural loudspeakers," **J.** **A***udio* **E***ng.* **S***oc.*, vol. 9, pp. 148–151, 1961.

5. G. S. Kendall and W. L. Martens, "Simulating the cues of spatial hearing in natural environments," in *Int. Computer Music Conf.*, Paris, 1984, Computer Music Association, pp. 111–126.

6. W. L. Martens, "The impact of decorrelated low-frequency reproduction on auditory spatial imagery: Are two subwoofers better than one?," in *Proceedings of the Audio Engineering Society 16th Int. Conf. on Spatial Sound Reproduction*, Rovaniemi, Finland, 1999, pp. 67–77.

7. J. A. Moorer, "About this reverberation business," *Computer Music Journal*, vol. 3, pp. 13–28, 1979.

8. J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," **J.** **A***cous.* **S***oc.* **A***mer.*, vol. 65, no. 4, pp. 943–950, April 1979.

9. D. H. Cooper and J. L. Bauck, "Prospects for transaural recording," **J.** **A***udio* **E***ng.* **S***oc.*, vol. 37, pp. 29–40, 1989.

10. P. Damaske, "Head-related two-channel stereophony with loudspeaker reproduction," **J.** **A***cous.* **S***oc.* **A***mer.*, vol. 40, no. 4, pp. 1109–1115, 1971.

11. G. S. Kendall, W. L. Martens, Daniel J. Freed, M. Derek Ludwig, and Richard W. Karstens, "Image model reverberation from recirculating delays," *Proc. Audio Engineering Society 81st Conv.*, 1986, Preprint No. 2408 (C–17).

12. W S Torgerson, *Theory and methods of scaling*, Wiley, New York, 1958.

13. S. S. Shiffman, M. L. Reynolds, and F. W. Young, *Introduction of multidimensional scaling*, Academic press, 1981.

14. J. D. Carroll and J. J. Chang, "Analysis of individual differences in multidimensional scaling via an n-way generalization of "Eckart-Young" decomposition," *Psychometrika*, vol. 35, pp. 283–319, 1970.

15. L R Tucker and S J Messick, "An individual differences model for multidimensional scaling," *Psychometrika*, vol. 38, pp. 333–368, 1963.

16. SPSS Inc., Chicago, WA, *SPSS Base 10.0 User's Guide*, 1999.