# An integrated system for dynamic control of auditory perspective in a multichannel sound field

Jason Andrew Corey

Graduate Program in Sound Recording

Faculty of Music

McGill University, Montreal

July 2002

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements of the degree of Doctor of Philosophy.

National Library
of Canada

Acquisitions and
Bibliographic Services

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque nationale
du Canada

Acquisisitons et
services bibliographiques

395, rue Wellington
Ottawa ON K1A 0N4
Canada

*Your file Votre référence*
*ISBN: 0-612-85695-X*
*Our file Notre référence*
*ISBN: 0-612-85695-X*

# Abstract

An integrated system providing dynamic control of sound source azimuth, distance and proximity to a room boundary within a simulated acoustic space is proposed for use in multichannel music and film sound production. The system has been investigated, implemented, and psychoacoustically tested within the ITU-R BS.775 recommended five-channel (3/2) loudspeaker layout. The work brings together physical and perceptual models of room simulation to allow dynamic placement of virtual sound sources at any location of a simulated space within the horizontal plane.

The control system incorporates a number of modules including simulated room modes, "fuzzy" sources, and tracking early reflections, whose parameters are dynamically changed according to sound source location within the simulated space. The control functions of the basic elements, derived from theories of perception of a source in a real room, have been carefully tuned to provide efficient, effective, and intuitive control of a sound source's perceived location.

Seven formal listening tests were conducted to evaluate the effectiveness of the algorithm design choices. The tests evaluated:

- loudness calibration of multichannel sound images
- the effectiveness of distance control
- the resolution of distance control provided by the system
- the effectiveness of the proposed system when compared to a commercially available multichannel room simulation system in terms of control of source distance and proximity to a room boundary
- the role of tracking early reflection patterns on the perception of sound source distance
- the role of tracking early reflection patterns on the perception of lateral phantom images

The listening tests confirm the effectiveness of the system for control of perceived sound source distance, proximity to room boundaries, and azimuth, through fine, dynamic adjustment of parameters according to source location. All of the parameters are grouped and controlled together to create a perceptually strong impression of source location and movement within a simulated space.

# Résumé

Un système intégré permettant le contrôle dynamique de l'azimut et de la distance d'une source sonore dans un espace acoustique simulé, ainsi que de la proximité des surfaces délimitant cet espace, est présenté pour être utiliser dans la production de musique et bandes sonores de film multi-canaux. Le système proposé utilise une configuration de haut-parleurs à cinq canaux (3/2) obéissant à la norme ITU-R BS.775. La simulation d'un espace sonore combinant des modèles physiques et perceptifs permet de placer des sources sonores virtuelles à des endroits arbitraires dans le plan horizontal.

Le système de contrôle inclut des modules pour la simulation des modes de résonances d'une salle, la simulation de sources floues, et le suivi des premières réflexions dont les paramètres sont ajustés dynamiquement selon la position de la source sonore dans l'espace simulé.

Sept tests d'écoutes ont été effectués dans le but dévaluer l'efficacité des algorithmes choisis. Ces tests avaient pour but d'évaluer:

- le calibrage de la sonie des images sonores reproduites par un système multi-canaux
- l'efficacité du contrôle de la distance

- l'efficacité du système proposé quant au contrôle de la distance d'une source sonore et de sa proximité d'une surface délimitant l'espace sonore en comparaison à un systéme multi-canaux de simulation acoustique disponible sur le marché
- le rôle joué par le suivi des premières réflexions du son sur la perception de la distance d'une source sonore et des images latérales fantômes
- la résolution obtenue pour le contrôle de la distance

Les tests d'écoute ont confirmé la capacité du système à contrôler efficacement la distance et l'azimut d'une source sonore de même que la proximité des surfaces délimitant l'espace simulé, grâce à l'ajustement dynamique des paramètres selon la position de la source. Tous les paramètres sont groupés et contrôlés ensemble afin de créer une simulation convaincante de la position et du mouvement d'une source sonore dans un espace virtuel.

# Acknowledgements

This thesis could not have been written without the help of many people. First to my main supervisor Prof. Wieslaw Woszczyk, who, with his enthusiasm for multichannel audio, provided guidance and many stimulating discussions. My thesis co-supervisor, Prof. Daniel Levitin, who became involved more recently, has provided help with listening tests and statistics. Thanks to George Massenburg for providing feedback about the sound of the algorithm.

None of the work could have been completed without the generous support of the following people and organizations: National Sciences and Engineering Research Council of Canada (NSERC); Kim Rishøj, Morten Lave, Thomas Lund and T.C. Electronic; Dr. Søren Bech, Poul Præstgaard and Bang & Olufsen; and Dr. Takeo Yamamoto and Pioneer Electronic Corp.

Dr. Geoff Martin, with whom I shared the MARLab for some time is thanked for asking interesting questions about audio and recording and for a well-developed sense of humour. Dr. René Quesnel helped translate the abstract and helped in the initial stages of the development of the thesis. To the most recent members of MARLab, Steve

Bellamy and John Usher, thanks for putting up with the listening tests.

To my parents and their respective spouses for constant encouragement and support and to my grandparents for instilling a curiosity about "how things work" from a very young age. Finally thanks to Jenn for her support, encouragement, and understanding.

# Contributions of Authors

Two conference presentations have been integrated to form part of the thesis. Because these presentations are multi-author, a brief description of the contributions of the primary author to the papers is outlined below.

- Corey, J., Woszczyk, W., Quesnel, R., and Martin, G. (2001a). Enhancements of room simulation with dynamic cues related to source position. Presented at the *19th International Conference of the Audio Engineering Society on Surround Sound Techniques, Technology and Perception*, Schloss Elmau, Germany.

  The primary author is responsible for all of the main ideas in the paper, specifically:

  - invention and formulation of a method of implementation for the "fuzzy" sources as a method of perceptually modeling the boundary effect.
  - development of dynamic control functions of physical parameters such as level, equalization, and delay time of the various components, that are mapped in a novel way to sound source location

- implementation of the entire system in software was performed by the primary author
- design, implementation of test automation, analysis of data and coordination of test subjects for the listening tests was performed by the primary author
- all writing was done by the primary author

The second author supervised the research and the third and fourth authors offered consultation in the implementation and general discussion of multichannel reverberation and control.

- Corey, J., Woszczyk, W., Martin, G., and Quesnel, R. (2001b). An integrated multidimensional controller of auditory perspective in a multichannel sound field. Presented at the *111th Convention of the Audio Engineering Society*, Preprint 5417, New York.

This conference presentation includes a description of the room simulation system nearly identical to the previous publication with the same contributions of the primary author. The primary author also implemented the tracking reflection algorithm. The paper presented new listening tests and results. The primary author designed, implemented the test automation, analyzed the data, and coordinated the test subjects for the listening tests. All writing was done by the primary author. Again the second author supervised the research and the third and fourth authors

offered consultation in the implementation and general discussion of multichannel reverberation and control.

# Table of Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

*If it sounds good, it is good.*

Duke Ellington

In the past few years there has been an increased interest in multichannel audio or surround sound among both audio professionals and the general public. Although the use of surround sound has been prominent in the film industry for many years (Hull, 1999), the use of surround sound for music applications has been limited until very recently. With the introduction of DVD and Super Audio CD (SACD) comes the possibility for widespread distribution of music and film productions in multichannel surround. Where traditional two-channel stereo once dominated the home music listening and movie-watching environments, discrete (as opposed to matrixed) multichannel surround is now becoming more common. Specifically 5.1 channels are typical, normally comprised of three front loudspeakers, two rear loudspeakers, and one low-frequency

effects (LFE) subwoofer channel. This configuration is also referred to as a 3/2 system with recommendations for loudspeaker placement by the International Telecommunications Union (ITU-R, 1994). The 1970's saw a rise and fall of so-called quadraphonic ("quad") configurations where a four loudspeaker system was promoted for the home listening room. The two extra channels were derived from different types of matrixing techniques (Hull, 1999).

With the current storage and reproduction media (e.g., DVD and SACD) it becomes possible for music and film producers to make use of five independent reproduction channels. As such there is more freedom for control over factors such as sound image locations and movement, the simulation of real acoustic spaces, sound image size and depth. For this reason, new and more advanced tools designed for manipulating the spatial dimension in a multichannel sound field need to be developed to be used by music and film soundtrack creators to take full advantage to the opportunities available in the five-channel format. Five independent (or discrete) channels of audio provide the possibility to position recorded sounds in many locations around the listener, and spatially separate the sound sources in ways not possible with stereo or matrixed surround. As Theile (2000) states, the new 3/2 format offers improved imaging characteristics such as envelopment, spatial impression, direction and depth, over traditional 2-channel (2/0) stereo. With spatial separation of sound sources there can be less masking of one source with another (Kidd et al., 1998; Saberi et al., 1991), therefore making the auditory system more sensitive to individual components of the sound field such as:

- frequency balance of the individual instruments
- spatial image size and spread
- subtleties in musical expression
- dynamic modulation, whether artificial or not
- recorded acoustic space or artificial reverberation and the imperfections therein

The main point here is that with multichannel reproduction, aspects of the sound field that were once hidden in two-channel stereo now become apparent. In addition, the production of music and film soundtracks is much more complex for multichannel reproduction than it is for two-channel reproduction. The number and complexity of parameter changes necessary to manipulate multichannel auditory perspective grows exponentially with the number of channels used, which require sophisticated mapping of control functions to take advantage of the reproduction format.

Besides the surround channels which provide the possibility for the listener to be fully immersed in the recording, the centre channel provides the possibility to have a more stable central sound image that is not as sensitive to listener location (Boone et al., 1995).

Music and film producers have used artificial reverberation for many years to add a sense of depth and spaciousness to recordings. This is especially true for recordings made in relatively dry acoustic settings where reverberation is added to recorded sound sources during the mixing process. Historically artificial reverberation units have provided users

with control over numerous physical parameters of the simulated reverberation such as reverberation decay time, pre-delay time, reverberation level, frequency and spatial modulation, and filters, to name a few.

The early use of reverberation employed echo chambers into which recorded sound was fed, with the returning signal from the chamber being mixed into the original recording (Blesser, 2001). Further developments led to the plate reverberation unit and eventually to the digital reverberation unit, initially developed by Schroeder (1962).

## 1.1 Goals and rationale

The goal of the thesis is to propose a new system or algorithm for control of sound source location within a multichannel reproduction environment where panoramic control (panning) of a sound source and adjustment of simulated room acoustics would fall under the same control. Most panning designs are based on a constant-power algorithm and do not provide the necessary control of auditory perspective required for multichannel reproduction. Multichannel panning through the manipulation of amplitude and phase, based on the mathematical principles of spherical harmonics has been developed by Gerzon (1992a,b), but thus far have found limited practical applications in audio engineering.

The impetus to develop this system is to improve the control of auditory perspective and sound source distance in current multichannel reverberation systems. Auditory space perception is considered to have two main aspects (Mershon and King, 1975): the

direction of the sound source and the distance of the sound source from the listener. Taking it one step further, the description would also include the proximity to room boundaries. While multichannel reverberation systems are quite capable of providing control over the perceived angular location of a sound source in the horizontal plane (or azimuth), they fall short in the area of providing efficient control over the perceived distance of a sound source. Up until now control of source location involved only the manipulation of source azimuth without consideration for its distance or proximity to a room boundary. If one imagines the location of a real source in an acoustic enclosure, both the azimuth and distance are considered in determining its location. The positioning of sound sources within the multichannel environment should be no different. The control system provided to the user (e.g., recording engineer) should allow efficient and intuitive control of sound source distance and azimuth within the same control algorithm.

Most current artificial reverberation devices consider room size and dimensions in the synthesis of room acoustics, and yet none allow the user to position a source near a room boundary within the simulated space. There are obviously room boundaries being considered if there is an early reflection pattern and reverberation. Recording engineers know from experience the effect of placing a musician or other sound source near a room boundary, and how that will create a different impression than the same sound source placed in the middle of the room. Formal listening tests have been conducted showing the influence of room boundaries on the placement of loudspeakers in a listening room. For instance Olive et al. (1994) found that loudspeaker location within a listening room

had a larger effect on listener preference than loudspeaker type. Bech (1994) also found that loudspeaker position in a room had a significant effect on perceived fidelity of the timbre of reproduced sound. This underlines the importance of modelling the perception of acoustic changes according to changes in source location.

The proposed auditory perspective control algorithm will provide user controlled sound source azimuth and distance within a given simulated room, through dynamic, coordinated adjustment of numerous physical parameters and perceptual models. The physical parameters that are manipulated include: level, equalization, delay time of the direct and indirect sound as well as reverberation level. A simplified model of room modes and a perceptual model of the boundary effect are also dynamically updated according to source location and help to strengthen the perceptual effect of the control algorithm. The proposed system can be integrated with any multichannel reverberation unit.

It is indeed possible to simulate a sound source changing distance with current artificial reverberation technology through manual adjustment of literally hundreds of physical parameters in a coordinated way. It requires that the reverberation unit be automated over time (and synchronized with timecode to the recording), but the time consumed by the engineer to program the proper adjustments would be enormous, especially with multichannel reproduction. The idea behind the current work is to have the adjustment of physical parameters happen in an automatic and effective way when the user moves a dot in the graphic interface, representing sound source position, to a new location. By conducting listening tests, it is possible to determine if the control algorithm

provides effective control of perceptual aspects of the sound image. Woszczyk (1993) also states that more sophisticated methods of control of individual signals is needed in order to take advantage of all the imaging possibilities allowed by multichannel reproduction.

It is important for the room simulation and source panning algorithms to have some basis in our perception of sound in the multichannel environment. While it may be possible to compute the behaviour of sound in a room for the duration of the decay time, this involves very complex mathematics to accurately model the details of a specific room. Simply mapping all of the surfaces (both shape and texture) in a real concert hall would be a monumental task. Smith (2002) suggests that accurately to model 1 second of reverberation decay in a typical room, one would need "at least 10 Pentium CPU's clocked at 3 GigaHertz, assuming they were doing nothing else, and assuming both the multiply and addition can be initiated each clock cycle, with no wait-states caused by the three required memory accesses (input, output, and filter coefficient)." Only in this way would we be sure that all of the important perceptual properties of the room acoustics would be modelled. By simplifying the physical model we are not sure what perceptual information is lost. By attempting to simulate the perception of sound in a room, the physical model is not accurate but the perception may be closer to that of a real space.

It may be considered that the creation of psychoacoustic illusions is more efficient in terms of signal processing than physical models and at the same time more convincing to the listener. Because of the complexity of modelling the acoustics of a real space, a perceptual or psychoacoustical model may result in a more realistic and less fatiguing

result. "It is generally true in creating psychoacoustic illusions that the reliability of the illusion, especially under adverse technical or listening conditions, is increased if the maximum number of auditory cues for the illusion is supplied, with a side-effect of low listening fatigue since the ears and brain are having to do less work to decide what is going on." (Gerzon, 1992a, p. 2)

The current generation of artificial reverberation algorithms is based on a simplified model of the behaviour of sound in a room or hall. The development of this reverberation generally does not consider the process of human perception. Because the model is simplified, we may conclude that perceptually important features of the sound field are unintentionally omitted. As such it becomes possible to easily distinguish between sound recordings using artificial reverberation and those using a natural acoustic venue. There are occasions when it is possible to create the illusion of a musical instrument recorded in a real acoustic space through the use of many layers of artificial reverberation mixed with temporally complex musical material in a stereo recording. Complex musical material may help mask the perceptual imperfections of the reverberation.

A second main reason to emphasize the perceptually most relevant features of multichannel room simulation is to help overcome some of the inherent limitations of the ITU-R BS.775 five-channel loudspeaker layout (ITU-R, 1994), such as the wide spacing between the front and rear loudspeakers. Although it is not flawless, it has become a standard and as such methods of signal processing and control need to be developed that work to improve the quality of spatial imaging characteristics.

## 1.2 Multichannel versus 3-D audio

Much research has been conducted on the testing and development of 3D audio systems which use headphones to deliver sound to listeners. (A good overview of 3-D audio can be found in Begault (2000)). The system described in this thesis involves the use of loudspeakers surrounding the listener as opposed to presentation over headphones. In 3-D or binaural audio systems the left and right ear signals are normally filtered with what are known as head-related transfer functions or HRTF's to simulate the natural spectral filtering that occurs to sounds due to interference with the head, shoulders, and outer ear, when listening to real sounds in a natural environment.

When identical signals are presented to both ears over headphones, there will be intracranial or in-head localization of the sound sources (Blauert, 1997). That is, the auditory event will occur inside the head. This phenomenon is rarely present in normal, everyday hearing mostly due to head and pinna filtering. Filtering of the headphone signals using HRTF's helps improve out-of-head localization. There are also transaural audio systems that present 3-D audio over stereo loudspeakers usually with the use of cross-talk cancellation and head-tracking devices. (A good reference for transaural audio is Gardner (1997)). Some of the problems associated with 3-D audio include (Hafter, 2001):

1. The sampling of the HRTFs is limited to a finite number of positions around the listener with interpolation required to position sounds between the sampled locations.

2. Usually 3-D audio uses sampled HRTFs from other people (or a mannequin such as KEMAR (Gardner and Martin, 1995)) which is not as accurate as using our own ears. Wenzel et al. (1993) found that the use of non-individualized HRTF's results in front-back and up-down confusions, due to distortion of spectral cues. Toole (1991) has suggested that although it is possible to present a very convincing auditory environment over headphones with the use of HRTFs, there are still front-back reversals.

3. The HRTF's need to change dynamically with head movements to provide a sense of realism

In multichannel audio there is no need for artificial HRTF filtering or head-tracking, it is all automatic. It is very possible to create a realistic sounding presentation over loudspeakers, where the more loudspeakers there is, the easier it is to make it sound realistic. In addition to this listeners can sit together to enjoy the same presentation in multichannel audio, whereas under headphones each person has his own isolated presentation. Multichannel is the next logical extension of traditional two-channel stereo.

## 1.3 Structure of the thesis

The thesis is structured as follows:

1. a review of literature related to the main areas of research is presented

2. a preliminary study of front and rear stereo

3. a description of the proposed system of control

4. psychoacoustic evaluation of the system to test control of perceived source distance, lateral placement of sound source images, and overall functionality of the system.

The dissertation will apply the known theories of distance perception to the control of auditory distance perception in a multichannel environment with simulated acoustics. The design of the system described here is supported by formal and informal listening tests.

# Chapter 2

# Review of Literature

The design and implementation of the system described in this thesis draws on a broad area of research. The current chapter will review some of the main areas related to the following three main areas of research:

- perception of sound source distance and azimuth
- principles of room simulation and artificial reverberation
- perception of sound in 3/2 multichannel reproduction

A great deal of research has been conducted concerning the physical aspect of room simulation and modelling. Similarly there is a large body of scientific writing documenting the results of experiments in the perception of sound source distance and location in a real environment. This chapter will outline a review of literature relevant to room acoustics modelling from a physical point of view and also the theories of distance perception

in a real room. Since the proposed system will function in 3/2 multichannel reproduction, a review of studies concerning the generation and perception of sound within such an environment becomes crucial. Being an artificial environment it is limited with respect to reality. Nonetheless it is important to understand auditory perception in a real environment, such that the same principles can be applied in the synthetic acoustic environment.

## 2.1 Perception of sound source distance and azimuth

### 2.1.1 Localization of sound sources

It is necessary to understand the nature of how real sounds in real environments are perceived in the development of a room simulation system. From the literature we know that the localization of real sources is determined by three main cues (Moore, 1997):

- interaural time differences (ITD)
- interaural level differences (ILD)
- direction-dependent filtering by the head and pinna, giving what is known as the head-related transfer function (HRTF)

ITD's can range from 0 s for sounds directly in front or behind a listener to about 690 $\mu$s for lateral sounds at 90°. The time difference is derived from the time it takes for sound to propagate from one ear to the other ear. For the localization of sinusoids,

ILD's are used for high frequencies and ITD's are used for low frequencies, and this is known as the "duplex theory" of sound localization (Rayleigh, 1907).

Sound sources located to the side of a listener in a position slightly ahead or behind, provide ambiguous interaural differences. This area is known as the cone of confusion. Listeners can resolve location judgments for sounds in the cone of confusion by moving their heads. Pinna and head filtering provide information to help discriminate between front and back and to help resolve vertical judgments (Butler, 1969). Blauert (1997) indicated that there is a greater localization blur for sources located laterally. Figure 2.1 illustrates the localization blur for four different locations around a listener.

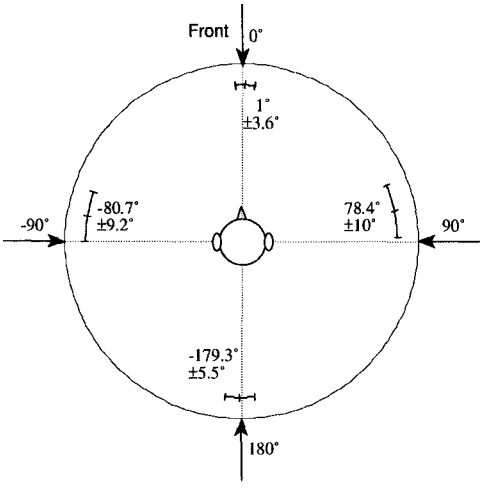


Figure 2.1: Localization blur and localization in the horizontal plane for loudspeakers emitting white-noise pulses of 100ms duration at $\phi = 0°, \pm 90°$, and 180°. Arrows indicate the location of the sources. (Diagram adapted from Blauert (1997), p. 41 , after Preibisch-Effenberger (1966) and Haustein and Schirmer (1970))

### 2.1.2 Spatial impression

Apparent or auditory source width (ASW) is used to describe the perceived size of an auditory image (regardless of the visual width) and has been found to be influenced by the amount of early reflections present in the first 80 ms of the impulse response (Rumsey, 2001). It has mostly been studied by concert hall acousticians and it is related to the interaural cross-correlation (IACC), specifically measured by 1-IACC (Beranek, 1996). The IACC, simply put, is a measure of the similarity between the signals received at the two ears. Usually a larger ASW is preferred in concert halls. As Rumsey (2001) states it is unclear how relevant ASW is for reproduced sound, since it usually depends on the aesthetics of the style of music being produced for loudspeaker reproduction.

ASW and IACC are also related to the sense of spaciousness in a concert hall. A second factor related to spaciousness is listener envelopment (LEV), which is judged to be highest when the reverberant sound arrives at a listener equally from all directions. In this case the reverberant sound is defined as the sound arriving 80 ms or more after the arrival of the direct sound (Beranek, 1996). In summary apparent source width is defined by early IACC, where listener envelopment is defined by late IACC

### 2.1.3 Distance perception in a real environment

Many researchers of auditory perception have attempted to understand how we determine the distance of a sound source in a natural environment (Coleman, 1962, 1963, 1968; Martens, 2001; Mershon and King, 1975; Mershon and Bowers, 1979; Mershon

et al., 1989; Nielsen, 1991, 1992; Zahoric, 1998). It is known that multiple acoustic parameters, or cues, co-vary with physical sound source distance in nearly all naturally occurring environments. Auditory perception scientists are interested in how listeners combine the information of these multiple cues to form a unitary distance percept. (Zahoric, 1998, p. 2)

Mershon and King (1975) indicate that the perception of distance of a sound source is dependent on two main factors:

- intensity of the sound source
- level of reverberation

Butler et al. (1980) state that the spectral balance of the sound source also contributes to the perception of distance. Sounds with attenuated high frequencies are perceived as being more distant than sounds with attenuated low frequency bands. This cue would be dependent on the listeners knowledge of the source. As a sound source moves away from a listener, high frequencies are naturally absorbed by air.

Nielsen (1991, 1992) states that a number of factors change as a sound source changes distance from a listener:

- loudness
- perceived frequency spectrum
- direct-to-reverberant ratio
- binaural differences

The important point is to determine which factors are most important in influencing the perception of distance. It has been found that listeners cannot judge the distance of loudspeakers placed at different distances from the listener, under anechoic conditions when the loudness of the loudspeakers remains constant (Nielsen, 1992; Coleman, 1962). Sheeline (1982) found that a 10 to 21 dB change in intensity was needed to change the perceived distance of a sound source by a factor of 2 under anechoic conditions. Listeners are more capable of judging distance when loudspeakers are placed at different distances in a reverberant environment. From this it can be concluded that early reflection patterns and the direct-to-reverberant ratio are both important conditions for the determination of distance of a sound source. Kendall and Martens (1984) also state that early reflections provide sufficient information for determining the distance of a sound source in a typical room.

In a free-field sound decays by 6dB for every doubling of distance which is known as the inverse square law. In acoustic enclosures, a change in direct-to-reverberant energy ratio is primarily due to the effect of the inverse square law on the direct (first arriving) portion of the sound field, since the energy in the later arriving reflected portion of the sound field is relatively constant for varying source distance (Chowning (1971); Moore (1983); Blauert (1997), cited in Zahoric (1998)).

Studies in distance perception have indicated a phenomenon known as the 'auditory horizon' effect (Begault, 2000; Bronkhorst and Houtgast, 1999). The idea is that there is a limit to the perceived distance of a sound source in a virtual or simulated environment even if the direct-to-reverberant ratio is decreased. When a sound source is moved more

than 2m away from a listener the perceived location increases much slower than the actual location. What this may mean for room simulation algorithms is that the effects need to be exaggerated to provide a perceptually robust control of sound source distance.

### 2.1.4 Craven hypothesis of sound source distance

Gerzon (1992a) outlines what he refers to as the Craven hypothesis, a theory of sound source distance perception in a room that was originally formulated by Peter Craven (unpublished other than in Gerzon (1992a)). The hypothesis states that the apparent distance of a sound source is derived by the ears according to the gain and time of arrival of early reflections arriving at the listener, relative to the gain and time of arrival of the direct sound. The hypothesis assumes the sources have omnidirectional radiation characteristics and the walls are perfect reflectors.

A conventional approach used by recording engineers to change the perceived distance of a recorded sound source is to use what is known as an "auxiliary send" to route a portion of the dry sound to a reverberation unit, and return the output of the reverberation unit to the overall mix. The level of the auxiliary send determines the level of reverberation on the dry signal, and it is intended that different reverb levels will simulate distance. This approach to changing apparent sound source distance does not satisfy the Craven hypothesis because only the overall gain of the reflections is changed. The gain and time of arrival of individual reflections remains constant, relative to the other reflections.

Despite this, Begault (1991) found when testing the perceptual effects of synthetic reverberation on 3-D audio systems in a listening test, that all subjects made relative increases in their distance judgements when reverberation was added to the stimuli. Begault (1987) investigated loudness-equalized speech sources and found that a virtual sound source placed at 2 metres was judged by listeners to be closer in a large modelled enclosure than in a small modelled enclosure. The physical difference between the two conditions was that the reflections were spread out over a greater time period and were lower in level relative to the direct sound, than in the small enclosure. We may deduce from this that the direct-to-indirect energy ratio is an important cue for distance perception in a synthetic acoustic environment.

The theories of distance perception are applied to sound source distance control in a simulated sound field, which will be described in Chapter 4.

## 2.2 Principles of room simulation

The acoustics of a performance space has a direct impact on the perception of the musical event for both the performers and the audience. A superior acoustic space can enhance the musical experience in some ways without detracting from it in others (Beranek, 1996). The same is true for artificial reverberation and room simulation. Even in everyday life, when a person talking walks across a room, the perception of the room effect changes for the listener. This is important to model for film soundtracks, especially when the dialogue is overdubbed after the film shooting. Often the dialogue

recorded during a film shooting is not used due to the high level of background noise present during the shooting. Therefore actors usually re-record the dialogue in a quiet recording studio. Unfortunately the original ambience and room effect mixed with the dialogue from the location sound is lost by recording the dialogue in studio. Any changes in perceived room effect due to movement of the actor across the set will need to be recreated artificially when dialogue is re-recorded.

Normally in room simulation the energy decay is broken down into two main parts representing the early and the late parts of the impulse response. The early part of reverberation decay is characterized by discrete echoes of the original sound source arriving typically in the first 80 ms after the direct sound. The late part, arriving more than 80 ms after the direct sound, is characterized by being diffuse. A typical simplified representation of the impulse response of a concert hall is depicted in Figure 2.2, illustrating the direct sound followed by early reflections, and then the decay of the reverberation.

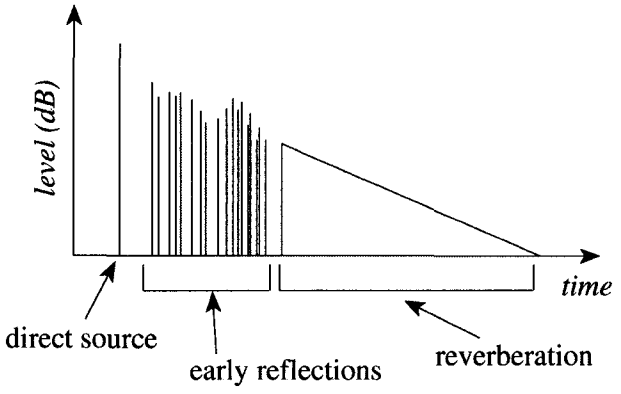


Figure 2.2: A typical representation of an impulse response of a concert hall, adapted from Rumsey (2001).

### 2.2.1 Modelling early reflections

As we know from the pioneering work of Haas (1972), a single echo of the original sound arriving less than about 30 ms after the arrival of the direct sound is perceptually disregarded by the auditory system in terms of direct sound localization. The echo will change the timbre of the direct sound, but even if its angle of incidence is different than the direct sound, it will not change the perception of azimuth of the direct sound. Early reflections in rooms are essentially a series of discrete echoes arriving after the direct sound, and the Haas-effect becomes slightly more complex in a real room. Early reflections contribute to source distance (as noted above) and also to the impression of room size.

Due to the "precedence effect" (Blauert, 1997), when listeners are presented with a direct source and a pattern of reflections, localization will be determined by the direct source location since it is the first arriving wavefront. When presented with two loudspeakers producing identical clicks, with an inter-loudspeaker delay of e.g., 5 ms, listeners will perceive the location of the click to be in the non-delayed loudspeaker. In an interesting experiment Clifton (1987) found that after switching the delayed and non-delayed signals between the loudspeakers, listeners were able hear two separate clicks for a brief period of time after which localization is determined by the leading signal. This effect is known as the "Clifton effect" and may have implications for dynamically changing reflection patterns where the relative delay time between reflections changes according to source location. It may make listeners more aware of the early reflection

patterns while a source is moving and less sensitive while the source is stationary.

Two of the typical ways of modelling early reflection patterns are:

- image source method (Allen and Berkley, 1979)
- ray tracing (Krokstadt et al., 1968)

Vorländer (1989) has also suggested an algorithm combining ray tracing and image source methods. Both methods are based on geometrical room acoustics.

In the image source method, the room is assumed to have flat walls producing specular reflections. Secondary sources (reflections or delayed versions of the direct source) that arrive after the direct source are specified by mirroring the original sound at the plane of a reflecting wall. Because there are multiple surfaces present in the room model, multiple-order mirror images can be calculated. The higher the order of a reflection, the more wall reflections it has undergone. An illustration of the image model is found in Figure 2.3. Rectangular rooms are the easiest to model and more complex geometric room shapes are more difficult (Borish (1984) cited in Lehnert and Blauert (1992b)). One drawback of the image model is that the computation time increases exponentially with length, but the achieved time resolution is high (Vorländer, 1989).

The ray tracing method is better for modelling sound scattering as it occurs when sound is reflected from rough surfaces (Vorländer, 1989). The idea of ray tracing can be thought of as an omnidirectional source radiating sound in the form of particles, where the path of each particle is traced, including the reflection off walls which represents some filtering and absorption of the particle. Figure 2.4 represents tracing one ray from

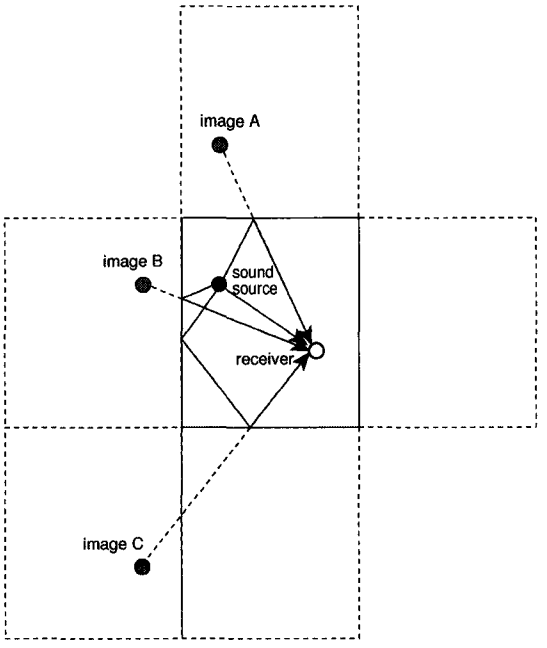


Figure 2.3: Image source diagram where Images A and B represent first-order reflections, and Image C represents a second order reflection.

source to receiver. The computational time increases proportional to the length of the impulse response, but the temporal resolution is limited (Vorländer, 1989).

The ray tracing and image models assume that early reflections behave in a manner similar to light, i.e., in a specular manner. In actuality a waveform will scatter energy in many directions in a diffuse manner especially at low frequencies (Begault, 1992). As such, diffusion modelling is important for a truer model of the behaviour of sound in a real room. Martin (2001) has suggested a model for the diffusion of early reflections.

Normally in artificial reverberation devices the early reflection patterns do not change if the sound source is panned to a different location, but as was found in the development of the system described herein, changing reflection patterns according to source location is very important for a more realistic perception of source distance.

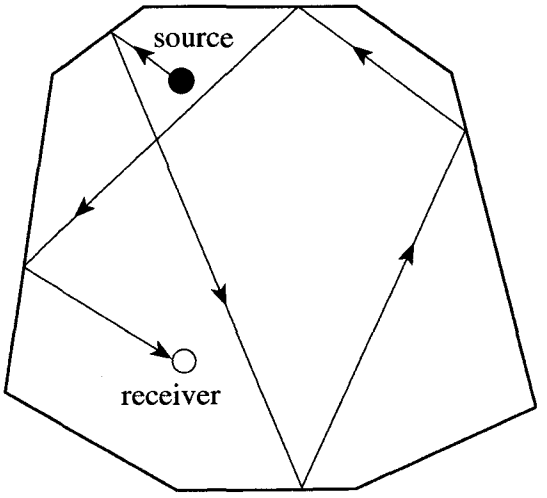


Figure 2.4: Tracing a ray from a source to a receiver.

### 2.2.2 Modelling reverberation

Most digital reverberation devices today involve the use of complex arrangements of Schroeder's original reverberation algorithm. This, together with complex signal processing and tuning by ear with a musical sensibility and a knowledge of psychoacoustics, provides the basis for most current reverberation algorithms (Pellegrini, 2001b). Although this method is not as accurate as convolution with a measured impulse, it is more efficient and provides a higher level of dynamic, parametric control of the synthetic reverberation (Jot, 1997).

Typically in artificial reverberation devices, the user is provided with control over such parameters as: reverberation decay time, pre-delay time, reverberation level, reflections level, and equalization. The reverberation decay time is defined as the amount of time in seconds that is required for 60 dB of sound decay after a sound is turned off. The pre-delay time is the amount of time (usually in milliseconds) that the artificial

reverberation is delayed relative to the direct sound, typically it ranges from 0 to about 100 ms. The reverberation level is the amplitude of the reverberation and essentially controls the direct-to-reverberant sound level. The level of the early reflections refers to the gain applied to the amplitude of the reflections ranging from 0 to unity gain. Equalization can be applied to the reverberation and early reflection patterns through control of such parameters as "High Soften," "Lo Damp," "Hi Cut," "Lo Cut." Usually these parameters translate into control of the level of high and low frequencies, with user-specified cut-off frequencies. There is usually also control over decay time multipliers for high and low frequency bands, relative to the main, mid-frequency decay time.

In this paradigm, the user needs to translate the given physical parameters into meaningful control over the perceived room simulation. There is an isomorphism that exists between the physical parameters and the desired perceptual effect. For instance to increase the perceived distance of a sound source, the user would need to adjust numerous physical parameters such as reverberation level, frequency equalization, reverberation time, and early reflection patterns. There is a translation that needs to take place between the perception of the sound field and the physical parameters that need to be adjusted to provide the intended auditory perspective. The situation is complex for stereo reproduction and becomes much more so with multichannel reproduction, due mainly to increased numbers of loudspeakers and the complex interrelationships between them. Because multichannel reproduction offers the possibility for increased listener envelopment and increased realism of room simulation over traditional stereo, there are many more physical parameters that need to be properly controlled to take

full advantage of the new sound reproduction paradigm.

The first digital reverberation algorithms were designed by Schroeder (1962) and Moorer (1979), and consisted of comb filters in parallel feeding allpass filters in series. It provided an exponentially decaying reverberant tail with the allpass filters and a very simplified model of the early reflections with the use of comb filters. An illustration of the original Schroeder reverberator can be found in Figure 2.5.

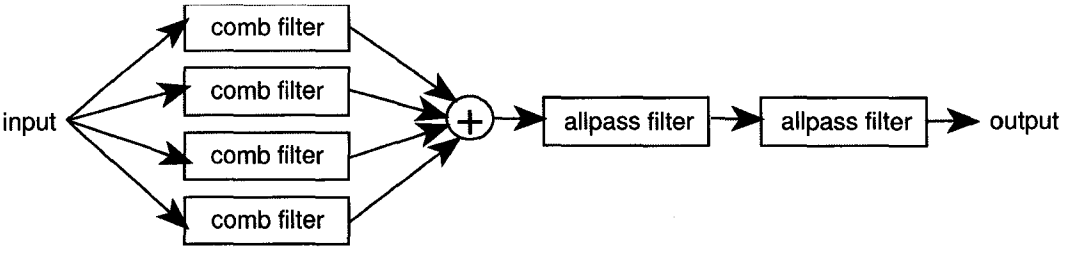


Figure 2.5: Block diagram of a Schroeder reverberator.

An allpass filter (shown in Figure 2.6) is an infinite impulse response (IIR) filter providing an exponentially decaying output for a given impulse. The frequency response of such a filter is flat but the phase response is altered.

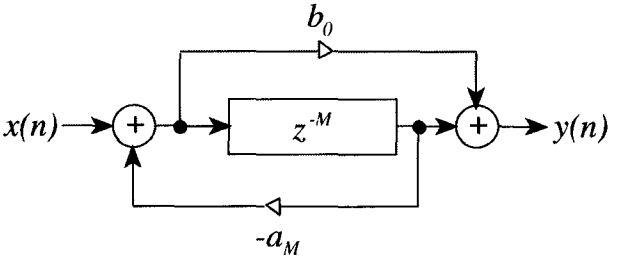


Figure 2.6: Block diagram of an allpass filter by combining a feed-forward and feedback comb filter where $b_0 = a_M$. (Smith, 2002)

A block diagram of a feedforward comb filter is shown in Figure 2.7.

A block diagram of a feedback comb filter is shown in Figure 2.8.

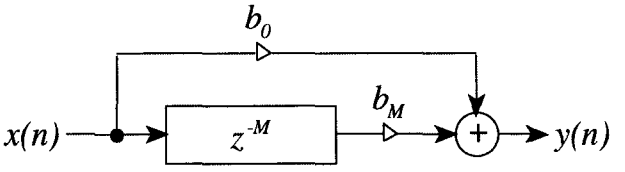


Figure 2.7: Block diagram of a feedforward comb filter where $b_0$ = feedforward coefficient, $b_M$ = delay output coefficient, and $M$ = delay-line length in samples. (Smith, 2002)

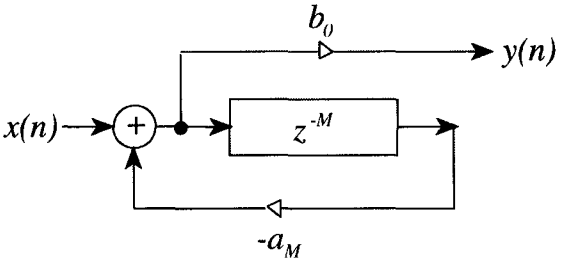


Figure 2.8: Block diagram of a feedback comb filter where $a_M$ = feedforward coefficient (need $|a_M| < 1$ for stability) and $M$ = delay-line length in samples. (Smith, 2002)

### Digital waveguide mesh

Digital waveguides have been used to model musical instruments with some success. The term digital waveguide refers to a bi-directional delay line (Smith, 1992; Murphy, 2001), whose fundamental structure plays an important role in models of musical resonators. A complex mesh composed of digital waveguides can be used to model three-dimensional resonances of acoustic spaces. A couple of major problems associated with this technique of room simulation include:

- requirement of large amounts of processing power, proportional to the update frequency of the mesh
- dispersion error (causing distortion of the audio signal)

The propagation speed of waves through a square mesh can be frequency dependent, causing dispersion error. This can be partly solved by changing the shape to a triangular mesh (Savioja et al., 1995). With these two major problems associated with it, there does not seem to be any real gain from using this method to model acoustic spaces with current processor technology.

### Convolution-based reverberation (sampled acoustics)

Newer commercially available digital signal processing units are powerful enough to provide real-time convolution (such as the Sony DRE-S777 sampling reverb). Impulse responses of real concert halls and other acoustic spaces are convolved with music recordings to create the impression that the music was actually performed in the space from which the impulse response was taken. This method, although providing a very high quality and realistic sounding reverberation, has a number of drawbacks. The main drawback is the inability to alter the reverberation characteristics easily. The impulse response from each acoustic space is recorded from a specific location. As such it is difficult to simulate a sound source in another location. In theory one would need to have an infinite number of impulse responses and have the ability to interpolate between them to move the source to different locations within the space. Often the recording engineer wants to change parameters of the reverb such as the decay time, early reflection patterns, decay time multipliers for different frequency bands, and predelay. Usually the only parameter available to change in convolution-based reverberation is the decay time, which simply changes the envelope of the reverberation decay. The early reflection

patterns and the overall character of the reverberation remain the same.

### 2.2.3 Perceptual control of reverberation

Recently some researchers have begun working on reverberation algorithms which are based on perceptual factors of acoustic spaces rather than constructing a pure physical model. This has the benefit of reducing processor requirements and at the same time increasing the perceived complexity of the simulation.

Pellegrini (2000, 2001a,b) has been developing so-called auditory virtual environments (AVE's) with the intention of having the design be perceptually-based as opposed to physically-based. Perceptually-based AVE's are designed to model the perceptually most important features of a simulated room. A physically-based AVE would be based solely on the physical behaviour of sound in a room, with no consideration of how perceptually relevant the components are. In the generation of early reflection patterns Pellegrini considers them as having two distinct parts: those used to simulate source distance and those used to give the impression of room size. The number of reflections used in distance control is four, and six reflections are used for room size perception, making a total of 10 reflections. For the source distance reflections, he chose to make the direction of incidence approximately $\pm 60°$ with respect to the direct sound. As we know from sound behaviour in a real room, early reflections arrive from all directions, not just $\pm 60°$. Therefore it is important to simulate reflections arriving from all directions surrounding a listener. The author has found, through informal experiments conducted

in the Multichannel Audio Research Laboratory comparing different numbers of early reflections, that it is important to have more than 10 reflections in a simulated space to give a convincing impression of source distance and room size.

Jot has designed and implemented a perceptually-based reverberation algorithm for use in MAX/MSP called Spat~, proposing efficient reverberation and distance rendering algorithms with the use of feedback delay networks (FDN's) (Jot and Chaigne, 1991; Jot and Warusfel, 1995; Jot, 1997). The feedback delay network, illustrated in Figure 2.9 provides a stochastic model of late reverberant decay. Stautner and Puckette (1982) claim that FDN's produce a less annoying colouration of the sound output than an equivalent number of comb filters. One problem with FDN's is that the energy builds up far more quickly than it does in a typical acoustic space.

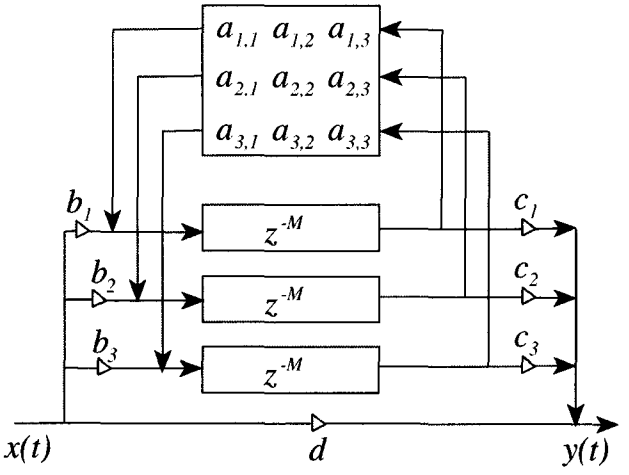


Figure 2.9: Block diagram of a feedback delay network. (Jot and Chaigne, 1991; Smith, 2002)

The Spat~ graphical interface provides control over nine perceptual factors of the room and source (Jot and Warusfel, 1995). There are three perceptual factors related

to the room effect:

- late reverberance
- heaviness
- liveness

The other six factors describe effects which depend on the position, directivity and orientation of the source:

- source presence
- brilliance
- warmth
- room presence
- running reverberance
- envelopment

This was one of the first implementations of multichannel reverberation with the idea of having perceptual controllers. It appears that most of the perceptual controllers are related to control of spectral filters and direct-to-reverberant ratios. Because this algorithm has been designed to operate without using much processing power, it is clearly audible as an artificial reverberation effect.

## 2.3 Perception of sound in 3/2 multichannel reproduction

A number of studies have been conducted regarding the perception of 3/2 multichannel reproduction (Figure 2.10). It is important to review some of this work since the perception of sound produced over loudspeakers in a multichannel configuration differs greatly from that in a real environment.

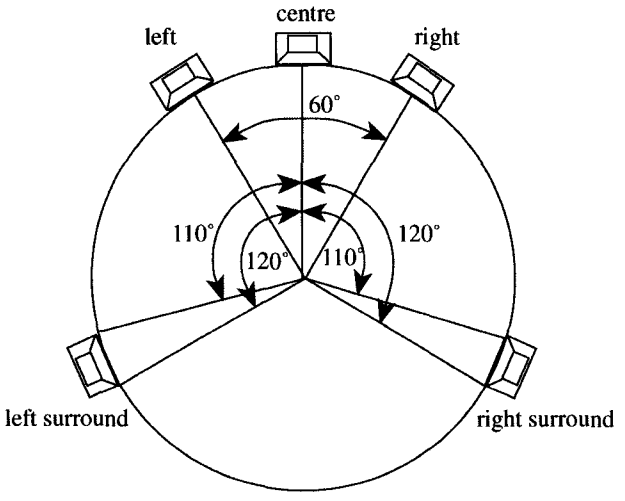


Figure 2.10: Diagram of the ITU-R BS.775-1 recommendation for a reference 3/2 loudspeaker arrangement (ITU-R, 1994).

Going back to stereo (2/0) reproduction for a moment, Kurozumi and Ohgushi (1983) found that an interchannel correlation of $r = 0$ produced the widest perceived sound image and $|r| = 1$ produced the least wide images. In other words a monophonic image, where the two loudspeaker signals are identical, would have a correlation of $r = 1$. With a low correlation, the image was perceived to spread the width of the loudspeaker spacing ($\pm 30°$). In a study of the effects of multichannel loudspeaker placement on increasing

the subjective diffuseness of incoherent signals, Damaske and Ando (1972) found that a 3-channel system is as diffuse as a system with 5, 6, or 7 channels, by comparing IACC of ear signals from a dummy head. However it is impossible with only 3 channels because complete diffuseness requires a good front/back balance. Wagener (1971) found that 5 incoherent channels were sufficient to produce optimal subjective diffuseness and that increasing the number of channels to 6 or 7 did not result in any improvement in subjective diffuseness. Subjective diffuseness is related to IACC, where a low IACC means a high degree of diffuseness.

Studies in sound localization in a four-channel reproduction system conducted by Ratliffe (1974), investigated the localization of phantom images generated by amplitude differences between pairs of loudspeakers. His conclusion was that it is generally difficult to pan sources smoothly from front to back along the side. Theile and Plenge (1977) studied lateral localization of phantom images and came to a similar conclusion.

One may ask, what do the rear loudspeakers contribute to the front loudspeakers? As Morimoto (1997) found through listening tests, the rear loudspeakers contribute to listener envelopment (LEV). The front loudspeaker array helps control apparent source width (ASW), but not LEV. By increasing the level of the rear loudspeakers, the LEV is also increased. This is assuming that the program material being reproduced is a recording of music in concert hall, where the ensemble is in front and the hall reverberation is mainly in the rear. Woszczyk (1993) writes that the surround channels should help to create a feeling of involvement for the listener without distracting too much from the events on the screen (in the case of audio-visual presentation).

In the assessment of multichannel sound systems, Miyasaka (1993) outlines possible subjective testing methodologies for investigating both impairments and quality. He suggests that attributes such as basic audio quality, front image quality and surround impression quality need to be assessed. Theile (1993) reports that the 3/2 loudspeaker arrangement offers enhanced directional stability and clarity of the frontal sound image and improved realism of auditory ambience. The addition of rear channels to the front three channels allows improved realism of auditory ambience.

Researchers have begun to develop methods of calibration of multichannel loudspeaker levels (Suokuisma et al., 1998; Zacharov et al., 1998; Bech and Zacharov, 1999; Zacharov and Bech, 2000). The calibration of levels in multichannel reproduction is much more critical than it is for two-channel stereo to ensure optimum reproduction of the intentions of the program maker. This series of papers reviewed the effects of signal type, loudspeakers placement, loudspeaker directivity, reproduction bandwidth on the calibration of levels. They also compared physical measures with subjective level calibration and found that the following signal/metric combinations provide the best prediction of subjective level calibration (Zacharov and Bech, 2000):

- constant specific loudness signal according to the Zwicker free field model
- Moore or Zwicker (diffuse or free field) loudness or B- or C-weighted SPL metrics
- highpass filtering of the test signal from 500 Hz

Woszczyk (1993), Mason et al. (2000), and Ford et al. (2002) presented the idea of having listeners represent spatial auditory impressions by graphical means. This method

has the advantage that the results avoid the potential ambiguity of language and listeners may feel it is a more intuitive method of representing their auditory perception. One difficulty with this method is in interpreting the graphical results especially when the graphical representation is done by hand and not on a computer.

Bech (1999) has suggested rigorous methods of subjective evaluation of the spatial characteristics of sound using descriptive analysis. In this case one of the goals is to develop a descriptive language that would allow listeners to meaningfully describe in words their impressions of spatial audio. The idea has been transposed from the food industry.

In the development of methods of subjective evaluation of quality of multichannel audio, Berg (2002) has researched relevant attribute scales originating from verbal elicitation of individual listeners' concepts of discrimination of different sounds. In this thorough study, he applied the Repertory Grid Technique for the first time in audio research. The results of the study indicate that for a list of attributes describing spatial audio with sufficient meaning, a group of experienced listeners can make significant judgments of spatial audio quality in multichannel systems.

# Chapter 3

# Preliminary study of front and rear stereo imaging

This chapter presents the results of informal listening and study comparing two channel stereo music in front versus behind a listener. Consideration of the quality of stereo program material from the rear of the listener is important in the exploration of music applications for multichannel reproduction systems to fully utilize the rear speakers. When producing music for multichannel reproduction, there is a need to be able to incorporate and integrate sound coming from the front with sound coming from the rear, to create a coherent sound image surrounding the listener. Part of this work lies in finding out how we perceive sound originating from the rear, and how this perception is different from front stereo. For instance when a mixing engineer is producing music for five-channel reproduction, will the optimal sound image be produced when a stereo image is positioned in the rear that was intended for the front? In other words does

the engineer need to treat the front stereo image differently than the rear image, for identical program material?

In the development of a system for control of auditory perspective in a five-channel environment, further examination of the differences between front and rear stereo needs to be undertaken. There has been a great deal of literature written on stereo as we hear it when the loudspeakers are in front of the listener, such as Blauert (1997) and Theile (1990, 2000). Additionally many have studied the subjective perception of multichannel audio reproduction, such as Nakayama et al. (1971); Damaske and Ando (1972); Wagener (1971) involving the use of four or five loudspeakers. Up to this point, there have not been many investigations of perception of stereo images produced by loudspeakers placed behind a listener. Because of the lack of literature on the topic, it is important to uncover and explore these specific differences to be able to fully exploit the rear stereo imaging capabilities of multichannel music.

The other main reason for this study was to become more familiar with multichannel reproduction to fully exploit the possibilities with the system described in the thesis. As the technical ear training course, required by sound recording students at McGill (Quesnel, 2001), trains listeners to hear and label different timbres, it may also be possible to train the auditory system to better and more quickly interpret the spatial dimension of audio specifically in multichannel reproduction. A great deal of the fine tuning of the proposed algorithm was done by ear, and a study of rear stereo helped to prepare for the task. This paper will present the findings of these preliminary informal listening experiments.

## 3.1 Method

A listening room was set-up with four Bang & Olufsen Beolab 4000 two-way active loudspeakers: two in front and two behind the listening position. The front speakers were positioned to create an equilateral triangle with the listener, with each speaker being placed 30° on either side of front center (0°), and the rear speakers were placed ±120° from front center 0°. The placement of the loudspeakers was chosen to be consistent with a set-up that would be used for 5-channel music reproduction as recommended by ITU-R BS.775-1 (ITU-R, 1994).

A number of commercially available compact discs were chosen, representing a variety of styles of music, to evaluate the differences between what we hear coming from the front and what we hear coming from the rear. Using computer software controlling a digital console, it was possible to switch easily between the front and rear pairs of loudspeakers.

For the comparison evaluations the focus was mainly on the following specific parameters:

- stereo image: consistency of placement of sounds in the sound stage
- depth and spaciousness
- frequency response

## 3.2 Results and discussion

### 3.2.1 Consistency of stereo image

In the examination of the stereo image, the focus was mainly on placement of sound images in the sound stage, and how consistent this placement was between front and rear. It was found that the stereo image in the rear was not as clear as that coming from the front. The rear image was less focused and more diffuse, exhibiting a real lack of clarity. It was possible to roughly localize sound sources in the rear but not with the same accuracy as in the front. Sounds that appeared to come from hard left or hard right in front stereo, were perceived to come from outside of the stereo image in the rear. For some music (e.g. pop music) with a strong center phantom image, the center image produced in-head localization when listening to the rear loudspeakers. This may be attributed to the fact that there is a wide spread between the loudspeakers, creating less cross-talk between ears and more head shadowing of the loudspeaker signals. This is similar to the effect of listening over headphones. One reason why the multichannel microphone technique using a dummy head developed by Klepko (1997) for the pick-up of surround channels is very effective is because it decreases in-head localization.

### 3.2.2 Depth and Spaciousness

The adjective 'depth' when used to describe music refers to the relative proximity of sounds in the stereo image. In a recording with a sense of depth, some sounds will appear to be closer than other sounds, giving a sense of distance between sounds at

the front of the sound stage and those at the rear. The question that was addressed in this evaluation is whether music originating from the rear has the same perceived depth as that originating from the front. It was found that there was less sense of depth in the rear stereo. Music coming from the back seems to be deflated and constricted, and lacking in depth.

Also related to depth is the sense of spaciousness in a recording. Rear stereo failed to communicate as much of a sense of spaciousness as did front stereo. The reverb was audible in the rear but it seemed to be less connected to the direct sound and served more to clutter the sound image. The reverb tails seems to have less detail.

### 3.2.3 Frequency response

The perceived frequency response of the rear stereo differed greatly from that of the front stereo. The front is much clearer and crisper, while the rear is muffled and dark. There are many frequencies that are affected, but it is clear that there is a roll-off of frequencies above 8 kHz. The sound coming from the rear was coloured to a large extent, most probably due to the effect of the pinnae. If we were to examine the frequency response of head-related transfer functions of sources behind a listener compared to those in front, the differences in spectral balance would be apparent.

## 3.3 Conclusions

In making evaluations about the overall impressions of front versus rear stereo a few things became apparent. The most obvious is that stereo coming from the rear is perceived much differently than stereo originating in the front. Sounds coming from the rear are coloured differently than sounds coming from the front. We are able to acquire more information from sound that comes from the front in terms of frequency response, distance of the sound source, and localization. This also makes sense from an evolutionary point of view in that we assume that a person, upon hearing a sound from behind, will turn around to gain a visual perspective and improved auditory perspective, allowing more accurate knowledge of a source's location.

In a multichannel context when front and rear stereo are perceived simultaneously, it is hypothesized that this integration of front and rear signals will provide some compensation for the lack of depth, skewed frequency response and the unfocused nature of rear stereo. The image in the front seems to be more coherent than it is in the rear. The sonic components that contribute to the stereo image are blended together more in the front give a sense of unity and integration. In the rear, sounds did not fuse together to create a unified whole and seemed to not have many internal connections. Some authors argue that only diffuse sound should be sent to the rear loudspeakers because we are less sensitive to sound sources at $\pm 120^{\circ}$ (Woszczyk, 1993).

There was also a sense that incoming sound from the front is open, where it was constricted in the rear. Front stereo gave more of a sense of definition to the music than

did rear stereo.

It is important to pose a few questions as a result of the investigation which might lead to further knowledge in this domain: How is our perception of rear stereo changed with the addition of related stereo music coming from the front? Do we pay more attention to the front stereo, making some of the above issues irrelevant? In other words, does front stereo mask or take precedence over rear stereo? These questions assume that front and rear stereo signals are somewhat correlated but are not exactly the same. There may also be applications for such a study in the area of automotive sound system design.

This investigation suggests that there is a relation between timbre and position. When listening to sound originating in the rear, not only was there a timbre change but there was also a change in the perceived position of the components in the stereo image, as compared to front stereo. This is consistent with Olive et al. (1994) and Bech (1994) who found that loudspeaker location has a significant influence over perceived timbre.

It should be noted here that the loudspeaker locations were chosen to represent the standard ITU loudspeaker layout for 5-channel reproduction. In this layout the rear loudspeakers have a wider aperture than the front loudspeakers. This difference in aperture alone may have contributed substantially to the differences in perception of front and rear stereo. Another interesting test would be to do a similar comparison with the rear loudspeakers at $\pm 150°$ as opposed to $\pm 120°$.

Because of the greater sensitivity to the front loudspeakers compared to rear loudspeakers, the production of stable lateral phantom images between front and rear loudspeakers is difficult. Chapter 5 presents results of a listening test concerning lateral

phantom images.

Due to the differences between the perception of front and rear stereo images, room simulation algorithms need to treat rear signals differently than front signals, and it is important to know in what way they differ. The preliminary study serves to inform the author about aspects of the multichannel sound field that were not apparent and which have not been researched extensively.

# Chapter 4

# Dynamic control of auditory perspective in a multichannel environment: System architecture

As mentioned in Chapter 2, the most common way for recording engineers to increase the perceived distance of a recorded sound source is to increase the "auxiliary send" level of the direct sound to an artificial reverberation device, and mix the return of the reverberation with the direct sound. This method changes only the level of the indirect sound relative to the direct sound. In a real room when a source changes location, delay time, direction, and amplitude of each early reflection changes, with respect to a given receiving point. In addition to this room modes (standing waves) are present in small rooms and alcoves of larger spaces, and the perception of room modes changes when a

source moves, with respect to a given listener position. The position of a sound source in terms of proximity to a room boundary also affects the perception of the sound source, in terms of its perceived size (localization blur and ASW) and spectral balance. All of these aspects are considered in the development of this dynamic control algorithm.

The main idea in the design of this control algorithm is to provide much more complex control of sound source distance than is currently available is multichannel reverberation devices taking into consideration the theories of distance perception as outlined in Chapter 2. As Woszczyk (1993) states, in order to take advantage of the increased imaging possibilities allowed in multichannel reproduction, more sophisticated methods of control of individual signals is needed. By coordinated adjustment of numerous physical parameters in real-time, according to source location, it is possible to create a strong impression of perceived source distance. Commercially available multichannel reverberation devices have recently started modelling early reflection patterns of real spaces (e.g., System 6000 by T.C. Electronic) rather than simply presenting a pseudo-random series of tapped delays of the original signal. In the System 6000 (which is generally considered by recording engineers to be the most sophisticated multichannel reverberation unit currently available), the reflection patterns are generated for a few fixed source locations but it is not possible to move the virtual sound source smoothly between these points and have the reflection patterns change in a similarly smooth way.

The proposed system of control of auditory perspective is considered dynamic because many physical parameters of the sound field are changed in an intelligent and coordinated way whenever a user adjusts the position of the virtual sound source. Pellegrini

(2001a) states that an auditory virtual environment is considered dynamic whenever the properties of the source, environment, or listener change over time. Although referring to 3-D audio systems, this is also applicable to multichannel systems.

Lehnert and Blauert (1992b) (in referring to Lehnert and Blauert (1992a)) states that the amount of information in a binaural room impulse response (reverberation time of 1 sec.) is 1000 times greater than the amount of information in the auditory environment which is processed in the conscious level of the human brain. This statement supports the idea of investigating and modelling only the perceptually most important aspects of a simulated room. A purely physical model would likely need to be very complex to include all of the important perceptual information.

With only five loudspeakers arranged at 0°, ±30°, and ±120° according to the ITU-R BS.775-1 recommendation ITU-R (1994), we are far from an ideal situation, in terms of ability to reconstruct a recorded or synthesized sound field. For instance, the wide lateral spacing between front and rear channels makes it difficult to accurately place phantom images to the side. As such it is necessary to compensate for the deficiencies by providing strong perceptual cues in the absence of accurate physical modelling.

## 4.1 Integrating perceptual models with physical models

This chapter will provide a detailed description of the proposed system of dynamic control of auditory perspective. The system that has been developed uses a combination of simplified physical models and perceptual models. The perceptual models are attempts to approximate what has been observed informally by the author but which has not been fully documented. Since the goal of this system is to provide a more realistic sounding multichannel reverb, the development of auditory models may be a more efficient means of achieving this goal than the development of physical models, due to the complexity of physical models necessary to provide a subjectively accurate model. Blesser (2001) points out that "ultimately, reverberation must use a perceptual language where scientific analysis intermediates between composer and listener. Quality reverberation derives from the cultural inheritance of historic musical forms and from the intrinsic properties of the human auditory system." He advocates the use of physical and perceptual models in the development of artificial reverberation.

In the development of this system, there has been a certain amount of fine tuning of the control functions by ear. Formally trained recording engineers learn basic microphone techniques such as ORTF, NOS, X-Y, A-B, and Blumlein from textbooks. Experience using any of the "textbook" techniques tells one that there is always room to improve the sound pick-up by listening to the result and making changes in the locations of the microphones or performers. Often times unorthodox microphone techniques

are developed through exploration and listening, that are far different than ones that provide so-called mathematically correct imaging characteristics. In the end it comes down to taste and it may be difficult to explain in mathematical terms why a certain microphone technique is more preferred than another. It has been found through informal experimentation in the development of this system that room simulation based on seemingly correct mathematical models, still requires fine-tuning "by ear" to achieve a natural or realistic sounding room simulation. Silzle (2002) found that tuning HRTF's by ear produced a significantly better result in an algorithm presenting surround program material binaurally over headphones. Those who work in computer modelling of visual images are also required to fine tune the rendered images by sight, despite the complexity of the mathematics behind the model (Lemley, 2001).

## 4.2 Architecture of the system

The system is composed of many components which work together to provide a unified multichannel sound field. Some of this material has already been reported in conference presentations by the author. (Corey et al., 2001a,b)

The system was implemented in software using MAX/MSP (Puckette and Zicarelli, 2001) running on two parallel Apple Macintosh G4's with CPU's of 733 MHz and 867 MHz. One G4 is able to perform the processing alone but the visual interface is much less responsive, therefore it was decided to use two machines in parallel. MAX/MSP provides a graphical programming environment for digital signal processing.

The components of the system, described in detail in this chapter, are as follows:

- Simulated axial room modes
- Multichannel tracking reflections
- Dynamic "fuzzy" sources for boundary effect simulation
- Dynamic equalization
- Multichannel panning
- Dynamic global gain control
- X-Y controller
- Graphical User Interface (GUI)
- Multichannel reverberation

### 4.2.1 Simulated axial room modes

Although room modes are present in real acoustic they have generally not been modelled in current digital artificial reverberation devices. Axial room modes are present in real acoustic spaces and are perceived more readily in small rooms and alcoves of larger rooms. As such it is important that they are present in room simulation systems. Research has shown that the placement of a loudspeaker in a listening room in relation to its adjacent boundaries has a strong effect on the perceived frequency response and sound quality (Olive et al., 1994; Bech, 1994; Allison, 1974). A loudspeaker placed near

a room boundary will have different perceived qualities than if placed at some distance from the room boundaries. We could infer that this is also the case for other sound sources in room, such as musical instruments. Recording engineers have been using boundary layer microphones for years, taking advantage of the acoustical qualities near a room boundary as differentiated from a free field.

Axial room modes represent a resonance that occurs between parallel walls in an acoustic space. Angus (1997) has indicated that in a real acoustic space, modes behave differently than diffuse sound. Modal decay may be longer than the diffuse sound decay, with the main difference being due to a reduction in absorption at the modal frequencies. Since the standing wave does not have a random incidence, absorption is specific to the surfaces involved instead of an average of all the surfaces in the room.

Room modes, particularly evident in smaller acoustic enclosures, help define, among other things:

- the perceived character of a given space,
- the perceived reverberation decay time, and
- the perceived location and movement of a sound source within the boundaries of that space.

The two dominant approaches to room simulation, the image source and ray tracing methods, fall short in modelling low frequency room resonances. Other researchers such as Savioja et al. (1995) and Murphy (2001), have proposed a method to simulate the resonant characteristic of acoustic spaces using 3-dimensional waveguide meshes. This

helps to augment the commonly utilized image source and ray tracing methods in the lower frequency band, but it has other problems associated with it as outlined in Chapter 2.

The addition of room modes to multichannel reverberation may provide the possibility for more control over the perceived sound source location, applying appropriate transformation of the spatial attributes due to source-boundary relationships and room dimensions. Efficient, real-time methods of room mode simulation are needed to enhance studio equipment.

The room mode module is essentially an infinite impulse response (IIR) comb filter with additional parallel feeds of the direct sound delayed and added to the output. The room modes are dynamically altered according to the distance between the direct sound and the virtual microphones, and provide comb filtering effects that change with changes in sound source location.

For the system being described here, a simplified model of axial room modes has been implemented to help provide more auditory information to the listener regarding the location and movement of the sound source. Simulated room modes may provide the possibility for more control over the perceived sound source location and movement, through appropriate transformation of the perceived spatial and spectral characteristics of the room and source.

The room modes are distributed to the five loudspeakers through what are referred to as virtual microphones. These microphones are assumed to be perfect omnidirectional point receivers and can be placed independently at any location within the room

boundaries. Sound sources are assumed to have omnidirectional radiation characteristics in the room mode module. Figure 4.1 illustrates the default positions of the virtual microphones in the room mode module and their respective correspondence to the loudspeakers.

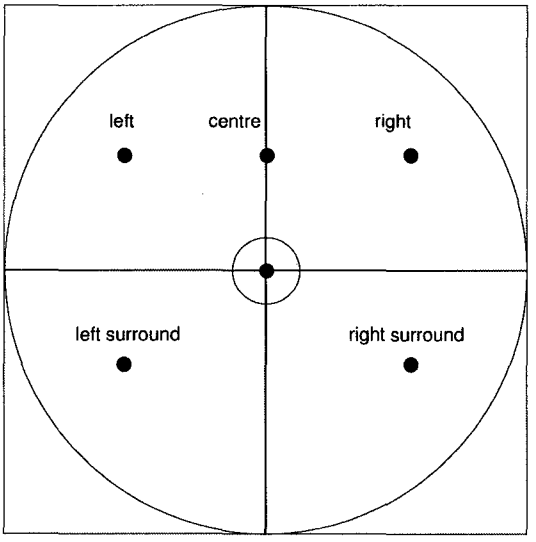


Figure 4.1: The default positions of the virtual microphones in the room mode module. This is not actually visible to the user.

We know from Kutruff (1991) that frequencies (in Hz) of normal modes of vibration of a room can be determined by Equation 4.1:

$$f = \frac{c}{2}\sqrt{\left(\frac{p}{l}\right)^2 + \left(\frac{q}{w}\right)^2 + \left(\frac{r}{h}\right)^2} \tag{4.1}$$

where $c = 344$m/s (speed of sound in air),

$l$ =length of the room,

$w$ =width of the room,

$h$ =height of the room,

$p, q, r$ =harmonic number (integer).

Figures 4.2(a) – 4.2(d) (below) illustrate the frequency response curves for four different source and microphone locations in the room mode module. The upper right-hand corner of each figure indicates a top view of the simulated room with microphone and source locations. The frequency response graphs indicate the automatic spectral modification of the simulated modes according to the source and microphone locations. These automatic changes in spectral balance are performed smoothly in real-time with changes in source location.

In addition to the automatic spectral modification related to source position, additional gain changes are applied to the modes according to source location. For example, when the source is in the centre of the simulated room, the room modes from all channels are set to unity gain. This helps to bring the perceived direct source image closer to the listener. When the sound source moves towards a room boundary, the room modes originating from that particular direction are increased moderately in level, while modes that originate from all other directions are attenuated. This appears to have the effect of accentuating the boundary effect when the source moves towards a wall.

Only axial modes in the two horizontal dimensions are implemented. Furthermore the modes from the two dimensions do not interact, but are simply summed at the output.

Figures 4.3(a) – (e) illustrate the sound transmission paths and distances that are used to calculate the respective delay times for a single dimension in each virtual mi-

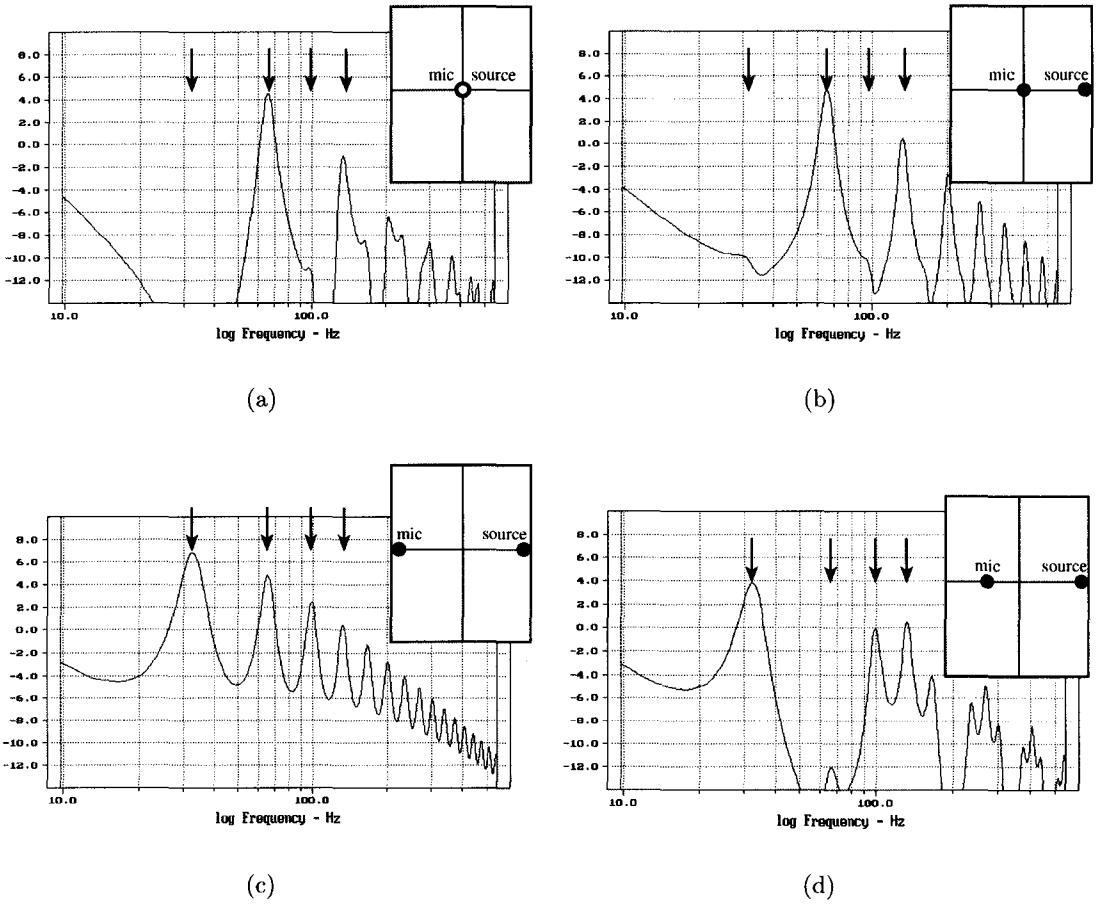


Figure 4.2: Frequency response plots of a room mode in one dimension for a single microphone for different source and microphone locations. The upper right corner of each figure indicates the locations of the source and microphone in the room for the particular frequency response. The four arrows indicate the frequencies of the first four harmonics of the mode. Note how the frequency response changes dynamically and automatically according to source and microphone locations.

crophone. These distances are calculated using the X-Y coordinates of the sound source and virtual microphone.

Figure 4.4 illustrates a block diagram of the room mode module. The signal is split into five paths and subsequently delayed, multiplied by the reflection coefficients (beta) and low-pass filtered. The fifth signal path is fed back to the input of the module, providing a recursive delay, to help model the resonance of the simulated room. The resonant frequencies are directly related to the room dimensions. The five delay times are calculated according to the five transmission paths in Figure 4.4, where:

$Delay_N = distance_N/c; c = 344m/s$

Low-pass filters have been added for two reasons: to reduce the amount of flutter echo and because of the fact that the room mode module is implemented to model only the low frequency components of the room simulation. The default cut-off frequency of the first-order low-pass filters is set to 200 Hz. This is related to the "Schroeder frequency" which is the crossover frequency marking the transition from individual, well separated resonances to many overlapping normal modes (Schroeder, 1996).

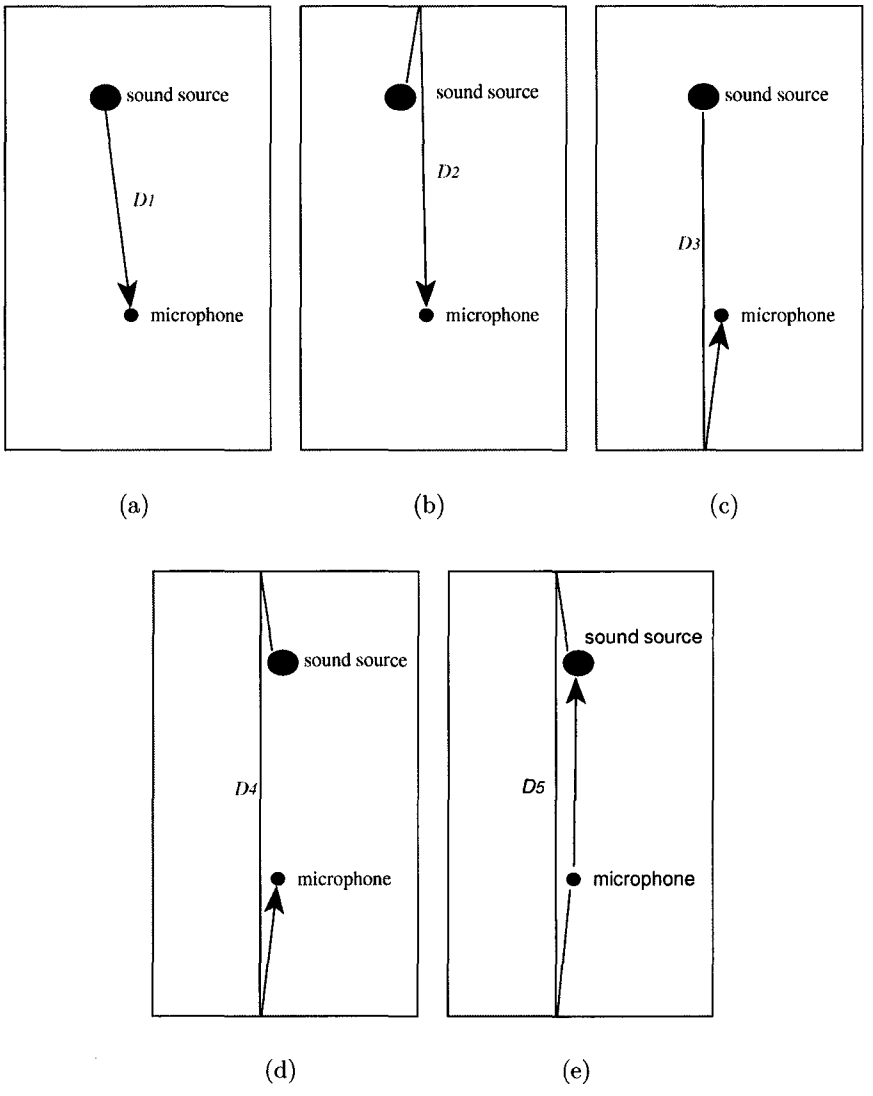


Figure 4.3: An illustration of the five sound transmission paths simulated by the room mode module.

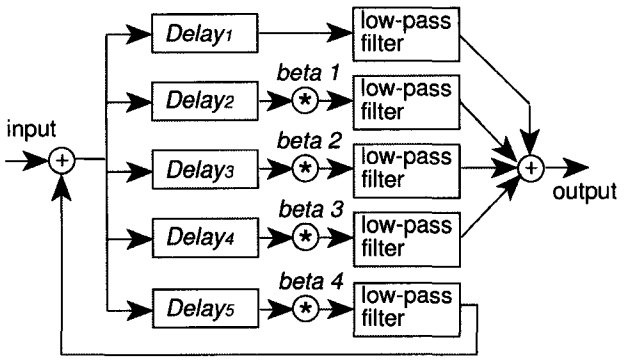


Figure 4.4: A block diagram of the room mode module. Delay times are calculated according to sound transmission paths as outlined in Figure 4.3a-e. Multiplication by beta represents an approximation of the reflection coefficient of the walls. Low pass filtering at the output reduces flutter echo.

### 4.2.2 Multichannel tracking reflections

Early reflections patterns that are updated smoothly in real-time according to source location are an essential component of any room simulation system but have not been implemented in any effective way in commercially available reverberation units. They help provide control over sound source distance, give a sense of the boundary effect and help give an impression of the room size. Horbach et al. (2000) states that "convincing room effects can hardly be achieved by means of externally inserted reverberation processors only, because the discrete reflections, the simulation of which is an important part of any auralization scheme, depend on position and distance of each individual sound source."

For the system proposed herein, early reflections have been implemented using a 4th-order image model (Allen and Berkley, 1979) providing 40 reflections, rendered

from all directions. The timing and angular location of the reflections are calculated according to the sound source position and room dimensions, as referenced to a central listening position. Because they are updated dynamically according to source location, they are referred to as tracking. Reflections are generated for the horizontal plane (two dimensions) of a geometrically simple room, i.e. rectangular.

Early versions of the system employed static reflections generated by an external hardware device (T.C. Electronic System 6000, a state-of-the-art multichannel reverberation unit). Although the reflection patterns generated by this device were carefully calculated and later tuned by the manufacturer, they represent a single sound source location. Early reflection patterns can be generated for a number of static source locations, but it is impossible to move smoothly between these static points and have the reflections smoothly change from location to location. On the other hand, in a real acoustic enclosure, for a given receiving point, the reflection pattern will change according to changes in the location of the sound source. An ideal room simulation system would also have this feature. Tracking reflections will help provide listeners with a clearer indication of the location of the sound source. Results from listening tests (Chapter 5) support this hypothesis. Tracking reflections are needed to influence the perception of distance of a source.

It has been found through informal experimentation and listening that tracking reflections alone help give a sense of the boundary effect when the source approaches a wall. There is a gain in the perceived low frequency band and a reduction in the perceived focus or clarity when the source is next to a room boundary, as opposed to being

placed at some distance from the boundaries.

The gain, propagation delay, and angular location of each early reflection are calculated in real-time according to the sound source X-Y position as related to the centre of the simulated room. The linear gain is calculated according to the distance of each reflection where:

$$gain = 1/distance \tag{4.2}$$

It is assumed that a distance of 1m represents unity gain, where the level is reduced by 6dB for every doubling of distance. This models the natural decrease in sound level by 6dB for each doubling of distance in a free field (Mershon et al., 1989).

The delay time for each incoming reflection is also calculated according to the distance travelled for each reflection where:

$$delay = distance/c; c = 344m/s \tag{4.3}$$

The delay time is updated at the sampling rate which provides a Doppler effect.

Reflections were low-pass filtered to approximate wall and air absorption, with each successive order of reflections having a lower cut-off frequency. Specifically the cut-off frequencies were:

- 1st order reflections: $f_c = 16$ kHz
- 2nd order reflections: $f_c = 12$ kHz

- 3rd order reflections: $f_c = 10$ kHz
- 4th order reflections: $f_c = 8$ kHz

Roughly speaking, air is a low-pass filter whose magnitude of the transfer function has a Gaussian shape and whose cutoff frequency depends on the distance travelled and on humidity. At 100ms delay, $f_c =$ 8000Hz, at 1000ms delay $f_c =$ 2000Hz. (from Lehnert and Blauert (1992b))

Each reflection is panned according to the calculated angle of incidence to the centre of the room. Pair-wise constant-power panning is used to distribute the reflections to the loudspeakers.

### 4.2.3 Dynamic "fuzzy" sources for boundary effect simulation

As a sound source moves towards a room boundary in an acoustic enclosure, the perceived sound image changes in several different ways. Empirical evidence seems to indicate that the proximity of a room boundary to a sound source affects, among other things:

1. the perceived spectral balance of the direct source and the reflected sound
2. the perceived size of the direct sound as it perceptually fuses with a room boundary

Control of sound source location in a multichannel room simulation needs to include the ability to place a sound source near a wall or room boundary of the simulated room. The location of a sound source in a real acoustic space has a direct effect on the spatial,

temporal and frequency characteristics produced at a given location in the room. A sound source placed next to the wall of a room will create an entirely different aural impression than if the same source is placed in the middle of the room. Any frequency components that are normally radiated towards the rear of a source will be directed out towards the middle of the room when the source is against a wall. The result is a loading of low frequencies and an increase in the perceived size of the sound source as it couples with the wall near it.

To create an impression of the "boundary effect", so-called "fuzzy" sources are generated and added to the direct sound in the multichannel sound field. This component of the system helps to provide the impression of a sound source moving towards a rigid room boundary. As opposed to constructing a physical model of the boundary effect, it was decided to construct a perceptual model that would approximate what has been observed but which would not necessarily reflect the physical behaviour of sound in a real room.

The four channels of fuzzy sources are produced by parallel feeds of the direct sound to stereo digital reverberation units (T.C. Electronic M3000's) that are configured to provide short impulse responses. Specifically the parameters are configured as follows (specific parameter settings are given in Appendix A):

- early reflection pattern is set to model that of a small room,
- the reverberation time set very short,
- the reverberation level set very low,

- the direct sound completely attenuated.

Figures 4.5 and 4.6 show the impulse response and FFT of the impulse response respectively. As can be seen from the time response, there is a dense series of echoes which help provide a diffuse sounding image.

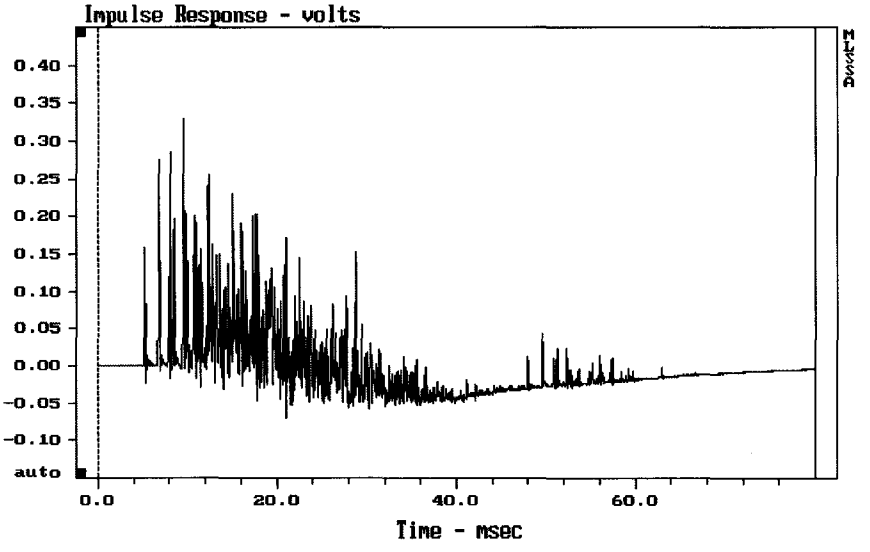


Figure 4.5: An impulse response measurement of a fuzzy source.

The fuzzy sources are treated as four separate moveable sound sources. They are located around the direct sound as in Figure 4.7 and move with the direct sound so as to maintain the same relative position and image continuity.

The fuzzy sources are fed to the multichannel processor and routed to the loudspeakers according to their X-Y location, which is relative to the direct sound coordinates. The levels of the direct sound feeds to the stereo reverberation units are controlled by the computer according to the location of the direct sound source. As the direct source approaches the front or back wall, fuzzy sources 1 and 2 are increased in level to a max-

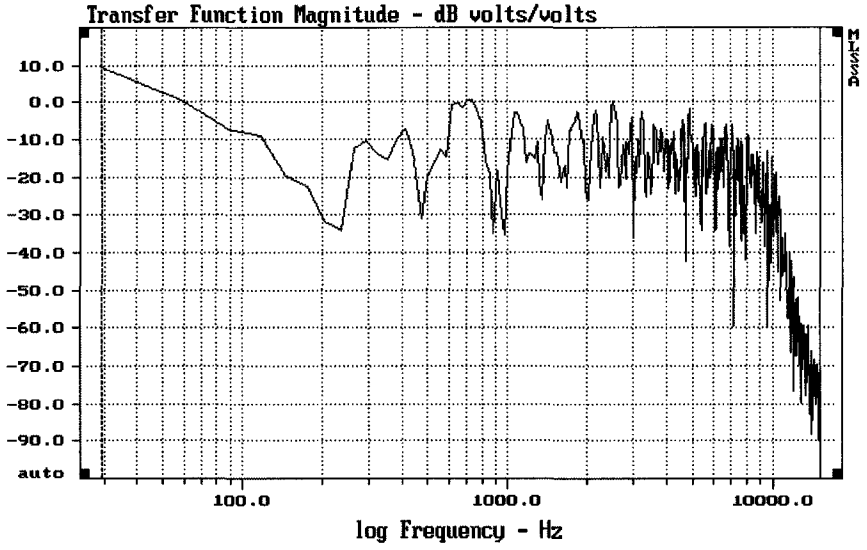


Figure 4.6: A 1024-point FFT of the impulse response measurement of a fuzzy source.

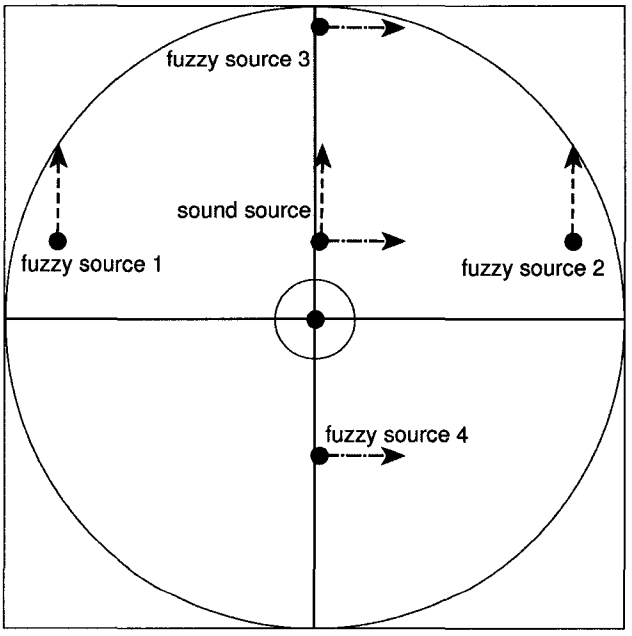


Figure 4.7: The locations of the four fuzzy sources relative to the direct sound. The arrows indicate the direction of movement that produces gain changes for each fuzzy source.

imum when the direct sound is against the wall. As the direct sound approaches a side wall, fuzzy sources 3 and 4 are increased in level. As the direct sound moves toward the centre of the room, the level of the fuzzy sources decreases until they are completely attenuated when the direct sound in the centre of the room. As an example of the gain function applied to the fuzzy sources, Equation 4.4 illustrates the gain applied to the left front fuzzy source, where $y$ refers to the source's location with respect to front-back, $l$ refers to the length of the room. The variable $k$ is defined in Equation 4.5, where $w$ refers to the room width. The variable $k$ is a scaling factor to translate the square control surface of the GUI into the rectangular shape of the simulated space.

$$g = \frac{8\left(\frac{|y-\frac{l}{2}|}{l}\right)^2}{k} \tag{4.4}$$

$$k = 1 + \left|-3\left(\sqrt{\frac{x^2 + (y-l)^2}{w^2 + l^2}}\right)\right| \tag{4.5}$$

The fuzzy sources have a low level of correlation in order to widen the sound source when they are added to the direct sound. The level of correlation of one fuzzy source to any other is approximately 0.15 as measured with a Brüel & Kjær 2035 Signal Analyzer.

### 4.2.4 Dynamic equalization

Dynamic filtering is implemented in software to enhance or exaggerate the effect of distance (high frequency attenuation) and proximity to a room boundary (low frequency boost). Filtering is performed on the direct and fuzzy sources according to the location of

the direct source within the room boundaries. Slightly different filtering is performed on the direct and fuzzy sources. For the case of the direct source, no filtering is applied when it is in the centre of the room. When it moves towards the room boundaries, equalization is applied to the upper frequency band and the lower frequency band. Fuzzy sources are boosted in both the low and high bands when the source moves towards a room boundary. Shelving filters are implemented with cut-off frequencies of:

- low shelf: $f_c = 175$ Hz
- high shelf: $f_c = 8000$ Hz

The cut-off frequencies were found empirically to provide the strongest impression without being to obtrusive. Extensive informal testing determined that the range of gain applied to the filters was no more than $\pm 3.5$ dB, with the greatest amount of filtering applied when the source is near a room boundary.

The control function for the gain applied to the shelving filters is illustrated in Equation 4.6, where $x$ and $y$ refer to the coordinates of the sound source and $l$ and $w$ refer to the length and width of the room respectively. Essentially the gain is determined by the proximity of the source to a room boundary.

$$gain = \frac{0.5 + \left(\sqrt{\frac{(x-w/2)^2+(y-l/2)^2}{(w/2)^2+(l/2)^2}}\right)}{\sqrt{2}} \tag{4.6}$$

### 4.2.5 Multichannel panning

Although other panning algorithms exist such as vector-base amplitude panning (Pulkki, 2001), Ambisonics (Gerzon, 1992b), polarity-restricted cosine (Martin et al., 1999a) the most commonly used panning algorithm, constant-power (sine/cosine), was chosen to position the direct sound. This method has the advantage of maintaining constant sound power no matter where the sound is positioned. Initial versions of the system utilized the panning function of a commercially available state-of-the-art multichannel reverberation unit and processor (T.C. Electronic System 6000). The sine/cosine panning function is scaled to fit the different loudspeaker apertures such that when a source is positioned midway between two loudspeakers, the gain of each loudspeaker signal is 0.707 (−3 dB). For instance to position a source between the front centre (0°) and right (30°) loudspeakers, the gains ($g$) applied to the loudspeaker signals would be as in Equations 4.7 and 4.8, where $\phi$ = intended angle ranging from 0–30° (0-$\pi/6$). In the equations, $\phi$ is multiplied by 3 to scale the loudspeaker aperture from 30° to 90°.

$$g_{centre} = cos(3\phi) \tag{4.7}$$

$$g_{right} = sin(3\phi) \tag{4.8}$$

The direct source gain is also changed according to its distance from the centre of the room, where:

$$gain = 1/distance \tag{4.9}$$

and the minimum distance is 1 m.

As with the early reflection patterns, the initial pre-delay time of the direct source is determined by its distance from the centre of the simulated room, where:

$$delay = distance/c; c = 344m/s \tag{4.10}$$

Because the delay time is updated at the sampling rate, the Doppler effect is modelled automatically.

### 4.2.6 Dynamic global gain

A global gain control is applied to the output bus of the system according to sound source location to compensate for the increase in sound power as the direct source moves close to a room boundary. The increase in sound power is due to the addition of fuzzy sources, room modes and frequency equalization when the direct source is near a room boundary.

### 4.2.7 Multichannel reverberation

For the generation of the reverberant tail a state-of-the-art multichannel hardware-based signal-processing device is used, namely the System 6000 by T.C. Electronic. Other sophisticated multichannel reverberation devices that provide at least 5 channels of decorrelated reverberation could be used as well. The development of a reverberation

algorithm is outside of the scope of this thesis.

The RT60, or reverberation decay time, is calculated using Sabine's equation, according to the room dimensions and global reflection coefficients, as specified by the user in the graphical interface. Reverberation time, being the amount of time in seconds that is required for 60 dB of sound decay after a sound is turned off, can be calculated using Sabine's equation:

$$RT60 = 0.163\frac{V}{\alpha S} \tag{4.11}$$

where $V$ is the cubic volume of the room in $m^3$ measured as though no seats are present and $\alpha$ is the average absorption coefficient of all surfaces, $S$ is the total surface are of the room in $m^2$ (Beranek, 1996). The calculated reverberation time is transmitted from the desktop computer providing the user interface to the multichannel reverberation hardware via MIDI.

The fuzzy sources feed the reverberation and therefore the level of the reverberation is proportional to the level of the fuzzy sources, which is directly dependent on the proximity of the sound source to a wall.

### 4.2.8 Graphical User Interface (GUI)

The GUI was implemented in software on the primary desktop computer (Figure 4.8), and it allows the user to have a visual representation of the sound source location and movement in the horizontal plane of the simulated room. The sound source is represented

by a red dot within the four boundaries of the room that moves in real-time with changes in auditory perspective, as controlled by the joystick or computer mouse. The listener location is represented by a blue dot in the centre of the room. The X-Y location of the dot in the interface, representing the source location, is transmitted to the early reflection generator, room mode module, fuzzy sources (via MIDI), dynamic equalization control, and dynamic level control, which are all changed in real-time when the source location changes.

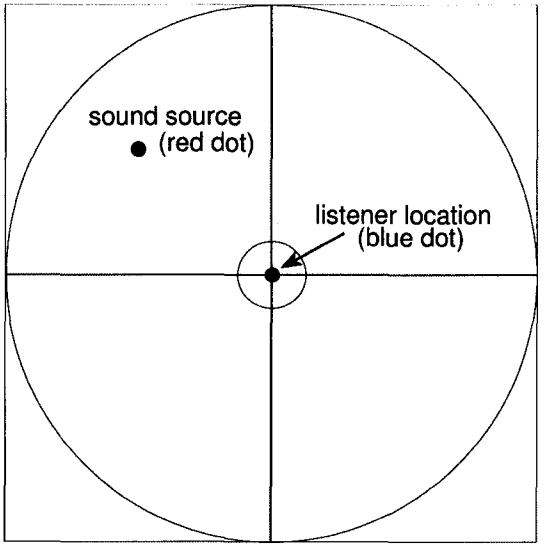


Figure 4.8: A diagram illustrating the GUI. A red dot represents the source position and the blue dot represents the listener position, which remains stationary in the centre of the virtual room.

### 4.2.9 Intuitive X-Y controller for the manipulation of auditory perspective

The initial versions of the system employed a computer mouse as the device to position the sound source within the boundaries of the simulated room. By clicking and dragging a red dot (representing the sound source) within the GUI on the computer screen, users of the system were able to move the sound source. This method proved to be cumbersome and inefficient. The GUI represents a top view of the simulated room with a blue dot in the centre of the room representing the location of the listener. It was found through informal experimentation that a more intuitive controller than the mouse was needed. A commercially available isotonic joystick (Figure 4.9) has been implemented to improve this deficiency.

Figure 4.9: A photograph of the joystick used for the control of auditory perspective.

The joystick has been configured to control the velocity of the sound source. The joystick normally rests in a central, vertical position. When pressure is applied to the top

part of the joystick, it leans in the direction it is being pushed. The further it leans, the greater the velocity of the sound source according to the direction of pressure applied to the joystick. The source location is represented both aurally (from the loudspeakers) and visually (in the GUI). The velocity controller has been carefully tuned through extensive informal experimentation and listening tests to allow a wide range of control from a slow, precise movement to a more accelerated movement. The joystick is also configurable to provide absolute sound source position control. That is, the two-dimensional location and movement of the sound source is mapped to the location and movement of the joystick.

### 4.2.10 Management of control functions

Two Apple Macintosh desktop computers, a primary and secondary, are employed to integrate the components of the system and to provide real-time DSP in software. The primary computer performs the following functions:

1. the transmission of parameter changes to the multichannel reverberator via MIDI communication according to sound source location

2. the transmission of the sound source X-Y coordinates and room dimensions to the secondary computer via ethernet

3. the simulated room modes implemented in software

4. a graphical user interface (GUI)

5. gain changes to the input signal according to source location

6. filtering of the input signal according to source location

7. an input for the joystick controller

The secondary computer performs one main function: real-time synthesis of tracking early reflections according to the source X-Y location, room size parameters, and reflection coefficients of the walls as received via ethernet from the primary computer.

A block diagram of the system is shown in Figure 4.10, outlining the signal flow of the system. The sound source is fed to the multichannel panner (software in MAX/MSP), room mode module, early reflection generator, and to four stereo reverberation units. Since the level of the fuzzy sources is changed dynamically according to source location and they are fed directly to the multichannel reverberation, it helps to change the direct-to-reverberant level dynamically as well. Having the fuzzy sources feed the reverberation unit directly was also chosen so that the input would be more temporally complex than simply using the direct source.

All of the control functions have been tuned carefully through thorough listening and experimentation to provide a strong impression of sound source perspective control. The following chapter will outline a series of listening tests conducted on the system to evaluate its ability to perform as designed.

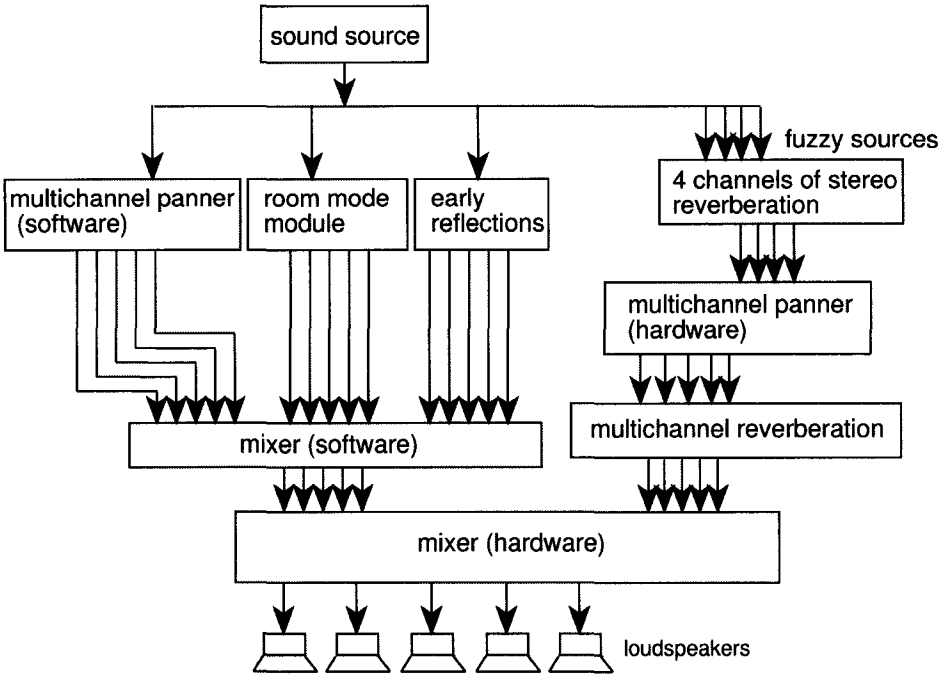


Figure 4.10: A block diagram of the entire system. The multichannel panner (software), room mode module, and mixer (software) are performed by the primary computer using MAX/MSP. The early reflection generator is performed by the secondary computer using MAX/MSP. The hardware mixer is a Yamaha O3D, and the hardware reverberation and multichannel panner is generated by a T.C. Electronic System 6000.

# Chapter 5

# Evaluation of the system with listening tests

In order to test the design of the described system, a series of listening tests were conducted. Since a perceptual model, as opposed to a physical model, was intended in the development of this control algorithm it was imperative to perform an evaluation of the system through a series of listening tests. The idea is to manipulate numerous physical parameters to provide a perceptual effect, and to find the method of control that best produces the intended perception.

## 5.1 Technical description of the listening room

Listening tests were conducted in the Multichannel Audio Research Laboratory (MAR-Lab), an acoustically damped listening room with a 3/2 loudspeaker layout according to

the ITU-R BS.775 recommendation (ITU-R, 1994). Five two-way active loudspeakers (Bang & Olufsen Beolab 4000) were placed at 0°, ±30°, and ±120° with a radius of 2.1 m from the central listening position. (See figure 5.1 for a top view of the listening room). An acoustically transparent, visually opaque curtain was hung directly in front of the loudspeaker array. It served to conceal the location of the loudspeakers from the listeners so that their auditory perceptions would not be influenced by the visual perception of speaker positions. It is well known that when making judgments about sound source location that the auditory system can be influenced quite strongly by visual judgments of source location (e.g., Jackson (1953)). A good literature review of auditory-visual interaction can be found in Storms (1998). A visually opaque yet acoustically transparent curtain is hung in front of the loudspeakers to conceal their visible location from the listener for the purposes of listening tests. Heavy curtains are hung from ceiling to floor against the walls along the perimeter of the room. Additional low-frequency absorption is provided by Helmholtz attenuators attached to three of the walls of the room. Above the ceiling panels, which are suspended on a grid, there is a 10cm layer of mineral wool to further dampen resonance of the room.

The ambient noise floor of the listening room is 25 dB SPL (A-weighted, slow time weighting), as measured with a Brüel & Kjær 2235 precision sound level meter. Calibration of the loudspeaker levels was performed using a Brüel & Kjær 2235 sound level meter, with A-weighting and fast time weighting. Distance of the loudspeakers from the central listening position was calibrated by comparing time delay measurements using a Maximum Length Sequence System Analyzer (MLSSA) (Rife, 2000). The room rever-

beration time is approximately 0.270 sec at 125 Hz, 0.172 sec at 250 Hz and 0.168 sec at 500 Hz as noted in Martin (2001).

Participants were seated behind a desktop computer display in the centre of the loudspeaker array. The computer was used to automate the test, record the participants responses, and perform the panning and generation of early reflection patterns in real-time. MAX/MSP was used to perform not only digital signal processing, but also automation of the listening tests and recording the participants' responses. The CPU of the computer was placed in an acoustically isolated machine room to minimize the noise floor of the room.

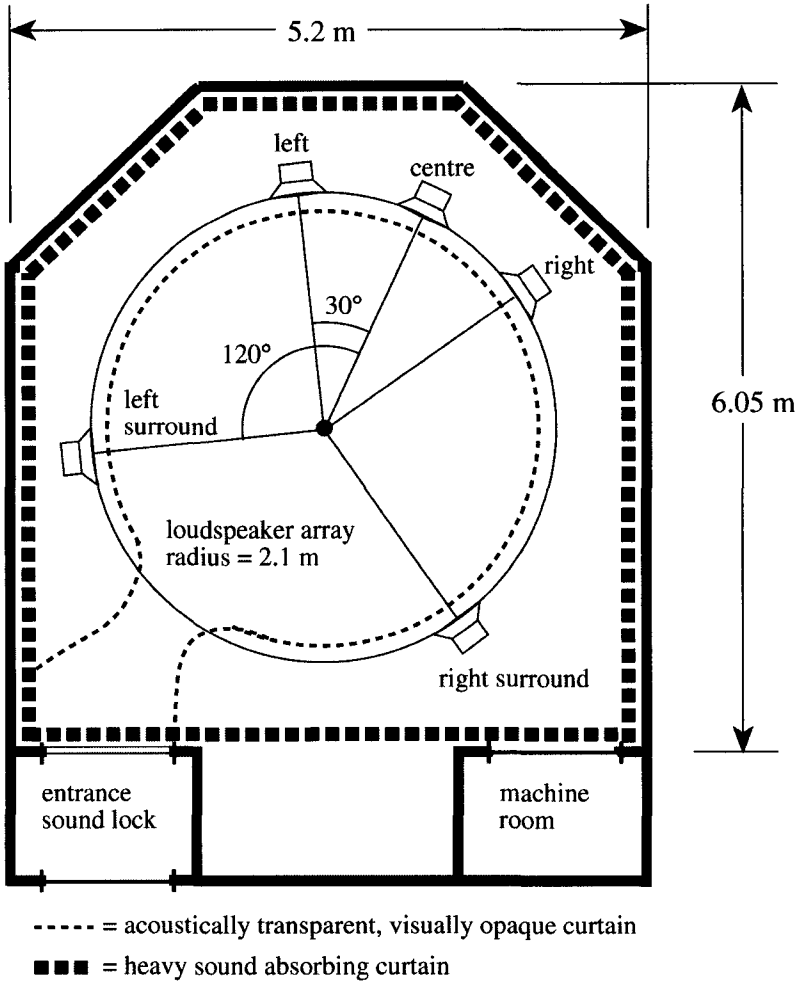


Figure 5.1: A top-view diagram of the MARLab illustrating the loudspeaker layout and room dimensions. The floor is carpeted and heavy curtains hang along the perimeter of the room (indicated by a thick dotted line in the diagram). An acoustically transparent yet visually opaque curtain hangs in front of the loudspeakers as indicated by a thin dotted line near the loudspeaker array.

## 5.2 Experiment 1: Subjective loudness calibration

Listening tests were conducted to compare a state of the art multichannel reverberator alone to the same device with the enhancements described above. Specifically, the two processing systems compared were configured as follows:

1. a multichannel panner and reverberation system (i.e., T.C. Electronic System 6000)

2. a multichannel panner and reverberation system with added room modes, fuzzy sources, and dynamic equalization

One of the goals of conducting the listening tests is to determine if subjects with no previous knowledge of the system find that the enhancements described in the previous chapter do in fact provide more control over sound source distance. The aim of this work is to design a system of control for multichannel reverberation systems based on the perceptual attributes of a multichannel sound field. Listening tests are performed to evaluate this design.

The aim of the first test was twofold: to have participants match the loudness of pairs of examples and to investigate listener consistency in loudness calibration. Their individual loudness calibrations were then used in the second test.

### 5.2.1 Participants

For the first four listening tests eight participants from McGill's Graduate Program in Sound Recording volunteered to participate. All participants have completed a well-

established course in timbral ear training (Quesnel, 1996), have experience recording and mixing music, and perform critical listening tasks on a near daily basis. As such they can be considered expert listeners (Stone and Sidel, 1993). Seven of them had no prior knowledge about the system being tested. One had some knowledge of the processing being done, but did not have experience listening to it. Listeners' heads were not fixed in one location. They were free to move their heads if needed. It was decided not to fix the listeners' heads because the intention was to conduct the tests under conditions close to those found in a natural listening environment. They were shown a mark on the ceiling indicating the centre of the room, from which they could judge their position and ensure they were correctly located in the centre of the room.

### 5.2.2 Method

For this experiment four sound source locations were tested (Figure 5.2), although the participants did not have a visual representation of the intended locations:

1. 0° front, near the centre of the simulated room
2. 0° front, near the front wall of the simulated room
3. 90° to the left, near the left wall of the simulated room
4. 45° to the right, near the front right corner of the room

The four sound source locations are intended locations of phantom images produced by the system. The four intended locations were not judged in terms of their absolute

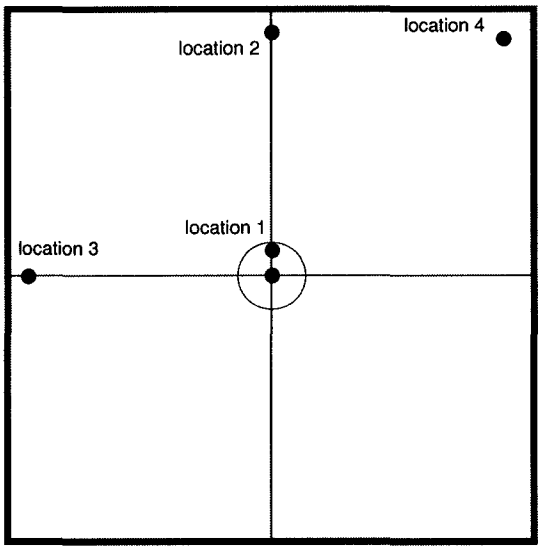


Figure 5.2: Source locations used to evaluate distance in the second test. Note: The participants did not see this display on their monitor.

perceived location by the listeners. Instead they simply served as four distinct locations that could be used to compare the commercial system with the proposed system. It was not necessary to prove that the intended locations matched the perceived locations, but to examine any differences in perception between the enhanced and non-enhanced systems.

The choice of sound sources fulfills three requirements:

1. percussion (bongos)

2. a musical source, acoustic guitar

3. a speech source (female, Danish) with a timbre familiar to all listeners

The speech and percussion samples were taken from the Bang & Olufsen *Music for Archimedes* compact disc (Bang & Olufsen, 1992), which were recorded in an anechoic

chamber. The percussion sample was a recording of bongos playing a fairly rapid series of notes. The guitar sample was recorded using a microphone placed very close to the guitar, providing isolation from the room. The guitar sample was a sustained arpeggiated melody with a strong sustaining bass component providing a relatively steady-state sound source.

The choice of sound sources represents the fulfillment of three criteria: a transient source (percussion), a more steady-state source (guitar), and a third source (speech) that exhibits a mixture of transient and steady-state characteristics. Sound sources were monophonic anechoic recordings, and were played from the hard disk of the computer. For all listening tests, the sound samples were repeated for each test question until the listener made a choice and moved on to the next example.

The combination of variables location (4) and sound source (3) make a total of 12 permutations. There were 48 trials in this test. The 12 permutations were repeated four times to determine listener consistency and to get an average calibration level for the second listening test. The test was fully-factorial for each listener. The average listener-calibrated levels for these 12 permutations were subsequently used for the second listening test. Thus, calibrations used in the second listening test were specific to each individual and represented each listeners choice of equal loudness. Participants were not told that their calibrations made in the first test would be used in subsequent listening tests.

The participants were presented with two stimuli, A and B, representing the enhanced and non-enhanced systems respectively. For this test A and B were identical sound

source inputs in the same X-Y locations within the simulated room. The only difference between them was that one was processed through the enhanced system and the other was processed through the original non-enhanced system. Participants were then asked to adjust B so that its loudness was equal to A. The resolution of level adjustment of B was 0.5dB, and the change in level was applied to the signal at the input to the system (see figure 5.3). Participants could adjust the level with the arrow cursor keys on the keyboard. The total range of adjustment was 10dB.

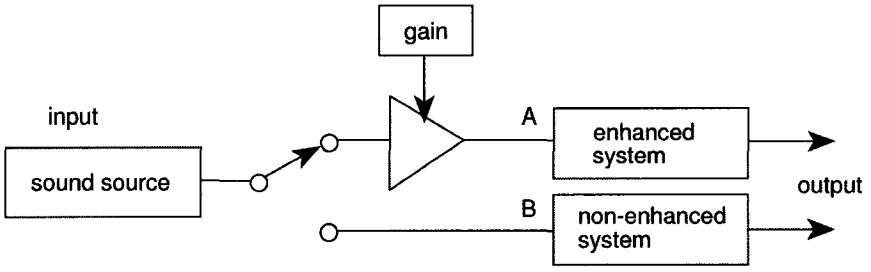


Figure 5.3: A block diagram illustrating the point where the gain is applied to the sound source.

Participants were instructed to consider the overall loudness of both conditions in matching the loudness. They were asked to evaluate only the loudness summation of the all-inclusive sound image and not just the direct sound, especially in cases where A and B had different direct-to-reverberant ratios.

The decision to have listeners equalize the loudness of A and B for themselves was made because objective measurements of loudness differences did not correspond to the perceived loudness difference in the multichannel sound field. It was found by the authors that when a sound level meter was used to calibrate the sound examples for equal level in dB SPL (A-weighted), the examples were not equal in loudness. This

is consistent with Aarts (1991) who conducted a test in which participants were asked to match the loudness of different loudspeakers, where the loudspeakers had different frequency responses. His results indicate that when loudspeaker pairs were calibrated objectively using a sound level meter to have equal SPL (dBA), this level was not consistent with the level that participants gave to indicate equal loudness. He concluded that the A-weighting method is not recommended for accurate loudness calibration. Since Aarts used human participants as ultimate judges of equal loudness with which he evaluated several methods of loudness calculation, it was decided to use the human subjects themselves to balance loudness in these multichannel experiments. Flanagan and Taylor (1999) also state that for the same measured sound pressure level, one type of loudspeaker was perceived to have a greater loudness than the other. In addition, Bech (1998) reported that program material had a significant effect on subjective loudness calibration. From this we can conclude that subjective loudness calibration is an effective method to deal with the complexity of the task of loudness matching when assessing very different multichannel stimuli.

Loudness is a perceptual estimate of sound source level and some of the factors influencing loudness perception are (Flanagan and Taylor, 1999; Zwicker and Fastl, 1999):

- sound intensity
- spectral balance
- temporal characteristics

In the test described here, the two conditions (A and B) being compared varied in a number of ways including spatial, temporal and spectral characteristics:

- frequency equalization
- perceived image size or width
- length and composition of the impulse response

### 5.2.3 Results

Figures 5.4–5.7 illustrate the average gains (with 95% confidence intervals) applied to B to make B equal in loudness to A for the 12 examples. The gain values represent each participant's average level of gain adjustment in B. The average levels were used for the gain calibration in the second part of the test.

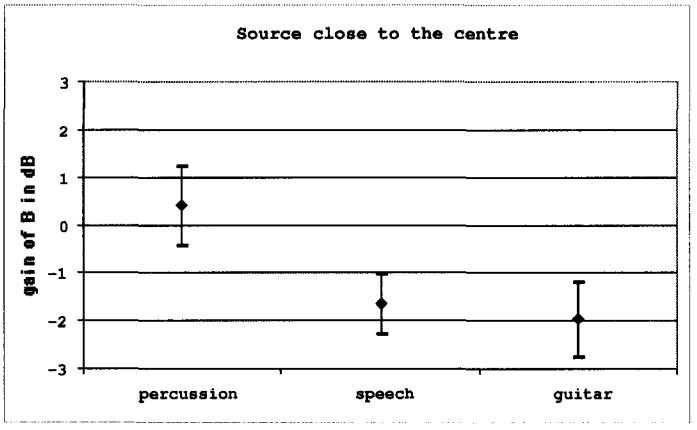


Figure 5.4: The average gain of B as set by the listeners with 95% confidence intervals for each sound type positioned close to the centre of the room.

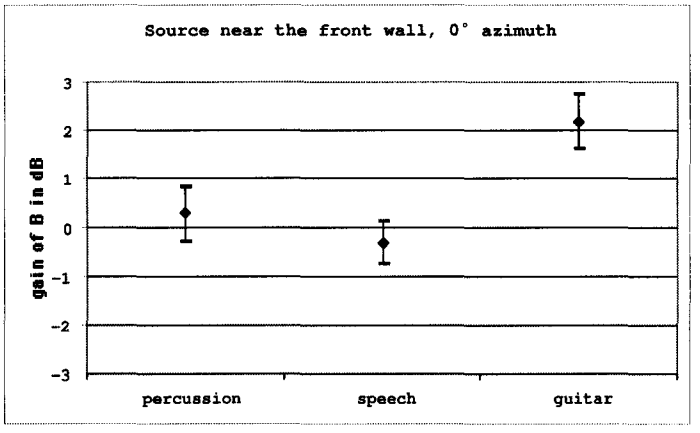


Figure 5.5: The average gain of B as set by the listeners with 95% confidence intervals for each sound type positioned near the front wall of the simulated room, 0° azimuth.

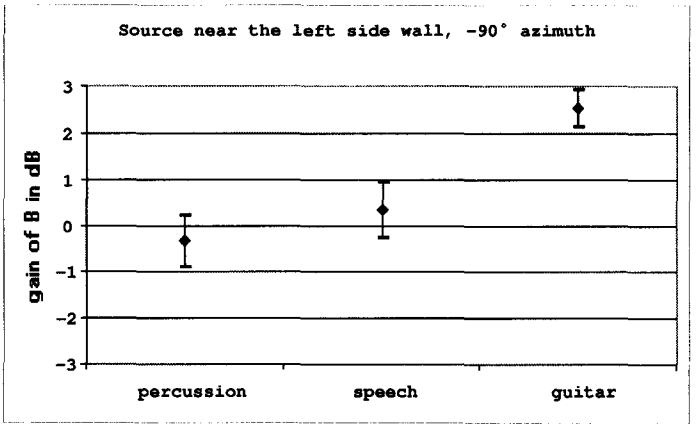


Figure 5.6: The average gain of B as set by the listeners with 95% confidence intervals for each sound type positioned near the left side wall, $-90°$ azimuth.

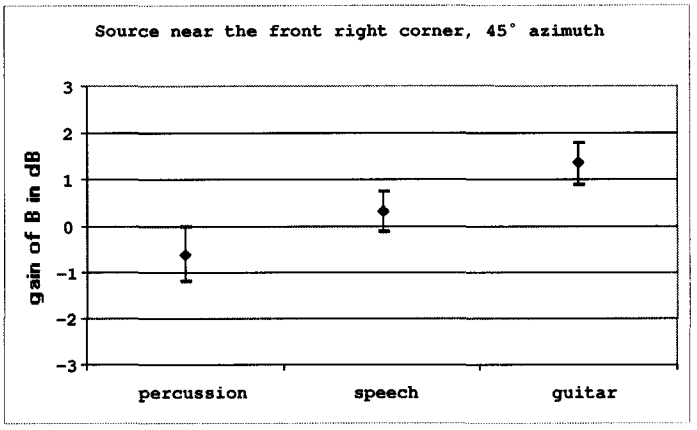


Figure 5.7: The average gain of B as set by the listeners with 95% confidence intervals for each sound type positioned near the right front corner, 45° azimuth.

### 5.2.4 Discussion

The graphs indicate the average gains of B as set by the listeners such that condition B was the same loudness as condition A. The graphs also provide information on the effect of using different sound sources for listening tests. One interesting point that these graphs indicate is that for sources located near a room boundary, the guitar sound source was set to a higher mean gain than the speech and percussion sources. This can possibly be explained by the fact that the speech and percussion were more transient in nature than the guitar. The more transient sounds would allow the listeners to perceptually "segregate" (Bregman, 1990) the direct sound from the reflections and reverberation and therefore match the direct sound levels only. The guitar sound was perhaps more perceptually integrated with the reverberant energy and therefore the entire sound field was taken into account when the loudness judgement was made. The enhanced system

(A) provides more reverberant energy when the sound source is near a room boundary than the non-enhanced system (B). Since the guitar has a high perceptual fusion with the reverb, the level of B had to be raised more to be the same loudness as A because B (non-enhanced) had less reverberant energy.

There is a possibility that listeners had more difficulty in integrating the loudness impression from a number of components heard in the percussion sound. Their attention could be divided between assessing the loudness of the room response and of the percussion sound once the transient nature of the stimulus promoted temporal and perceptual separation of the two.

Flanagan and Taylor (1999) state that the perception of loudness is more strongly influenced by the perceived distance than sound pressure level. For instance if two sounds of equal sound level, but different direct-to-reverberant ratios are compared, participants might perceive the example with a lower direct-to-reverberant ratio as being louder. This is related to what Zahorik and Wightman (2001) refers to as loudness constancy. The hypothesis states that loudness is computed from the sound intensity at the ear and a perceptual estimate of source distance. The explanation is analogous to size constancy in vision with changes in viewing distance (Boring, 1964).

## 5.3 Experiment 2: Judging sound source distance

The aim of this test was to investigate how listeners evaluated the capability of the enhanced system to improve the perception of source distance. The enhancement aimed

to simulate a wider range of auditory perspectives, from close to distant. The hypothesis was twofold:

1. a source placed near the centre sounds closer with the enhancements than without
2. a source placed near a wall of the simulated room sounds more distant with the enhancements than without

It was hypothesized that by increasing the level of room modes from all loudspeakers that this would help contribute to a perception of the source being closer than without the modes. Informal listening by the author before the tests indicated that this would also be true for the participants.

### 5.3.1 Method

After the participants calibrated the loudness of the two conditions, an average of their individual calibrations for each permutation was used to calibrate the system used to assess distance. This listening test involved choosing which of the two conditions, A or B, had a more distant sounding source. The listeners from Experiment 1 also participated in this experiment. The sound sources and source locations (Figure 5.8) were identical to those in the first experiment. Again the subjects did not see the location of the source and had to rely exclusively on aural impressions.

The 12 permutations were repeated three times in this listening test for a total of 36 trials. Conditions A and B were randomized such that for half of the examples, A represented the enhanced system, and for the other half of the examples, B represented

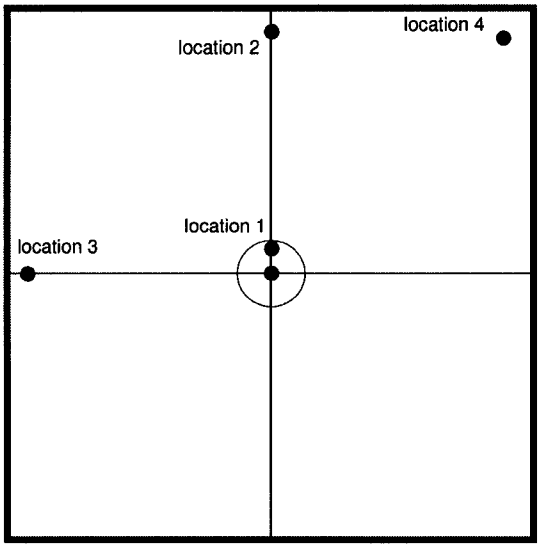


Figure 5.8: Source locations used to evaluate distance in the second test. Note: The participants did not see this display on their monitor.

the enhanced system. The participant had already ensured that the two conditions were equal in loudness in the first part, thus they had to make only the judgement of distance based on cues other than loudness. Since differences in loudness will affect the perception of distance (Blauert, 1997; Flanagan and Taylor, 1999; Zwicker and Fastl, 1999), the first part of the test helped to eliminate this variable.

### 5.3.2 Results

Table 5.1 illustrates the percentage of responses indicating that the enhanced version provided the closest sounding source, for a sound source located at a position near the listening position.

Table 5.2 illustrates the percentage of responses indicating that the enhanced version provided the most distant sounding source, for a sound source located at positions near

| Sound source position | Percussion | Speech | Acoustic Guitar |
|---|---|---|---|
| near the centre | 95.65% | 82.61% | 95.65% |

Table 5.1: Percentage of responses indicating that the enhanced system provided the closest sounding source when the source is near the centre of the simulated room for three different sounds. For all results $p < .01$; responses have to be greater than 65% to have significance at $p < .05$.

the room boundaries.

| Sound source position | Percussion | Speech | Acoustic Guitar |
|---|---|---|---|
| near the front wall | 100% | 100% | 100% |
| near the side wall | 100% | 100% | 100% |
| near the right front corner | 100% | 95.65% | 78.26% |

Table 5.2: Percentage of responses indicating that the enhanced system provided the most distant sounding source when the source is near the front wall, the side wall and the right front corner of the simulated room for three different sounds. For all results $p < .01$; responses have to be greater than 65% to have significance at $p < .05$.

### 5.3.3 Discussion

Results indicate strongly that the subjects chose the enhanced version as providing the most extended range of source distances from those perceived as the most distant to the closest sounding ones. The enhancements not only made the more distance location sound more distant but it also made the close location sound closer. From this we can conclude that the enhancements provide a stronger impression of distance control than was found in the original multichannel processor. We can be sure that loudness differences have been eliminated through individual listener calibration of loudness for each example in the previous experiment, and as such this most likely did not play a

strong role in distance perception. It is interesting to note that, although all results were significant, there was a slight reduction in the number of responses indicating the enhanced provided the most distant source for the guitar in the front corner. This may perhaps be related to the steady-state nature of the source. There is a similar but smaller reduction for the speech source in the front corner. Again because this source represents a mixture of steady-state and transient characteristics, we see the results falling between percussion (transient) and guitar (more steady-state). One way to improve the algorithm might be to include a module that would determine the temporal characteristics of the sound source and apply appropriate processing.

## 5.4 Experiment 3: Judging the presence of a wall effect

This experiment was performed to investigate whether listeners perceived the simulations of the boundary effect to be similar to the sound of a wall near a sound source. It also served as a training preparation for the fourth listening test. It was hypothesized that by increasing the low frequency energy and by increasing the apparent source width (ASW) through the use of fuzzy sources, that participants would have the perception that a wall exists behind or beside the sound source.

### 5.4.1 Method

For the "wall effect" listening test, listeners were instructed to choose which of A or B sounded most like there was a wall near (beside or behind) the sound source. For this part of the test the four sound source locations were chosen to be close to the boundaries of the simulated room (see figure 5.9):

1. 90° near the left side wall
2. 45° near the left front corner
3. 0° near the front wall
4. 45° near the right front corner

Again the listeners and sound sources were identical to the previous experiments.

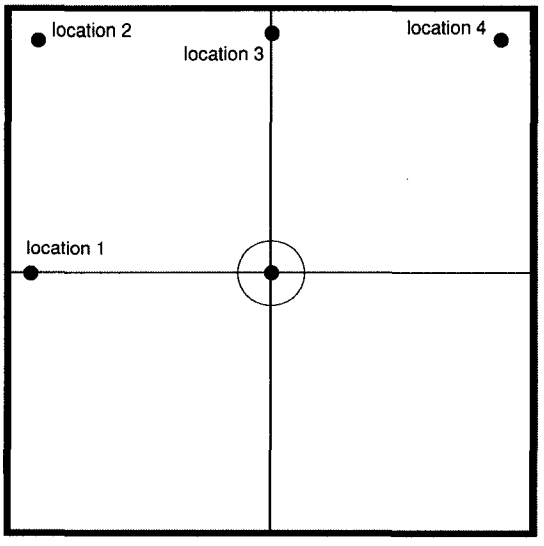


Figure 5.9: Source locations used to evaluate the presence of simulated room boundaries. Note: The subjects did not see this display on their monitor.

Please note that the subjects did not see the locations of the source and only used the aural impressions to indicate the presence of walls.

The assignment of A and B to the enhanced and non-enhanced versions was randomized so that for half of the trials, A represented the enhanced and for the other half, A represented the non-enhanced. This was done so that listeners would not learn which label (A or B) corresponded to which condition, and therefore would have to rely on what was heard for each example rather than what was presented in previous examples. The test was fully-factorial for each participant.

### 5.4.2 Results

From the test data, it was found that 97% ($p < .0001$) of the responses indicated that listeners perceived the simulations of the boundary effect to be similar to the sound of a wall near a sound source.

### 5.4.3 Discussion

It has been shown that loudspeaker location in a real room significantly affects the perceived timbre of the loudspeaker (Bech, 1994; Olive et al., 1994). From this we can conclude that the room boundary likely has an effect on the perception of the source. In experiments conducted by Shively and House (1996), it was found that different types of material create significantly different perceived boundary effects when testing loudspeakers in the interior of an automobile. From this type of research it is clear that

the boundary effect is audible to listeners and should be modelled in room simulation systems. The current listening test shows that the system provides a strong impression of a source being near a room boundary.

## 5.5 Experiment 4: Evaluating the dynamic properties of the system

This test served as a general evaluation of the enhanced system versus the non-enhanced system. Subjects were asked to decide which of two auditory presentations best illustrated the apparent source location and movement visible on the screen of the graphical interface. Figure 5.10 shows the view of the limiting boundaries of the simulated room with source and listener positions indicated by red and blue dots respectively.

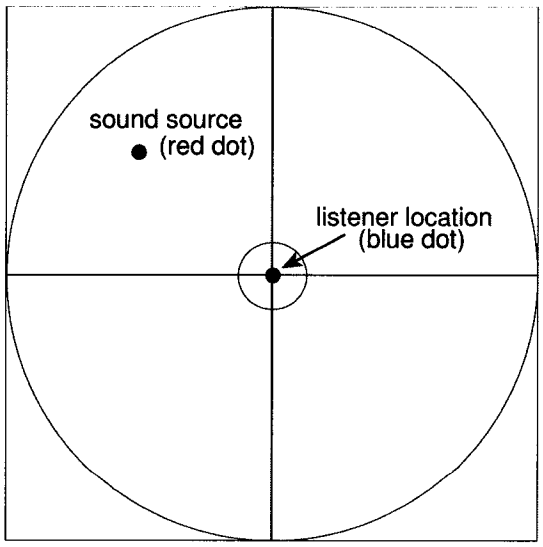


Figure 5.10: An illustration of the limiting boundaries of the simulated room with source and listener positions indicated by red and blue dots respectively.

### 5.5.1 Method

In this fourth test, the subjects were given a visual representation of the simulated room on a computer screen (see figure 5.10), and asked to use the mouse to move a red dot (representing the sound source) freely within the four boundaries of the visible space surrounding the listener (blue dot). They were instructed to explore and evaluate the movement of the sound source across different trajectories, such as from the centre of the room to the front wall, and compare two conditions, A and B, for each particular trajectory. The two conditions were:

1. enhanced,

2. non-enhanced

Their task was to decide which of the two auditory conditions best illustrated the visual representation of the sound source position and movement. Participants compared the two conditions using four different anechoic sound sources:

- percussion
- speech
- acoustic guitar
- electric guitar

The sound sources were identical to the previous tests with the addition of a fourth sound source: electric guitar, also a monophonic, anechoic recording. The electric guitar

sample was a sustained melody played on the lower strings, providing a relatively steady-state musical sample.

The participants were the same as in the previous experiments.

It was ensured that the loudness of the sound field did not change according to the distance from the centre of the room. The level of the direct sound was not altered, despite changes in location. Through extensive listening and calibration by the author, the other elements of the system were adjusted so as not to change the perceived level of the sound field. A decrease in loudness as a sound source moves away from the centre of the room can help to create a stronger impression of increasing distance, therefore this cue was eliminated so that subjects could only rely on the changing spatial, temporal and spectral balance. The enhanced system that has been developed uses loudness-varying cues to magnify the impression of changing distance and perspective. However, since loudness balancing is not included in the non-enhanced system, it was decided to disable this feature for this experiment, as well as all previous tests. In designing this system for a production environment, loudness cues would be included to make the impression of changing distance that much stronger and more natural. After the subjects had completed the first three listening tests it may be assumed that they had been trained to a certain degree to listen for source distance and proximity to a room boundary.

### 5.5.2 Results

All subjects, without exception, chose the enhanced version as the one that best illustrates the visual representation of the sound source location and movement. Some subjects commented afterwards that the enhanced version gave more control over the sound source distance and auditory perspective.

### 5.5.3 Discussion

By changing the auditory perspective according to the visual representation of the source on an interface, it is possible to create a strong impression of source location and movement. It is apparent from the test that listeners prefer the proposed system of control because it provides strong auditory feedback of the visual user interface. The dynamic adjustment of parameters in a coordinated way other than simple panning is a much better method of control of source location. Results indicate that the enhancements provide a greater range of control over perceived sound source location in a simulated room than the non-enhanced system, and that the enhanced version better corresponds to the location illustrated on the graphical interface.

## 5.6 Experiment 5: Tracking versus static reflections in distance perception

The next two listening tests (Experiments 5 and 6) investigated more closely the control of perceived sound source distance within the proposed system. The first test in this set investigated the influence of early reflection patterns on distance perception. The second test investigated if the entire system could provide a robust perception of small changes in source distance. Relative judgments of source location were made as opposed to absolute judgments. In terms of music and film sound production, there is no real need to produce sources at absolute distances; relative placement is more practical.

This test was developed to provide some evidence that dynamic, tracking early reflections are important to increase the control of perceived sound source distance. The hypothesis is that early reflections that track the location of the direct source will allow subjects to better judge source distance than static reflections. Michelsen and Rubak (1997) and Nielsen (1991) have indicated that early reflection patterns play an important role in auditory distance perception. They suggest that spatial perception results from a decoding of the reflection patterns, where the decoding is based on pattern recognition (Theile (1980), cited in Michelsen and Rubak (1997)).

### 5.6.1 Participants

Four unpaid subjects volunteered to perform the listening tests. Three of the subjects were audio engineers or researchers in audio and the fourth subject was a formally trained

musician.

The following two listening tests (Experiments 5 and 6) were performed on the system to evaluate the effectiveness of tracking reflections and to ensure the strength of control of the perception of source distance in a 5-channel environment. The duration of each test was approximately 20 minutes.

### 5.6.2 Method

For the listening test two conditions were tested to assess listeners abilities to judge source distance:

- static reflections
- tracking reflections

The early reflections of a room sized 6.5m (long) by 4.5m (wide) were synthesized with a 4th order two-dimensional image model representing the horizontal plane only. In total 40 early reflections were generated. The static reflections were generated by a commercially available 5-channel reverberation system (i.e., T.C. Electronic System 6000) for a room of the same size. For this test only the direct sound and early reflections were used, and the level of the direct sound remained constant.

Two different source locations were compared using three different sound sources. The two locations are indicated in figure 5.11. Both positions were at an angle of 0° in front of the listener, one near the front wall and the other close to the centre of the simulated room.

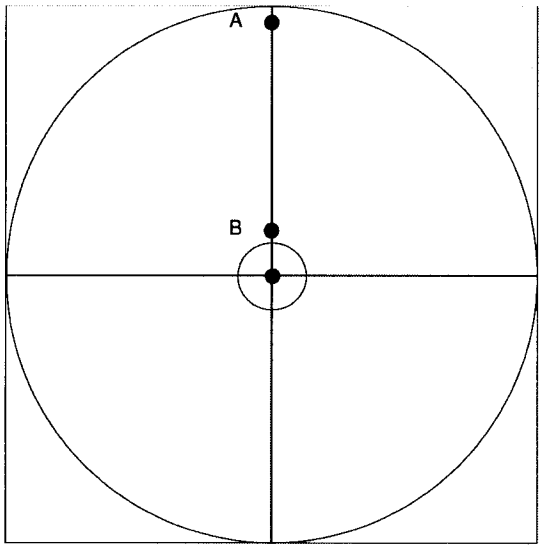


Figure 5.11: Sound source locations compared in the first listening test. It should be noted that listeners did not have a visual representation of the sound source location.

The sound sources used for the test were:

1. speech female

2. percussion (bongos)

3. electric guitar

All sounds were monophonic anechoic recordings.

Half of the examples presented subjects with static reflections and the other half employed tracking reflections. Subjects were first asked if there was any change in the sound source distance between A and B. If they heard a change then they were then asked to choose which of A or B had the most distant sound source, where A and B represented the near and far positions.

### 5.6.3 Results

90.7% of the responses indicated that listeners could hear a changing source position between A and B with dynamic reflections. This level is significant since $p < .01$.

94.5% of the responses indicated that listeners did not hear a difference in sound source distance between A and B in the condition with static reflections. This value is significant with $p < .01$.

58.4% of responses indicated that listeners heard the more distant source as sounding more distant with tracking reflections. Although this may appear to be a low percentage it is in fact significant with $p = 0.039$.

### 5.6.4 Discussion

From this we can conclude that listeners hear a difference in sound source location with tracking reflections. Tracking reflections are therefore important in a multichannel room simulation system since they help listeners form a clearer perception of the location of the sound source. This is consistent with the literature about the role of early reflection patterns in the perception of source distance. Therefore tracking reflections are a critical component of any room simulation system concerned with providing control over the perceived distance of a sound source.

Although the results were significant there may be ways to improve the significance level. One way might be to test on a larger room model. Being a relatively small room it provides an effective test of the influence of tracking reflections on perceived

distance. The diffusion of reflections might also improve the effect, although this has the disadvantage of being processor intensive.

## 5.7 Experiment 6: Calibrating relative perceptual distances

The aim of this test was to determine if the system provides a strong impression of changing distance for relatively small changes in source position. Furthermore we want to test whether four different sound source positions, representing different distances in the same angular location, will be perceived as having the same relative positions as represented by the visual positions (figure 5.12).

### 5.7.1 Method

For this test, the entire system was employed. Four source locations were tested in six different pairs. The room size that was simulated was 6.5m (long) x 4.5m (wide). Subjects were presented with a sound source and were allowed to switch freely between two positions, A and B. They were asked to choose which of A or B had a more distant sound source.

The subjects who participated in Experiment 5 also participated in the current one, and the same sound sources were used.

Figure 5.12 shows the four positions that were tested. The figure shows positions A,

B, C, D, and these were compared in pairs, with the pairs being represented as A and B in the listeners interface. The listeners did not have a visual representation of the sound source locations. The pairs of positions were compared as follows:

1. A and D
2. A and C
3. A and B
4. B and D
5. B and C
6. C and D

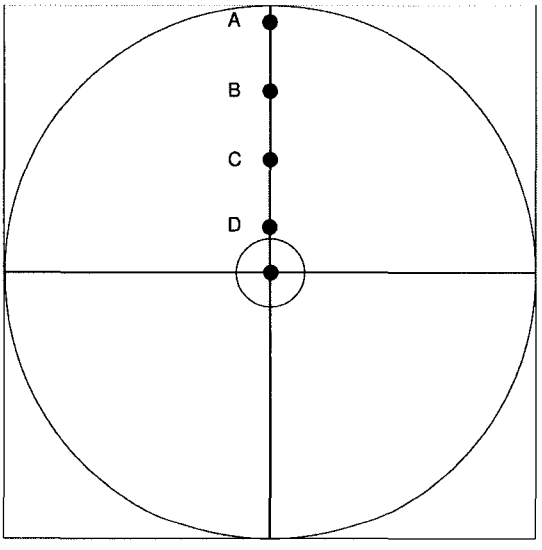


Figure 5.12: Four discrete positions that were tested against each other to determine the strength of the perceived distance. It should be noted that listeners did not have a visual representation of the sound source location.

The test was fully factorial for each listener. The number of permutations is 18, representing 6 source location combinations and 3 sound sources.

### 5.7.2 Results

Location pairs:

1. A and D: Listeners were 100% correct in indicating that position A was more distant than D.
2. A and C: Listeners were 100% correct in indicating that position A was more distant than C.
3. A and B: Listeners were 100% correct in indicating that position A was more distant than B.
4. B and D: Listeners were 100% correct in indicating that position B was more distant than D.
5. B and C: Listeners were 83% correct in indicating that position B was more distant than C.
6. C and D: Listeners were 94% correct in indicating that position C was more distant than D.

In all tests $p < .01$, therefore we can conclude that these results are significant.

### 5.7.3 Discussion

The results of this test indicate that the described system provides a strong perceptual indication of small changes in distance for the 0° angular position. Furthermore the relative auditory distances are identical to the relative visual distances in the GUI. Listeners clearly perceive whether the sound source is moving closer or further away from them. We can also be sure that as the source is moved further away, that the control functions provide a continuous adjustment of the physical parameters supporting the intended effect.

## 5.8 Experiment 7: Perception of lateral phantom images with and without simulated early reflection patterns

When producing music and film soundtracks for five-channel (3/2) reproduction, there is a possibility to create the impression of sound sources emanating not only from the front but also from the side and behind the listener. Because of the wide aperture between front and rear loudspeakers (80-90°) in a five-channel configuration (ITU-R BS.775) ITU-R (1994), and due to the loudspeakers projecting sound to one side of the listener's head, it is difficult to create a stable lateral sound image using only interchannel level differences.

Simple constant-power pair-wise panning, perhaps the most commonly used panning

algorithm, is not as reliable for lateral positions as it is for the front. Localization of phantom images between loudspeakers in front of a listener (summing localization) has been studied by Blauert (1997) and Griesinger (2002). Phantom images can be produced in conventional two-channel stereo when the loudspeakers are symmetrically arranged (normally ±30°) across the median plane in front of a listener. Phantom images are perceived between loudspeakers emitting identical signals, equidistant from a listener. By changing the amplitude of one loudspeaker relative to the other, it is possible to change the perceived location of the phantom image. Figure 5.13 illustrates the signal path from each speaker to the ears of the listener. Normally constant-power (or sine/cosine) panning is used to position phantom images in stereo reproduction systems.

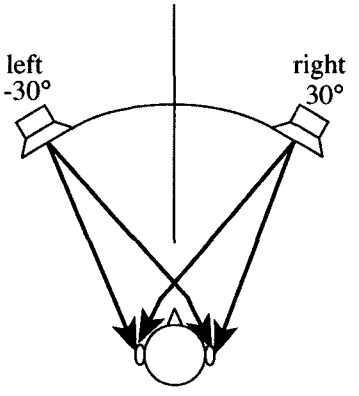


Figure 5.13: Signal paths from loudspeakers to ears in a conventional stereo (2/0) system.

Studies in auditory perception have determined that localization of real sound sources is accomplished through interaural time differences (ITD), interaural level differences (ILD), and head-related transfer functions (HRTF) (Moore, 1997). It was reported by Blauert (1997) that there is a large variability in localization of real sources for lateral

positions as shown in Figure 5.14. This may in part be explained by the "cone of confusion," where front-back confusions are made in the localization of lateral sources that are slightly ahead or behind the listener. There is considerable localization blur for sources at ±90° which indicates some variability in source localization and a perception of the source being wider than it is physically. Even before phantom source localization is considered it needs to be emphasized that there is a substantial localization blur even for real sources.

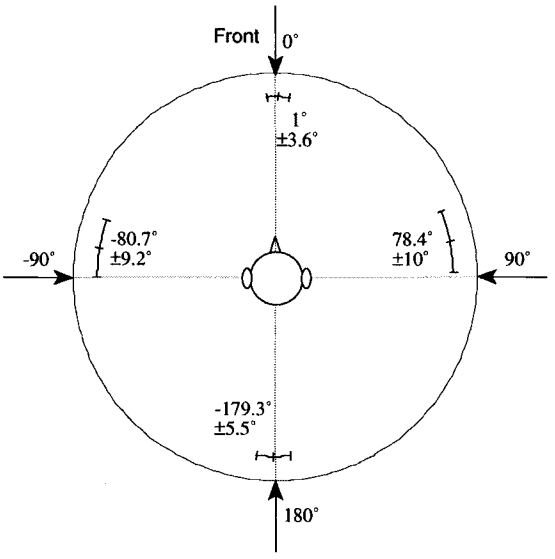


Figure 5.14: Localization blur and localization in the horizontal plane for loudspeakers emitting white-noise pulses of 100ms duration at $\phi = 0°, \pm 90°$, and 180°. Arrows indicate the location of the sources. (Diagram adapted from Blauert (1997), p. 41 , after Preibisch-Effenberger (1966) and Haustein and Schirmer (1970))

In an investigation of localization of lateral phantom images Theile and Plenge (1977) found a large variations in the perceived location of lateral phantom images. Damaske and Ando (1972) have also indicated that it is generally difficult to create lateral phan-

tom images. No matter what the amplitude difference between the loudspeakers, the ipsilateral ear will always receive the signals from both loudspeakers at a higher amplitude and with less time delay than the contralateral ear. Perhaps the auditory system is not relying on ILD and ITD for localization because there is little change in the interaural level or time difference as a source is power panned from a loudspeaker at 30° to one at 120°. Pinna filtering may play a larger role in this case. Figure 5.15 illustrates the sound propagation paths from side loudspeakers to the ears of a listener seated in the centre of the array. Whenever pairs of adjacent loudspeakers are not arranged symmetrically across the median plane, it is difficult to rely on interaural differences to determine location.

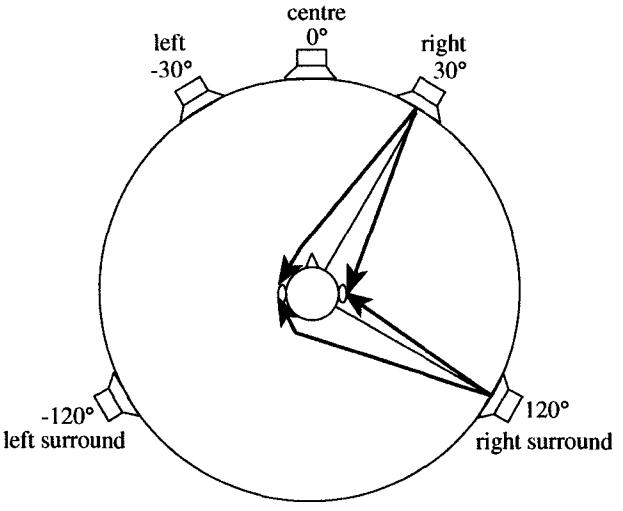


Figure 5.15: Signal paths from loudspeakers to ears for a lateral power-panned phantom image in a 3/2 channel system.

Despite the claimed difficulty in generating lateral phantom images, Lund (2000)

has suggested that the stability and localization certainty of lateral phantom images in a five-channel reproduction environment can be improved by rendering early reflection patterns from all loudspeakers. As this was a preliminary investigation, the current paper presents additional experimental methods and results to help understand the relationship between simulated early reflection patterns and the perception of lateral phantom images.

Rather than suggest the addition of more channels on the side, the challenge here is to work with the existing ITU loudspeaker layout standard and find better methods of signal processing to improve imaging. Because of the difficulty in producing stable lateral phantom images, early reflection patterns need to be generated from all loudspeakers that "support" and stabilize the perceived location of the direct sound. It was deemed that adding early reflection patterns, calculated according to the direct source location, would produce a less ambiguous perceived source location because there will be more information for the auditory system to judge the location of the direct source. This is consistent with Casey (1998) who states that there is an inverse relationship between the complexity of a stimulus and its respective perception. The more complex a stimulus is, the easier it is for a listener to draw conclusions about it. With a simple stimulus it is more difficult to determine information about the source through the perception of the physical event. A power-panned source with rendered early reflections will create a much more complex stimulus than the source without reflections, and perhaps will create a more stable perceived source location.

The research questions for the experiment are thus:

- how well does constant-power panning work for the localization of lateral sources?
- how do simulated early reflection patterns (rendered according to source location in a given room model) influence the localization of lateral phantom images in a five-channel environment?
- can lateral localization in a five-channel system be improved by rendering early reflection patterns to support the direct sound?
- does an image model provide the best location support for the direct sound?
- was the image model chosen, the correct one?
- can listeners localize sounds in a five-channel environment with simulated early reflection patterns alone (no direct sound)?
- is listener localization accuracy related to certainty about the image location?

### 5.8.1 Participants and Test Duration

Fifteen subjects, all students and faculty lecturers in sound recording at McGill, participated in the experiment. The test duration was around 30–45 minutes, and all participants were remunerated for their time. All participants completed the test twice at two different times. For this reason, it is considered a repeated measures experiment. During the test, subjects could mute the sound at any time to take a short break if needed. Each subject went through a few examples before beginning the test to familiarize themselves

with the procedure and to have a chance to ask questions for clarification of the task. Participants had normal hearing although it was not tested.

### 5.8.2 Method

Although other panning algorithms exist such as vector-base amplitude panning (Pulkki, 2001), Ambisonics (Gerzon, 1992b), and polarity-restricted cosine (Martin et al., 1999a), the most commonly used panning algorithm, constant-power (sine/cosine), was chosen to position the direct sound. Sources were positioned to locations to the side of the listening position, i.e., between left and left-surround and between right and right-surround loudspeakers. Gains ($g$) applied to the front and rear loudspeaker signals were derived from Equations 5.1 and 5.2 where $\phi$ = intended angle (from 30°–120°):

$$g_{front} = cos(\phi - (\pi/6)) \tag{5.1}$$

$$g_{rear} = sin(\phi - (\pi/6)) \tag{5.2}$$

Early reflection patterns were generated using a two-dimensional (horizontal plane) image model (Allen and Berkley, 1979) for each source location and panned to the loudspeakers using the same sine/cosine panning algorithm. Although the direct sound always originated from at most two loudspeakers, the early reflections were rendered through all five loudspeakers. Early reflections, generated in real-time using a desktop computer, were calculated up to 4th-order in the horizontal plane only, giving 40 reflec-

tions in total. Reflections were low-pass filtered to approximate wall and air absorption, with each successive order of reflections having a lower cut-off frequency. Specifically the cut-off frequencies were:

- 1st order reflections: $f_c = 16$ kHz
- 2nd order reflections: $f_c = 12$ kHz
- 3rd order reflections: $f_c = 10$ kHz
- 4th order reflections: $f_c = 8$ kHz

The delay time and gain according to signal propagation for each reflection was calculated according to the distance travelled, where:

$$delay = distance/c;\ c = 344m/s \tag{5.3}$$

$$gain = 1/distance \tag{5.4}$$

Participants were given a graphical user interface on a computer display representing a top view of the room (Figure 5.16) and were asked to position a black dot to a location on the interface that best represents the location of the sound source. This is nearly identical to the method used by Martin et al. (1999b) for testing perceived location of phantom images, where listeners were asked to place a dot on a graphical interface to indicate the perceived location of a phantom image.

Preliminary versions of the interface included dots to indicate ±30°, ±60°, ±120° with corresponding markers placed around the room to give listeners points of reference. It was decided that these dots and markers would influence the listeners by introducing some unwanted, artificial quantization in the responses and it might also give some information away about what was being tested. Markers around the room would require the subjects to turn around for a visual indication of the marker locations, thus making it difficult to make an accurate decision about their auditory perception.

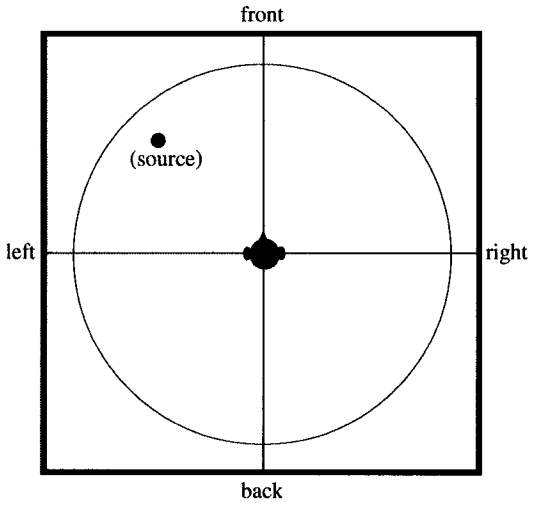


Figure 5.16: Graphical interface in which participants indicated, with a black dot, the perceived location of the sound image.

As Evans (1998) states, having listeners place a dot to indicate source location has the advantage of not requiring the listener to translate a direction into a verbal response. The method of having a graphical interface with which to indicate source location may be better suited than having a pointer because participants can indicate rear locations

without turning around to look. This is especially important because the method of generating lateral phantom images is sensitive to head orientation.

Participants heads were not fixed, but they were instructed to indicate perceived source location while facing forwards. A marker was placed at 0° front centre on the ceiling near the acoustically transparent curtain to indicate this location in the room. By facing the computer screen directly, they could be sure that they were facing forwards. They were shown a mark on the ceiling above their head indicating the centre of the room.

The reproduction level was approximately 65 dB SPL, linear frequency weighting, slow time weighting.

### 5.8.3 Independent variables

For the test three monophonic, anechoic sound sources were used:

1. speech, female Danish
2. percussion, bongos
3. electric guitar

The speech and percussion samples were taken from the Bang & Olufsen *Music for Archimedes* compact disc (Bang & Olufsen, 1992), which were recorded in an anechoic chamber. The guitar sample was recorded using a microphone placed very close to the guitar amplifier, providing isolation from the room. The choice of sound sources

represents the fulfillment of three criteria: a transient source (percussion), a more steady-state source (electric guitar), and a third source that exhibits a mixture of transient and steady-state characteristics.

Three "room effect" conditions were presented:

1. dry, anechoic source, positioned using constant-power panning (sin/cos)
2. dry sound with early reflections (4th order image model, horizontal plane only)
3. early reflection pattern only

Eight source locations were tested (Figure 5.17):

| | | | |
|---|---|---|---|
| 1. | $-45°$ (left) | 5. | $45°$ (right) |
| 2. | $-65°$ | 6. | $65°$ |
| 3. | $-80°$ | 7. | $80°$ |
| 4. | $-100°$ | 8. | $100°$ |

It was decided to treat the left and right sides separately and not to average the two together. In this way, the test results might illustrate differences in responses for left and right, indicating asymmetrical room effects in the listening room and/or differences in perception of the two sides.

The intended angle $45°$ was chosen to represent a location just outside of the front loudspeaker at $30°$. As it was known by the author from initial experimenting and listening that this position tended to pull towards the closest loudspeaker, it was decided that this would be a crucial location to test.

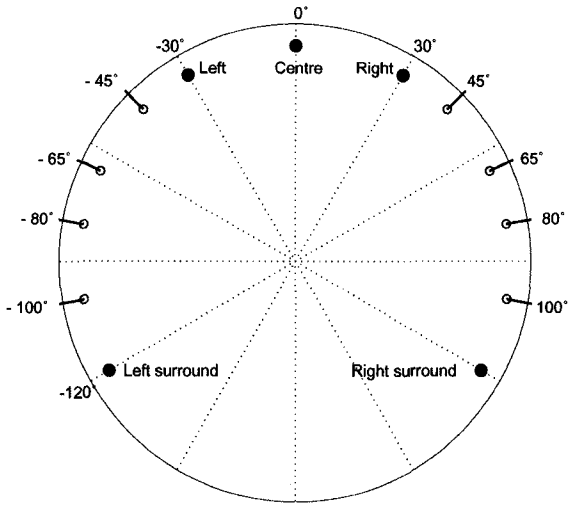


Figure 5.17: The eight lateral source locations tested, relative to the five speaker locations (indicated by black dots).

The intended angle 65° was chosen to determine if listeners could differentiate between 45° and a more lateral location. Also it was found through informal listening before the test that this location was slightly peculiar in that the sound source appeared to pull out from the loudspeakers towards the listening position. It was thought that this might influence the perception of azimuth differently for the "room effect" conditions.

Finally, to test listeners' abilities to differentiate between locations slightly ahead and behind 90°, positions at 80° and 100° were chosen. We wanted to find out if these locations would be quantized to the same lateral auditory event. Or if an intended angle of 100° would be perceived to pull back towards the rear loudspeaker.

Three sound sources, three room effect conditions, and eight locations makes a total of 72 permutations. The listening test was fully factorial for each participant, with

random order presentation of the variable combinations. The test was repeated once for each listener.

To provide additional information about the listeners' indications of source location, they were also asked to rate their "certainty" of the source location on a five-point scale, where 5 is most certain, and 1 is least certain. Specifically they could choose one of the five following levels of certainty:

- 5 – no doubt
- 4 – high
- 3 – good
- 2 – some
- 1 – poor

Lund (2000) used a similar listening test design where listeners were asked to rate certainty of source location on a five point scale. There they also rated "robustness" and "diffusion" to determine a composite "consistency score". By using the certainty rating scale in addition to asking for a specific indication of perceived location, it is possible to compare certainty with accuracy. It was hypothesized that as certainty ratings increase, accuracy of localization would also increase.

Listeners response times to place the dot were also recorded without their knowledge. It was deemed that response times should correlate negatively with certainty ratings, and

that it might provide additional information about the difference between localization of direct sound and direct with reflections.

### 5.8.4 Results

An 8(intended angle) * 3(room effect) * 3(sound source) repeated measures analysis of variance of within-subjects effects was performed with the results presented in Table 5.3.

The following independent variables and interactions had significant effects on the dependent variable perceived angle (see also Table 5.3):

- intended angle $[F(14, 194) = 93.943, p < .001]$
- sound source $[F(14, 54) = 3.107, p = .023]$
- intended angle * room effect interaction $[F(28, 390) = 2.079, p = .001]$
- intended angle * sound source interaction $[F(28, 390) = 5.186, p < .001]$

The following independent variables and interactions had significant effects on the dependent variable response time:

- intended angle $[F(14, 194) = 7.002, p < .001]$
- intended angle * room effect interaction $[F(28, 390) = 2.097, p = .001]$
- intended angle * sound source interaction $[F(28, 390) = 1.609, p = .028]$
- intended angle * room effect * sound source interaction $[F(56, 782) = 1.519, p = .010]$

The following independent variables and interactions had significant effects on the dependent variable certainty:

- intended angle $[F(14, 194) = 10.906, p < .001]$
- room effect $[F(4, 54) = 2.524, p = .050]$
- intended angle * room effect interaction $[F(28, 390) = 2.191, p = .001]$
- intended angle * sound source interaction $[F(28, 390) = 1.766, p = .011]$
- intended angle * room effect * sound source interaction $[F(56, 782) = 1.410, p = .029]$

| Source | Dependent Variable | Wilk's Lambda | F | Hypothesis dF | Error dF | Sig. |
|---|---|---|---|---|---|---|
| A: intended angle | perceived angle | .017 | 93.943 | 14.000 | 194.000 | < .001 |
| | response time | .441 | 7.002 | 14.000 | 194.000 | < .001 |
| | certainty | .313 | 10.906 | 14.000 | 194.000 | < .001 |
| B: room effect | perceived angle | .964 | .250 | 4.000 | 54.000 | .909 |
| | response time | .815 | 1.456 | 4.000 | 54.000 | .228 |
| | certainty | .710 | 2.524 | 4.000 | 54.000 | .050 |
| C: sound source | perceived angle | .661 | 3.107 | 4.000 | 54.000 | .023 |
| | response time | .791 | 1.675 | 4.000 | 54.000 | .169 |
| | certainty | .936 | .455 | 4.000 | 54.000 | .768 |
| A * B | perceived angle | .757 | 2.079 | 28.000 | 390.000 | .001 |
| | response time | .755 | 2.097 | 28.000 | 390.000 | .001 |
| | certainty | .747 | 2.191 | 28.000 | 390.000 | .001 |
| A * C | perceived angle | .531 | 5.186 | 28.000 | 390.000 | < .001 |
| | response time | .804 | 1.609 | 28.000 | 390.000 | .028 |
| | certainty | .788 | 1.766 | 28.000 | 390.000 | .011 |
| B * C | perceived angle | .873 | .962 | 8.000 | 110.000 | .469 |
| | response time | .832 | 1.325 | 8.000 | 110.000 | .238 |
| | certainty | .940 | .433 | 8.000 | 110.000 | .899 |
| A * B * C | perceived angle | .865 | 1.048 | 56.000 | 782.000 | .382 |
| | response time | .813 | 1.519 | 56.000 | 782.000 | .010 |
| | certainty | .825 | 1.410 | 56.000 | 782.000 | .029 |

Table 5.3: The results of an 8(intended angle) * 3(room effect) * 3(sound source) repeated measures analysis of variance. Tests of within-subjects effects dependent variables: perceived angle, response time, and certainty.

Figures 5.18 to 5.29 illustrate the perceived locations (with 95% confidence intervals) of the sound sources relative to the intended angle and loudspeaker positions, plotted according to sound source and room effect condition. To make the results easier to read only four intended locations (not all eight) are plotted on each figure, alternating between a plot with $\pm 45°$ and $\pm 80°$ and a plot with $\pm 65°$ and $\pm 100°$. Figures 5.18 and 5.19 indicate the perceived locations of all three sound sources for the anechoic condition. Figures 5.20 and 5.21 indicate the perceived locations of all three sound sources for the anechoic with early reflections condition. Figures 5.22 and 5.23 indicate the perceived locations of all three sound sources for the condition of early reflection patterns only.

Figures 5.24 and 5.25 indicate the perceived locations of the speech source for all three room effect conditions: anechoic, anechoic with reflections, and reflections only. Figures 5.26 and 5.27 indicate the perceived locations of the percussion source for all three room effect conditions. Figures 5.28 and 5.29 indicate the perceived locations of the electric guitar source for all three room effect conditions.

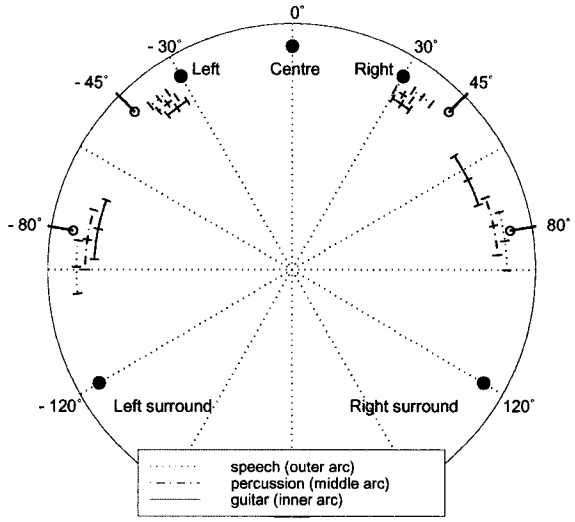


Figure 5.18: Plot of mean perceived locations of sound images (with 95% conf. int.) for anechoic condition and all sounds. Black dots indicate the loudspeaker locations. Intended image locations were ±45° and ±80°.

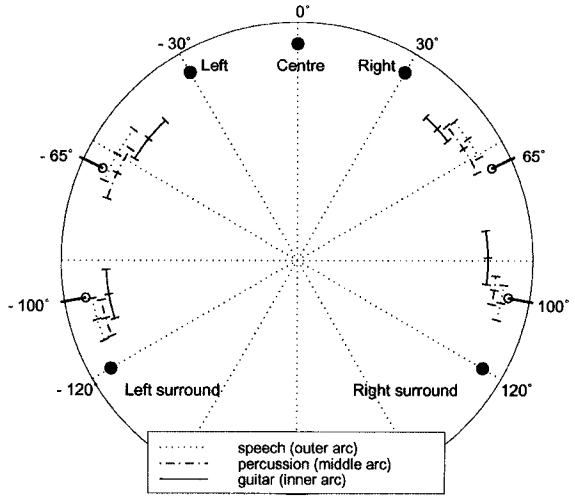


Figure 5.19: Plot of mean perceived locations of sound images (with 95% conf. int.) for anechoic condition and all sounds. Black dots indicate the loudspeaker locations. Intended image locations were ±65° and ±100°.

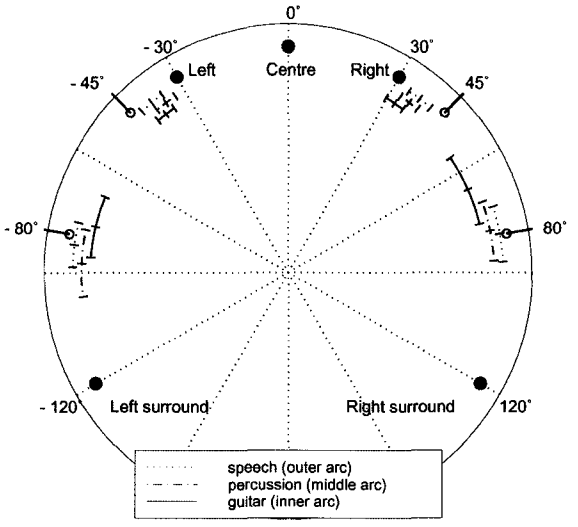


Figure 5.20: Plot of mean perceived locations of sound images (with 95% conf. int.) for anechoic with reflections condition and all sounds. Black dots indicate the loudspeaker locations. Intended image locations were ±45° and ±80°.

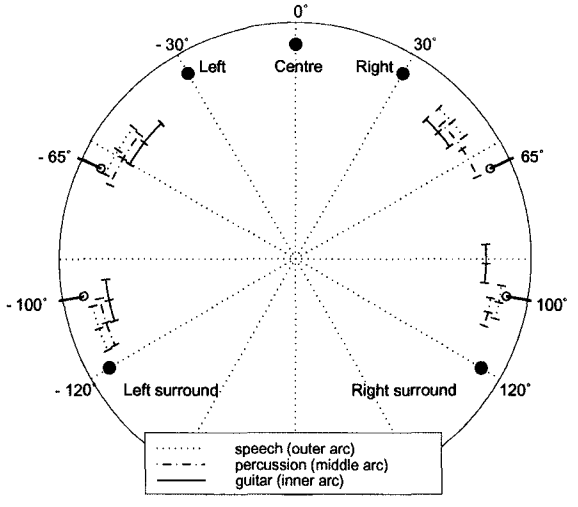


Figure 5.21: Plot of mean perceived locations of sound images (with 95% conf. int.) for anechoic with reflections condition and all sounds. Black dots indicate the loudspeaker locations. Intended image locations were ±65° and ±100°.

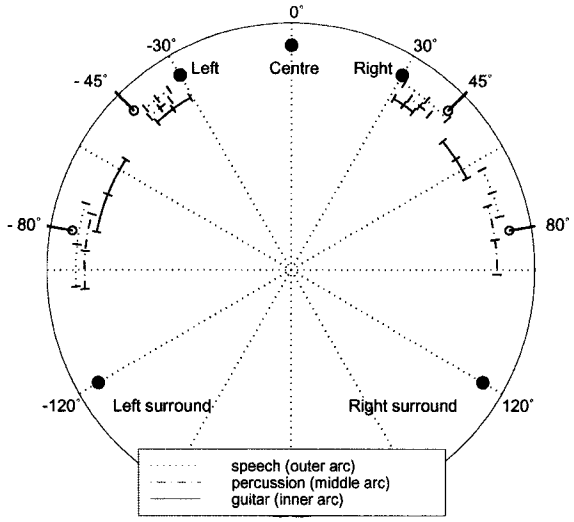


Figure 5.22: Plot of mean perceived locations of sound images (with 95% conf. int.) for reflections only condition and all sounds. Black dots indicate the loudspeaker locations. Intended image locations were ±45° and ±80°.

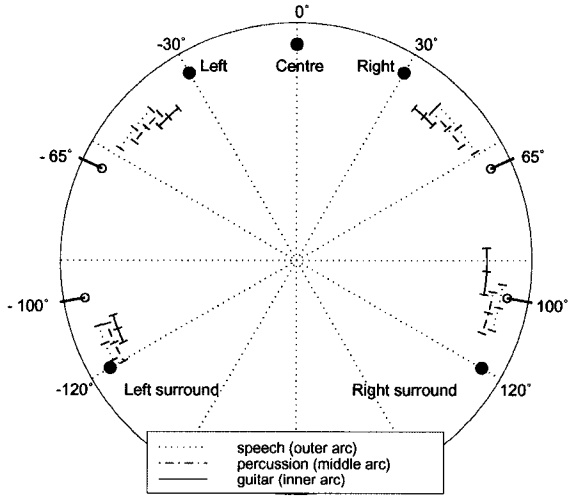


Figure 5.23: Plot of mean perceived locations of sound images (with 95% conf. int.) for reflections only condition and all sounds. Black dots indicate the loudspeaker locations. Intended image locations were ±65° and ±100°.

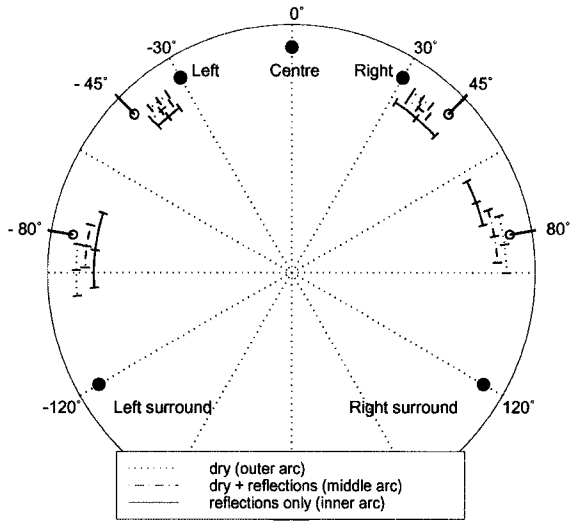


Figure 5.24: Plot of mean perceived locations of sound images (with 95% conf. int.) for speech and all room effect conditions. Black dots indicate the loudspeaker locations. Intended image locations were ±45° and ±80°.

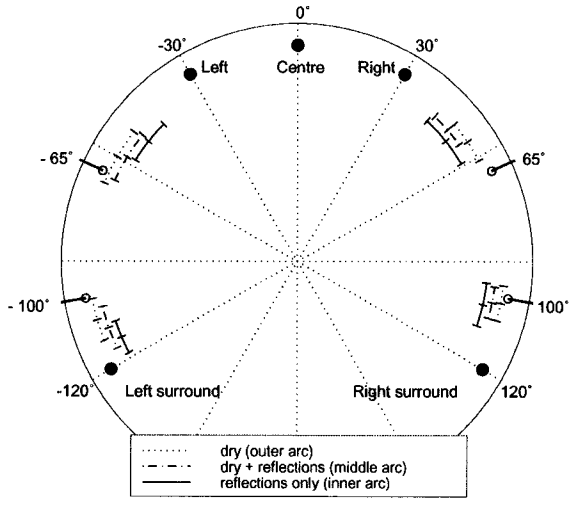


Figure 5.25: Plot of mean perceived locations of sound images (with 95% conf. int.) for speech and all room effect conditions. Black dots indicate the loudspeaker locations. Intended image locations were ±65° and ±100°.

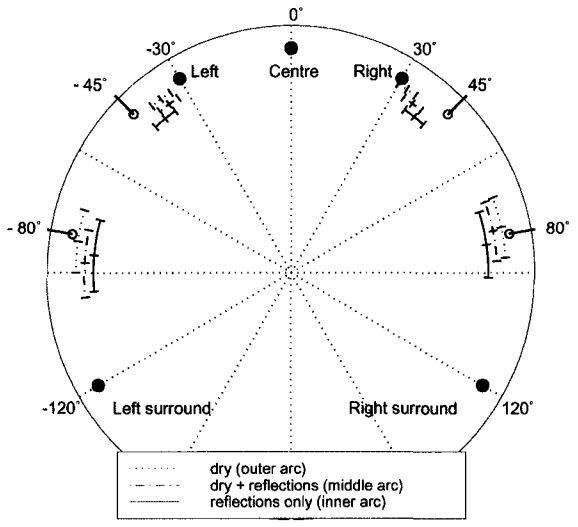


Figure 5.26: Plot of mean perceived locations of sound images (with 95% conf. int.) for percussion and all room effect conditions. Black dots indicate the loudspeaker locations. Intended image locations were ±45° and ±80°.

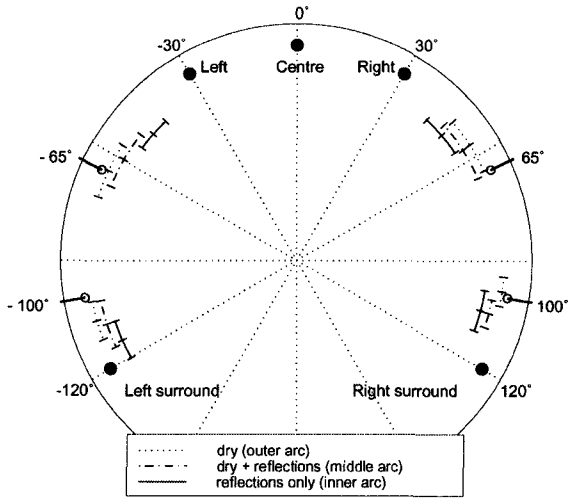


Figure 5.27: Plot of mean perceived locations of sound images (with 95% conf. int.) for percussion and all room effect conditions. Black dots indicate the loudspeaker locations. Intended image locations were ±65° and ±100°.

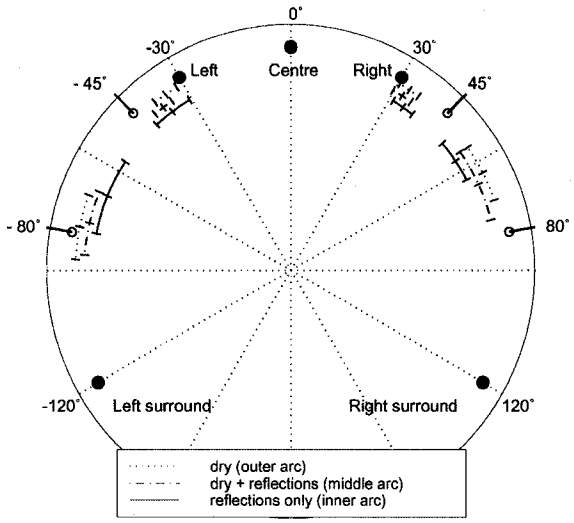


Figure 5.28: Plot of mean perceived locations of sound images (with 95% conf. int.) for electric guitar and all room effect conditions. Black dots indicate the loudspeaker locations. Intended image locations were ±45° and ±80°.

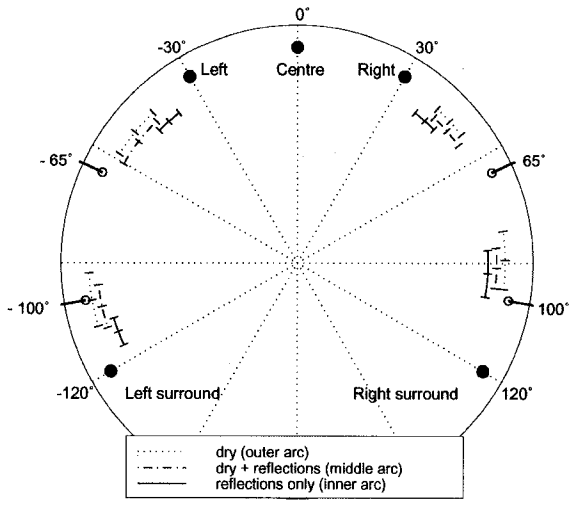


Figure 5.29: Plot of mean perceived locations of sound images (with 95% conf. int.) for electric guitar and all room effect conditions. Black dots indicate the loudspeaker locations. Intended image locations were ±65° and ±100°.

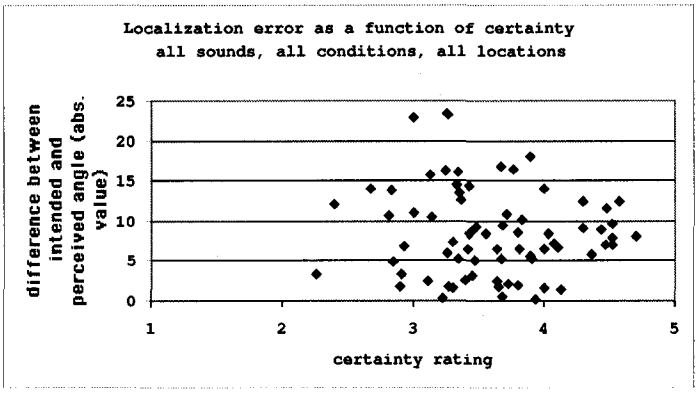


Figure 5.30: Plot of localization error (absolute value) as a function of certainty for all three sound sources, all eight locations and all conditions; $r = -0.09$, $p = 0.22$.

Figure 5.30 shows a plot of localization error as a function of certainty, $r = -0.09$, $p = 0.22$.

Figures 5.31–5.33 show the mean certainty ratings (with 95% confidence intervals) for each of the three room effect conditions, plotted according to intended angle.

Figures 5.34–5.36 show the mean response time (with 95% confidence intervals) for the three room effect conditions, plotted as a function of intended angle.

Figure 5.37 show the response time as a function of certainty for all three sound sources, all eight locations, and all three conditions, $r = -0.80$, $p < .001$.

Figure 5.38 shows the average response time as a function of example number.

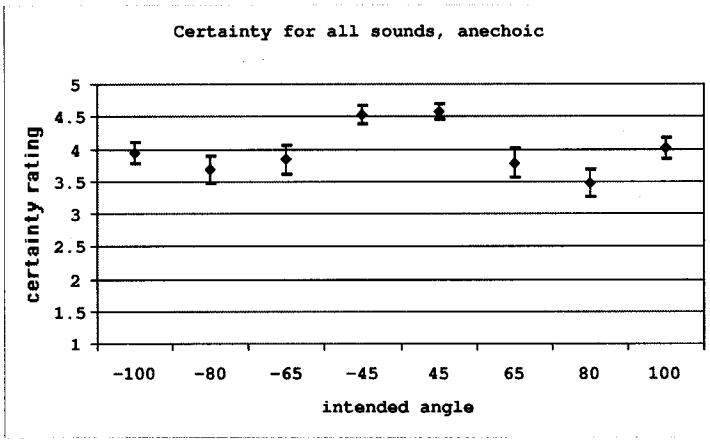


Figure 5.31: Plot of mean certainty rating (with 95% conf. int.) as a function of location for all three sound sources for all eight locations in the anechoic sound condition.

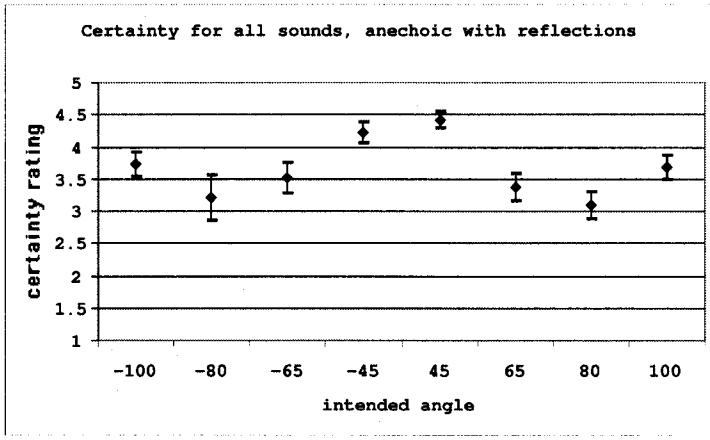


Figure 5.32: Plot of mean certainty rating (with 95% conf. int.) as a function of location for all three sound sources for all eight locations in the anechoic sound with reflections condition.

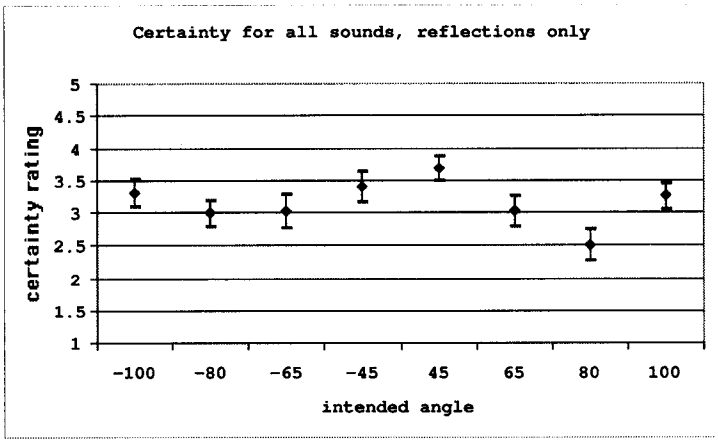


Figure 5.33: Plot of mean certainty rating (with 95% conf. int.) as a function of location for all three sound sources for all eight locations in the reflections only condition.

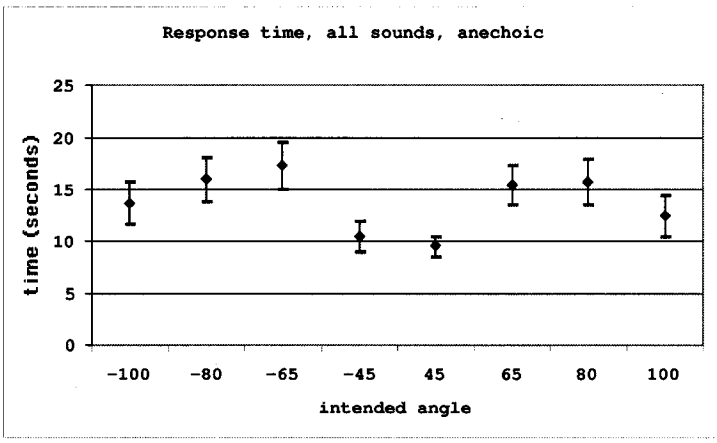


Figure 5.34: Plot of mean response time (with 95% conf. int.) as a function of location for all three sound sources for all eight locations in the anechoic sound only condition.

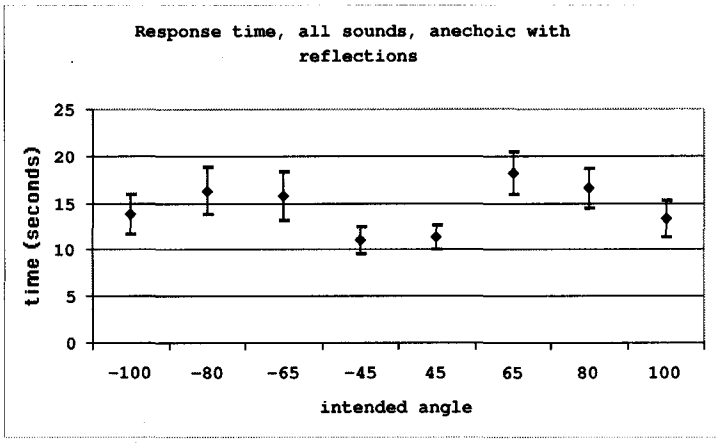


Figure 5.35: Plot of mean response time (with 95% conf. int.) as a function of location for all three sound sources for all eight locations in the anechoic sound with reflections condition.

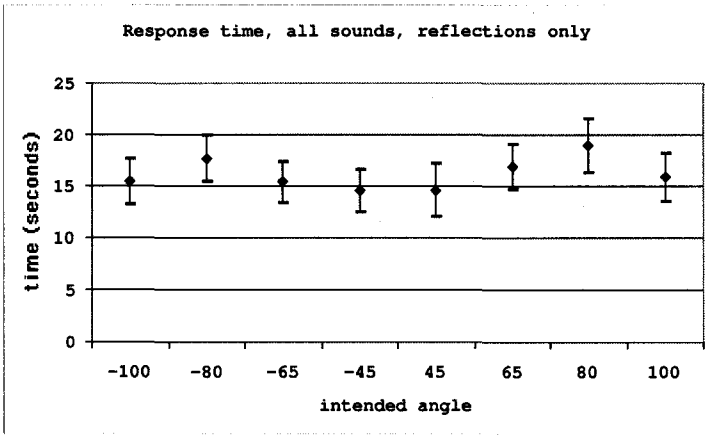


Figure 5.36: Plot of mean response time (with 95% conf. int.) as a function of location for all three sound sources for all eight locations in the reflections only condition.

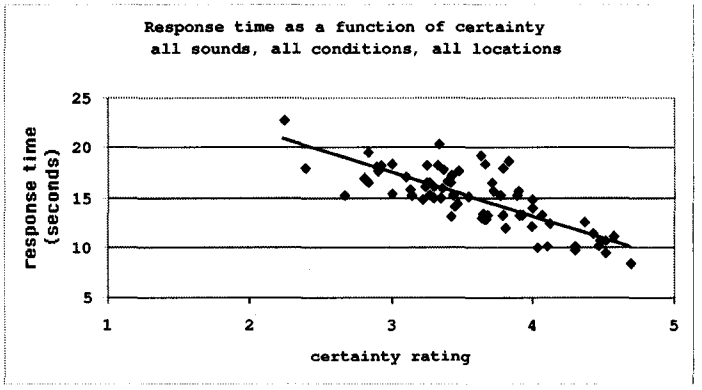


Figure 5.37: Plot of response time as a function of certainty rating for all three sound sources, all eight locations, and all three conditions; $r = -0.80$, $p < .001$.

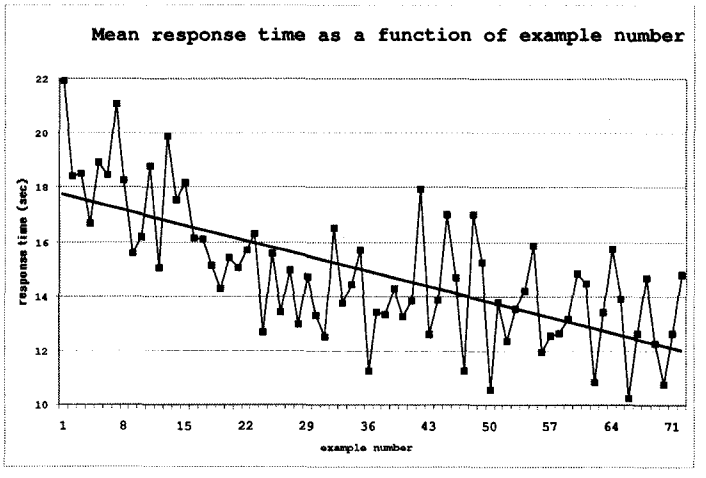


Figure 5.38: Plot of the response time (averaged across all listeners) as a function of example number; $r = -0.678$, $p < .001$.

### 5.8.5 Discussion

As the results indicate the independent variable intended angle had significant effects on the perceived angle. It is expected that listeners should perceive the eight distinct intended locations. Intended angle also had significant effects on certainty and response time. Figures 5.31–5.36 illustrate the differences in response time and certainty as a function of intended location for the three room effect conditions.

We would expect there to be a negative correlation between response time and certainty of source location which there was (Figure 5.37). As was consistent with Lund (2000), the room effect condition had a significant effect on certainty.

The localization blur of intended locations ±65°, ±80°, and ±100° is quite close to what Blauert (1997) reported for real sources (see Figure 5.14). Often the localization blur is similar for the anechoic and anechoic with reflections conditions.

There was a significant interaction between intended angle and room effect on the dependent variable perceived angle. This can be observed in Figures 5.24-5.29 where the room effect variable has differing effects on localization blur depending the intended angle. This may be related to the frequency-dependent image blur for lateral phantom images where the high frequency band is perceived as originating from the front loudspeaker and the low frequency band is perceived as coming from the rear loudspeaker, for an intended location of about ±90°. The localization blur of the phantom image produced laterally changes according to the intended location. The room effect condition appears to have different effects depending on intended angle of the source. For instance

adding early reflections to the dry source sometimes has the effect of increasing the blur and other times it decreases the blur depending on intended angle. Furthermore the perceived location is generally "pulled" more and more towards the loudspeaker closest to the intended angle as the room effect changes from dry to dry with reflections to reflections only. What seems to be a common occurrence is that for locations ahead of $\pm 90°$ the image is pulled forward, for those behind $\pm 90°$ it is generally pulled to the rear.

It was hypothesized that by adding simulated early reflection patterns to the constant-power panned anechoic sound, that this would decrease the localization blur of the phantom image. Generally it does not seem to be the case although there are exceptions such as for speech at $\pm 80°$, $\pm 65°$ and $100°$ right in Figures 5.24 and 5.25, for electric guitar at $100°$ right in Figure 5.29. The question is: why not? Some possible explanations are as follows.

The production of lateral phantom images using constant-power panning is much more difficult than it is in a traditional two-channel stereo setup due to the fact that the signals from both speakers are reaching the ipsilateral ear with a shorter time delay and with greater intensity that the contralateral ear (Figure 5.15). As such, when a source is panned between the side loudspeakers, the image is not necessarily perceived as having the same "width" and different frequency bands may be localized differently. For instance it has been found empirically by the author (and also reported by some participants) that for a single source panned to a lateral position (e.g., $90°$), the high frequency band (typically above 1500 Hz) is localized near the front loudspeaker and the

low frequency band is localized towards the side or rear. As such the exact location of the phantom image can be ambiguous. Participants involved in the listening test were required to place a single dot to indicate the perceived location of the phantom image. They may have had trouble deciding how to represent a wide phantom image with a small dot. Perhaps the relative balance of the high and low frequency bands influenced their decision.

The results also indicate evidence of what Gerzon (1992b) refers to as the "detent" effect where sound images that are intended to be near a loudspeaker but not directly at a loudspeaker (e.g., left 45°) are pulled towards that loudspeaker (left 30°). This is apparent in Figures 5.18, 5.20, and 5.22 for all sound sources and room effect conditions. The speech sound source, for all room effect conditions, is generally perceived to be closest to the intended angle of ±45° of the three sounds. In addition it may be noted that the ±45° intended locations resulted in smaller standard deviations of the perceived locations for all sound sources and all room effect conditions, as compared to other intended locations. The certainty ratings are higher (Figures 5.31–5.33) and the response times are generally shorter (Figures 5.34–5.36) for the intended location of ±45°. The intended angle 65° also exhibited the "detent" effect as perceived locations were pulled towards the loudspeakers at ±30°.

From the results it is apparent that different types of sources are localized differently. This is consistent with the findings of Blauert (1997) and Griesinger (2002) who found that the perceived position of a phantom image is strongly dependent on the spectrum of the source. For instance the electric guitar source was in many cases (e.g., Figures

5.18 – 5.23) localized more towards the front loudspeaker than the other two sound sources. This may be explained by the fact that it has a higher spectral centroid than both the percussion and speech samples. Due to the nature of pinna filtering, we are more sensitive to high frequencies located in front.

Figure 5.39 illustrates the magnitude spectrum of the electric guitar sound sample, calculated using a 1024-point Hanning window DFT averaged over the entire sound file. Comparing the magnitude spectrum of the speech (Figure 5.40) and percussion (Figure 5.41) samples, it is possible to see that there is considerably less energy (approximately $-30$dB) in the frequency band from 2–3 kHz, than there is in the electric guitar sample. The temporal characteristics of the sound sources may also have influenced the judgment of location.

As can be seen from Figures 5.22 and 5.23, participants were generally able to localize the sound sources with the early reflection patterns only (no direct sound). In this case the electric guitar is not as accurately localized as the speech and percussion.

In Figure 5.24 it is evident that by adding reflections to the dry source, the perceived location is closer to the intended location for $\pm 80°$.

In comparing the localization accuracy with certainty, it was found that there was not a significant correlation between the variables (Figure 5.30). From this we can conclude that confidence in source location does not always translate into accurate or consistent localization ability.

From Figure 5.38 it is apparent that response time decreased over the course of the experiment. It may be that listeners were getting more accustomed to the test procedure

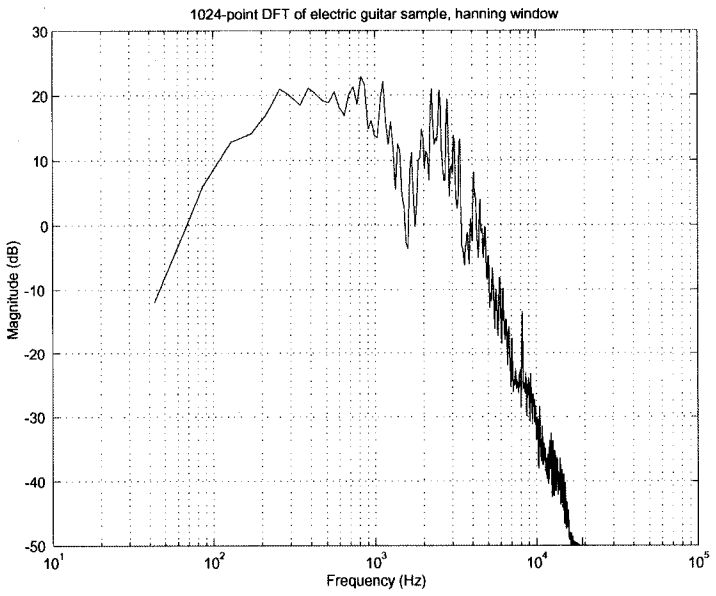


Figure 5.39: Magnitude spectrum of the electric guitar sample, averaged over the length of the sample.

and could more quickly identify where they heard the sound source. Since the test was presented in randomized order for each participant, we can be sure that this effect is not due to sound source, intended location or room effect condition. It would be interesting to see if the results of perceived location would change substantially if listeners performed the test numerous times.

Some subjects reported that a few examples were perceived as having increased elevation over others. Although it is not known which conditions elicited this perception, it would be an interesting study to pursue.

A couple of participants indicated that the visual modality may have influenced their location judgments. It may have been difficult for listeners to translate from

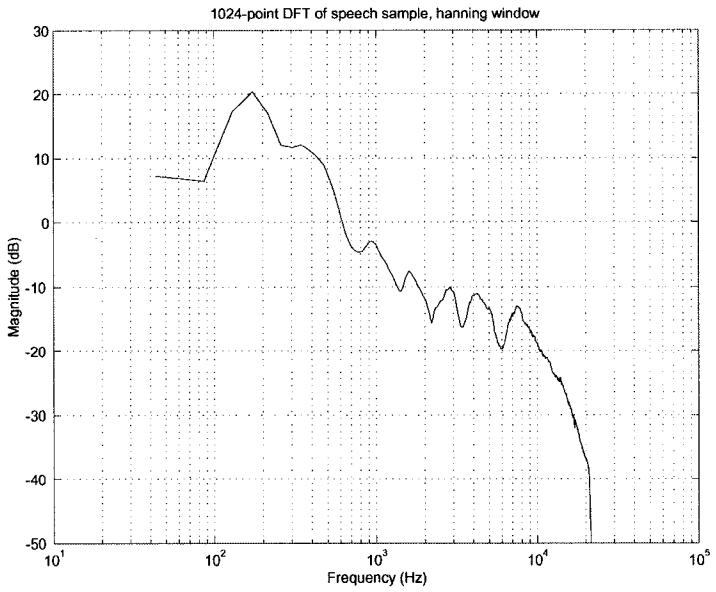


Figure 5.40: Magnitude spectrum of the speech sample, averaged over the length of the sample.

the horizontal auditory plane to the visual interface which was in front of them and angled nearly vertical. Perhaps markers around the room and corresponding markers on the visual interface would have helped listeners make more accurate and consistent responses. There may also have been some confusion about where true $\pm 90°$ was. One participant indicated that there was a tendency to use the outside limit of peripheral vision to determine $\pm 90°$, and this may have skewed the judgments slightly forwards for the intended angle of $\pm 80°$.

One participant noted after the test that in a couple of the examples the sound source was perceived to be in close proximity to the centre of the room, pulling out from the opaque curtain. Although it is not known for certain which conditions elicited

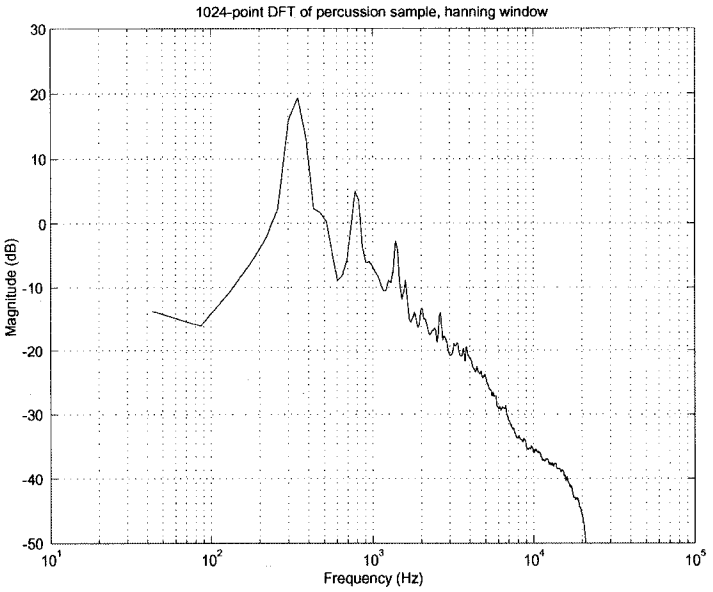


Figure 5.41: Magnitude spectrum of the percussion sample, averaged over the length of the sample.

this perception, we would guess that it would be the condition with a dry sound source panned to + or $-65°$.

As with Blauert (1997), it should be noted here that the results cannot confirm whether the deviations in localization result in errors in participants judgments about direction or whether they accurately reflect their auditory localization.

### 5.8.6 Conclusions

The results of the investigation indicate that the room effect condition had a significant interaction with intended location influencing the perceived location and localization

blur of the source, and it had a significant effect on the certainty rating. Sound source type had significant effects on the perceived location, underlining the importance of using different program material for localization tests. Early reflection patterns alone rendered according to intended source location result in fairly accurate localization of sources.

One possible way to increase the influence of reflections on localization of a dry source might include diffusion of the reflections. Since the room boundaries are treated essentially as mirror reflectors (with some attenuation due to distance and filtering) in this model, simulated diffusion might create more realistic sounding reflections.

In addition to this a more complex room geometry might also prove effective, although it becomes more processor intensive as the room model becomes more complex. The image model chosen was a rectangular room model, arbitrarily chosen. It is possible that a more suitable room model exists that would provide reflections that are more "supportive" to localization. Alternatively a perceptually based reflection pattern (i.e., one that is not based on the physical dimensions of a real room) might prove to have benefits as well. As Lund (2000) points out, appropriate digital signal processing may enable the positioning of sources within a simulated 5-channel sound field that is more robust perceptually than multichannel microphone techniques in a real room.

The constant-power panning algorithm might also be hindering the localization ability. There may be a more appropriate method of panning the direct sound and reflections. Perhaps there needs to be a different panner for the direct than for the reflections. Additionally a multi-band panner might prove to be useful as well. This would apply

different panning curves to the high and low frequency bands. The reason for using such an algorithm would be to compensate for the perception of high frequencies located near the front and low frequencies located near the rear when producing a phantom image to the side.

Future listening tests might include testing the perceived image width and extent, for the three different room effect conditions. An alternate method of testing localization might be to ask listeners to position the sound source at specific locations as indicated by a graphical interface. In this way they would have control over the panner and could change the perceived location of the sound. It is not known whether this would result in different responses than those found here. There was also the issue raised of perceived elevation of the source for some examples. More work needs to be conducted to investigate the perception of source elevation from loudspeaker arrays with an elevation of 0°.

# Chapter 6

# Conclusions

We have a lifetime of listening skills which we bring to the surround sound environment. We are accustomed to hearing sounds from all directions, whether they be a conversation behind us, a car honking a block away, or a performing musician in front of us with a sense of the acoustic space all around us. The implementation of a room modelling system in a surround sound environment should not require the user to develop a new set of listening skills to make sense of it; the technology needs to meet the user or the artist (ideas developed after reading Buxton (1992, 1997)). Since we normally hear sounds originating from any location surrounding us in a natural setting, it makes sense to want to re-create this sensation in the reproduction of recorded music and film soundtracks. Surround reproduction can make music sound more natural and also provide a sonic experience that is not possible in a natural acoustic setting. But according to Nielsen (1991) we do not hear as accurately as we think we do, and this might require an exaggeration of perceptual cues for control of auditory perspective.

With current multichannel audio technology, every approach to room simulation and multichannel reverberation involves a compromise in one way or another. For example, because of the complexity of calculations involved in making accurate physical models of acoustic spaces with digital waveguides, it is not possible to provide a dynamic multichannel room simulator based on this technique with current technology. Although IIR-based reverberation provides flexibility to tune numerous parameters of the reverberation, it is still an approximation (and gross simplification) of the impulse response and reverberation decay of a real room. Many IIR-based reverberation units employ source image models and ray-tracing methods to simulate early reflections patterns. As Savioja et al. (1995) have suggested, these methods fall short in the area of modelling the low frequency resonances of real acoustic spaces. Recent innovations have introduced sampling reverberators that use FFT-based convolution of impulse responses of real acoustic spaces. Although these units offer truly realistic sounding room response and reverberation, the amount of dynamic control over the spatial characteristics is limited.

Multichannel room simulation devices and reverberators need to take into account human perception of the multichannel environment. At this point in time, mathematical models of acoustic spaces, no matter how complex, still need to be fine-tuned by an expert with a trained ear. A room simulation system that is based on research into the perception of spatial attributes could offer more intuitive, effective control over auditory perspective by manipulating physical parameters that provide the strongest perceptual changes. The system has to perform complex, dynamic balancing of multiple component layers of sound and has to take into account spatial, temporal and spectral masking. It

must also consider the wavering attention of the listener whose ear is constantly seeking changes in the auditory stimuli.

## 6.1 Applications of the system

The system of dynamic control proposed herein has many applications in the production of music and film sound for 5-channel (3/2) reproduction. The system has been designed to be the next generation of multichannel reverberation offering increased control of sound source location within a simulated acoustic environment. For instance in most non-classical recordings, instruments are recorded using close microphone techniques with artificial reverberation being added in the mixdown process. The system can be used to place the recorded sound sources within a simulated acoustic space and have more efficient control over their locations. Each sound source can be placed at a virtual location within the virtual, simulated space, with its own unique azimuth and apparent distance. If the music producer wishes to make a recorded sound source appear to move forward within the mix, e.g., for a solo, it is very easy to do this, and all of the simulated acoustic components will change dynamically as the intended location of the recorded source is changed.

During a film shooting the dialogue that is recorded can not always be used in the final mix for the film because of high levels of background noise. In this case it is necessary to have the actors re-record the dialogue in a quiet studio environment. Unfortunately the "room sound" that was originally recorded on-site is now lost and has to be recreated.

For instance when an actor walks across a room, the perceived room acoustics will change according to the location of the actor. When the dialogue is re-recorded the original changes in perceived acoustics need to be simulated and the proposed system could effectively provide the control that is needed for such a simulation. If there were computer algorithms that could track the visual location of an actor, this might provide a way to automate the auditory perspective control.

## 6.2 Original contributions

This thesis provides evidence that overall gain, boundary effects, and modal effects are perceptually relevant components of a room modelling algorithm, allowing control over sound source distance and proximity to a wall impression in a multichannel environment. The system of control that has been developed provides the user with effective control over auditory perspective within a simulated acoustic space through the movement of a representative dot on a visual interface.

The system described here provides automatic, dynamic, coordinated and complex control of numerous physical parameters corresponding to source location, to give a strong perceptual impression of sound source distance control including proximity to a room boundary, for multichannel reproduction. The system incorporates physical and perceptual models to keep processor requirements down and still make the impression strong and unambiguous. It has been found to provide much stronger auditory feedback of the visual interface for source movement and location, as compared to a commercially

available state-of-the-art multichannel reverberation and signal processing unit. Listening tests confirm the need for dynamic control of auditory perspective in a multichannel environment.

The proposed system is the first room simulation algorithm to suggest a perceptual model of the boundary effect. The use of fuzzy sources, simulated room modes, dynamic equalization, dynamic level control, and tracking early reflection patterns all work together to help provide effective, efficient and intuitive control of the boundary effect. Commercial manufacturers (e.g., T.C. Electronic) are beginning to realize that the boundary effect is important to model especially for in film sound applications, and have begun to include this component in their multichannel room simulation algorithms. The use of tracking reflections has been proven through listening tests to be necessary for control of source distance. It has been proven that listeners can localize sound sources with only tracking reflections (without direct sound) for lateral positions.

Although the room mode simulation was not tested directly it helps provide a sense of the room dimensions, and as a source moves there is a natural comb filtering provided by the algorithm. This helps to give a cue regarding source motion, but on its own does not contribute directly to source localization. It is important to model since that is what is found in real acoustic spaces and it is often overlooked by developers of reverberation algorithms.

The mapping of source location to the processing of physical parameters presented here is unique. By simply moving a dot on a visual interface, representing the source location, hundreds of parameters of perceptual and physical models are changed au-

tomatically and dynamically. The combined total of all parameters changing together presents a strong impression of source location and movement.

The thesis presents original listening tests that evaluate the perceptual strength of the control functions as related to auditory perspective within the multichannel environment. The results of listening tests conducted on various aspects of the algorithm confirm that the control of physical parameters provide unambiguous perceptual cues as a function of source location.

## 6.3 Future work

It was found by Martin et al. (2001) that adding jitter to the source azimuth in a room simulation system helps make the perceived source width more realistic, especially for monophonic recorded sources. Future development of the current system will include modulation of the source location according to the level of the signal. Although this presents a completely artificial widening of the perceived source image, it helps model the natural movement of a performer on stage who always moves slightly while performing. Even a stationary instrument such as a piano will be perceived to have a sense of movement in a recording, since different notes are radiated in different directions. This will also take advantage of the auditory system's natural inclination to focus attention on that which is changing.

The perceived location of the reverberation also must change. It has been found by the author that by standing outdoors in a location surrounded by distant hills or

tall buildings, it is possible to hear the echo (and reverberation) of a loud shout or hand clap pan around the listening position (i.e., change azimuth). This has also been experienced by the author in informal experiments in an empty concert hall. There is a subtle sense of the reverberation flowing around the room, changing both elevation and azimuth. Artificial reverberation devices try to mimic this by providing some spatial modulation, which turns out to sound unnatural and therefore clearly audible as an undesirable artifact.

Through informal experimentation by the author it was found that modulation of the reverberation signal was useful to further exaggerate the dynamics of the musical signal. By using the level of the musical signal as a control signal for the modulation of feedback level of allpass filters it was possible to change the perceived width of the reverberation according to signal level. Modulation of any aspect of sound field has been found to be optimum when derived from the input signal itself. Modulation that is based on sinusoids or random noise creates an artificial and non-musical effect which detracts from the program material. The technique of source-based modulation may be useful as an added effect on future developments of the system.

As was found from the listening tests, the temporal and spectral characteristics influence the perceived effect of the room simulation. An ideal room simulation system would be able to analyze in real time the incoming signal (through FFT and beat tracking) and tailor the processing according to its temporal and spectral characteristics.

In a 5-channel or surround-sound audio environment listeners may not always be seated in the "sweet spot" or exact centre of the loudspeaker array. The sweet spot

is known as the ideal listening position that is equidistant from all loudspeakers. It is necessary to enlarge the listening area to provide optimal sound-image quality for different seating positions within the loudspeaker array. One of the results of sitting off-centre is that the loudspeaker closest to the listener becomes the dominant sound source, pulling the multichannel sound image towards that particular loudspeaker because of the precedence effect (Blauert, 1997) and because of the difference in relative loudness between the loudspeakers. This can be distracting if the listener is seated near one of the surround loudspeakers. Future listening tests will be conducted to find ways to effectively increase the size of the sweet spot.

A further improvement to the system would be to model sound source directionality in a simplified way. Currently the source is treated as omnidirectional. Making the source more directional would make the spectrum of the reflection dependent on the angle of incidence, which may be considered as a type of azimuth-dependent weighting. For instance high frequencies are normally not radiated to the rear of a source, therefore reducing the amount of high frequency content in the reflections coming from behind the sources. It is believed that this would help make the simulation more realistic and effective.

As mentioned in the conclusion of the final experiment, the design of clusters of reflections that support and enhance the direct source location might be a better method of improving lateral phantom image localization. These clusters although not based on physical models of reflections in a real space could help control apparent source width (ASW) and provide additional cues about source location to compensate for constant-

power panning.

In future developments to the system, there will be control of the listener location through different types of simulated rooms. For example, one possibility would be to simulate the sound heard as a person walks down a hallway, through a lobby and into a concert hall. There would be various sound sources such as conversation, phones ringing, doors closing, and musical instruments, which would excite the acoustics of the different spaces. The simulated sound fields would have to cross-fade from one to another providing a sense of motion to the listener.

To paraphrase Duke Ellington, in the end no matter what the mathematics and physics tell us about room simulation, the auditory system is the final judge.

# Appendix A

# Parameter settings for generating fuzzy sources

Below is a table showing the parameter settings on two T.C. Electronic M3000's used to generate one fuzzy source. The two units were connected digitally in serial, with Machine 1 feeding Machine 2. In total four M3000's were used to generate the four fuzzy sources. As is apparent from the table, the reverberation is completely attenuated and the early reflection pattern was chosen to give a very dense series of echoes through the selection of a small room early reflection type.

| | Machine 1 | Machine 2 |
|---|---|---|
| Algorithm | VSS-FP | VSS-FP |
| Mix | Mix | Mix |
| Routing | Parallel | Dual Mono |
| Decay | 0.01 s | 0.01 s |
| EarlyLev | 0.0 dB | 0.0 dB |
| RevLev | -100 dB | -100 dB |
| Mix | 100% | 100% |
| Out Level | 0 dB | 0 dB |
| Rev Delay | 0 ms | 0 ms |
| Pre Delay | 0 ms | 0 ms |
| Early Type | Conf Room | Studio |
| Early Size | Medium | Small |
| Early Pos | Distant | One Pos |
| Early Bal | Center | Center |
| Hi Color | -3 | 15 |
| Lo Cut | 50 Hz | 50 Hz |
| Reverb Modulation | Off | Off |
| Space Modulation | Off | Off |

Table A.1: The parameter settings for two M3000's connected in serial (Machine 1 to Machine 2) that would produce one fuzzy source.

# Bibliography


Aarts, R. M. (1991). Calculation of the loudness of loudspeakers during listening tests. *Journal of the Audio Engineering Society*, 39:27–38.

Allen, J. B. and Berkley, D. A. (1979). Image method for efficiently simulating small-room acoustics. *Journal of the Acoustical Society of America*, 65(4):943–950.

Allison, R. F. (1974). The sound field in home listening rooms influence of room boundaries on loudspeaker power output. *Journal of the Audio Engineering Society*, 22:314–319.

Angus, J. A. S. (1997). The behaviour of rooms at low frequencies. In *106th Convention of the Audio Engineering Society, Preprint 4421*, Munich.

Bang & Olufsen (1992). *Music for Archimedes.* CD B&O 101.

Bech, S. (1994). Perception of timbre of reproduced sound in small rooms: Influence of room and loudspeaker position. *Journal of the Audio Engineering Society*, 42:999–1007.

Bech, S. (1998). Calibration of relative level differences of a domestic multichannel sound reproduction system. *Journal of the Audio Engineering Society*, 46(4):304–313.

Bech, S. (1999). Methods for subjective evaluation of spatial characteristics of sound. In *AES 16th International Conference on Spatial Sound Reproduction*, pages 487–504, Rovaniemi, Finland. Audio Engineering Society.

Bech, S. and Zacharov, N. (1999). Multichannel level alignment, Part III: The effects of loudspeaker directivity and reproduction bandwidth. In *106th Convention of the Audio Engineering Society, Preprint 4909*, Munich. Audio Engineering Society.

Begault, D. R. (1987). *Control of Auditory Distance.* PhD thesis, University of California, San Diego.

Begault, D. R. (1991). Perceptual effects of synthetic reverberation on 3-d audio systems. In *91st Convention of the Audio Engineering Society, Preprint 3212*, New York. Audio Engineering Society.

Begault, D. R. (1992). Binaural auralization and perceptual veridicality. In *93rd Convention of the Audio Engineering Society*, San Francisco. Audio Engineering Society.

Begault, D. R. (2000). *3-D Sound for Virtual Reality and Multimedia.* National Aeronautics and Space Administration, Ames Research Center, Moffett Field, CA. Originally published in 1994, but now available in PDF format from `http://human-factors.arc.nasa.gov/ihh/spatial/personnel/begault.html`.

Beranek, L. L. (1996). *Concert and Opera Halls: How They Sound.* Acoustical Society of America through the American Institute of Physics, Woodbury. Leo Beranek.

Berg, J. (2002). *Systematic Evaluation of Perceived Spatial Quality in Surround Sound Systems.* PhD thesis, School of Music, Luleå University of Technology, Sweden.

Blauert, J. (1997). *Spatial Hearing: The Psychophysics of Human Sound Localization.* MIT Press, Cambridge, Mass., revised edition.

Blesser, B. (2001). An interdisciplinary integration of reverberation. In *111th Conference of the Audio Engineering Society, Preprint 5468*, New York.

Boone, M. M., Verheijen, E. N., and Van Tol, P. F. (1995). Spatial sound-field reproduction by wave-field synthesis. *Journal of the Audio Engineering Society*, 43(12):1003–1012.

Boring, E. G. (1964). Size constancy in a picture. *American Journal of Psychology*, 77:494–498.

Borish, J. (1984). Extension of the image model to arbitrary polyhedra. *Journal of the Acoustical Society of America*, 75:1827–1836.

Bregman, A. S. (1990). *Auditory Scene Analysis: The Perceptual Organization of Sound.* MIT Press, Cambridge, Mass.

Bronkhorst, A. W. and Houtgast, T. (1999). Auditory distance perception in rooms. *Nature*, 397:517–520.

Butler, R. A. (1969). Monaural and binaural localization of noise bursts vertically in the median sagittal plane. *Journal of Auditory Research*, 3:230–235.

Butler, R. A., Levy, E. T., and Neff, W. D. (1980). Apparent distance of sounds recorded in echoic and anechoic chambers. *Journal of Experimental Psychology: Human Perception and Learning*, 6:745–750.

Buxton, W. (1992). Snow's two cultures revisited. In Jacobson, L., editor, *Cyberarts: Exploring art & technology*, pages 24–31. Miller Freeman, San Francisco.

Buxton, W. (1997). Artists and the art of the luthier. *Computer Graphics: The SIGGRAPH Quarterly*, 31:10–11.

Casey, M. A. (1998). *Auditory Group Theory with Applications to Statistical Basis Methods for Structured Audio.* PhD thesis, MIT.

Chowning, J. M. (1971). The simulation of moving sound sources. *Journal of the Audio Engineering Society*, 19:2–6.

Clifton, R. K. (1987). Breakdown of echo suppression in the precedence effect. *Journal of the Acoustical Society of America*, 82:1834–1835.

Coleman, P. D. (1962). Failure to localize the source distance of an unfamiliar sound. *Journal of the Acoustical Society of America*, 34:345–346.

Coleman, P. D. (1963). An analysis of cues to auditory depth perception in free space. *Psychological Bulletin*, 60:302–315.

Coleman, P. D. (1968). Dual role of frequency spectrum in determination of auditory distance. *Journal of the Acoustical Society of America*, 44:631–632.

Corey, J., Woszczyk, W., Martin, G., and Quesnel, R. (2001a). An integrated multidimensional controller of auditory perspective in a multichannel soundfield. In *111th Convention of the Audio Engineering Society, Preprint 5417*, New York.

Corey, J., Woszczyk, W., Quesnel, R., and Martin, G. (2001b). Enhancements of room simulation with dynamic cues related to source position. In *Proceedings of the AES 19th International Conference on Surround Sound Techniques, Technology and Perception*, pages 84–97, Schloss Elmau, Germany. Audio Engineering Society.

Damaske, P. and Ando, Y. (1972). Interaural crosscorrelation for multichannel loudspeaker reproduction. *Acustica*, 27:232–238.

Evans, M. J. (1998). Obtaining accurate responses in directional listening tests. In *104th Conference of the Audio Engineering Society, Preprint 4730*, Amsterdam.

Flanagan, S. and Taylor, V. (1999). Investigation into the relationship between subjective loudness and auditory distance perception. In *107th Conference of the Audio Engineering Society, Preprint 5049*, New York.

Ford, N., Rumsey, F., and Nind, T. (2002). Subjective evaluation of perceived spatial differences in car audio systems using a graphical assessment language. In *112th Conference of the Audio Engineering Society, Preprint 5547*, Munich.

Gardner, W. G. (1997). *3-D Audio Using Loudspeakers*. PhD thesis, MIT.

Gardner, W. G. and Martin, K. D. (1995). HRTF measurements of a KEMAR. *Journal of the Acoustical Society of America*, 97:3907–3908.

Gerzon, M. A. (1992a). The design of distance panpots. In *92nd Convention of the Audio Engineering Society, Preprint 3308*, Vienna.

Gerzon, M. A. (1992b). Panpot laws for multispeaker stereo. In *92nd Convention of the Audio Engineering Society, Preprint 3309*, Vienna.

Griesinger, D. (2002). Stereo and surround panning in practice. In *112th Convention of the Audio Engineering Society*, Munich.

Haas, H. (1972). The influence of a single echo on the audibility of speech. *Journal of the Audio Engineering Society*, 20(2):146–159.

Hafter, E. (n.d./2001). Simulated Open Field Environment (SOFE). Web published at `http://ear.berkeley.edu/auditory_lab/sofe.html`, University of California at Berkeley.

Haustein, B. G. and Schirmer, W. (1970). Messeinrichtung zur Untersuchung des Richtungslokalisationsvernögens [A measuring apparatus for the investigation of the faculty of directional localization]. *Hochfrequenztech. u. Electroakustik*, 79:96–101.

Horbach, U., Karamustafaoglu, A., Pellegrini, R. S., and Corteel, E. (2000). Implementation of an auralization scheme in a digital mixing console using perceptual parameters. In *108th Convention of the Audio Engineering Society, Preprint 5099*, Paris.

Hull, J. (1999). Surround sound past, present, and future: A history of multichannel audio from mag stripe to dolby digital. Technical report, Dolby Laboratories.

ITU-R (1994). Multichannel stereophonic sound system with and without accompanying picture. Recommendation BS.775-1, International Telecommunication Union Radiocommunication Assembly.

Jackson, C. (1953). Visual factors in auditory localization. *Quarterly Journal of Experimental Psychology*, 5:52–65.

Jot, J.-M. (1997). Efficient models for reverberation and distance rendering in computer music and virtual audio reality. In *Proceedings of the International Computer Music Conference*, Thessaloniki, Greece. International Computer Music Association.

Jot, J.-M. and Chaigne, A. (1991). Digital delay networks for designing artificial reverberators. In *90th Convention of the Audio Engineering Society, Preprint 3030*, Paris.

Jot, J.-M. and Warusfel, O. (1995). Spat~: A spatial processor for musicians and sound engineers. In *Proceedings of CIARM 95*, Ferrara, Italy.

Kendall, G. S. and Martens, W. L. (1984). Simulating the cues of spatial hearing in natural environments. In *Proceedings of the International Computer Music Conference*, pages 111–125, Paris.

Kidd, G., Mason, C. R., Rohtla, T. L., and Deliwala, P. S. (1998). Release from mask-

ing due to spatial separation of sources in the identification of nonspeech auditory patterns. *Journal of the Acoustical Society of America*, 104(1):422–431.

Klepko, J. (1997). 5-channel microphone array with binaural head for multichannel reproduction. In *103rd Convention of the Audio Engineering Society*, New York.

Krokstadt, A., Strøm, S., and Sørsdal, S. (1968). Calculating the acoustical room response by the use of a ray tracing technique. *Journal of Sound and Vibration*, 8(1):118–125.

Kurozumi, K. and Ohgushi, K. (1983). The relationship between the cross-correlation coefficient of two-channel acoustic signals and sound image quality. *Journal of the Acoustical Society of America*, 74(6):1726–1733.

Kutruff, H. (1991). *Room Acoustics*. Elsevier Applied Science. Elsevier Science Publishers, Essex, 3rd edition.

Lehnert, H. and Blauert, J. (1992a). Aspects of auralization in binaural room simulation. In *93rd Convention of the Audio Engineering Society, Preprint 3390*, San Francisco.

Lehnert, H. and Blauert, J. (1992b). Principles of binaural room simulation. *Applied Acoustics*, 36:259–291.

Lemley, B. (2001). Virtual you. *Discover Magazine*, 22(7).

Lund, T. (2000). Enhanced localization in 5.1 production. In *109th Convention of the Audio Engineering Society, Preprint 5243*, Los Angeles.

Martens, W. L. (2001). Uses and misuses of psychophysical methods in the evaluation of spatial sound reproduction. In *110th Convention of the Audio Engineering Society, Preprint 5403*, Amsterdam.

Martin, G. (2001). *A Hybrid Model for Simulating Diffused First Reflections in Two-dimensional Synthetic Acoustic Environments.* PhD thesis, McGill University, Montreal, Canada.

Martin, G., Corey, J., Woszczyk, W., and Quesnel, R. (2001). A computer system for investigating and building synthetic auditory spaces: Part II. In *Proceedings of the AES 19th International Conference on Surround Sound Techniques, Technology and Perception*, pages 75–83, Schloss Elmau, Germany. Audio Engineering Society.

Martin, G., Woszczyk, W., Corey, J., and Quesnel, R. (1999a). Controlling phantom image focus in a multichannel reproduction system. In *107th Convention of the Audio Engineering Society, Preprint 4996*, New York.

Martin, G., Woszczyk, W., Corey, J., and Quesnel, R. (1999b). Sound source localization in a five-channel surround sound reproduction system. In *107th Convention of the Audio Engineering Society, Preprint 4994*, New York.

Mason, R., Ford, N., Rumsey, F., and de Bruyn, B. (2000). Verbal and non-verbal elicitation techniques in the subjective assessment of spatial sound reproduction. In *109th Convention of the Audio Engineering Society, Preprint 5225*, Los Angeles.

Mershon, D. H., Ballenger, W. L., Little, A. D., and McMutry, P. L. (1989). Effects of

room reflectance and background noise on perceived auditory distance. *Perception*, 18:403–416.

Mershon, D. H. and Bowers, J. N. (1979). Absolute and relative cues for the auditory perception of egocentric distance. *Perception*, 8:311–322.

Mershon, D. H. and King, L. E. (1975). Intensity and reverberation as factors in the auditory perception of egocentric distance. *Perception & Psychophysics*, 18(6):409–415.

Michelsen, J. and Rubak, P. (1997). Parameters of distance perception in stereo loudspeaker scenario. In *102nd Convention of the Audio Engineering Society, Preprint 4472*, Munich.

Miyasaka, E. (1993). Methods of quality assessment of multichannel sound systems. In *Proceedings of the AES 12th International Conference on the Perception of Reproduced Sound*, Copenhagen. Audio Engineering Society.

Moore, B. C. J. (1997). *An Introduction to the Psychology of Hearing.* Academic Press, San Diego, Calif., 4th edition.

Moore, F. R. (1983). A general model for spatial processing of sounds. *Computer Music Journal*, 7(3):6–15.

Moorer, J. A. (1979). About this reverberation business. *Computer Music Journal*, 3(2):13–28.

Morimoto, M. (1997). The role of rear loudspeakers in spatial impression. In *103rd Convention of the Audio Engineering Society*, New York.

Murphy, D. (2001). Surround-sound reverberation using digital waveguide mesh modeling techniques. In *Proceedings of the AES 19th International Conference on Surround Sound Techniques, Technology and Perception*, Schloss Elmau, Germany. Audio Engineering Society.

Nakayama, T., Miura, T., Kosaka, O., Okamoto, M., and Shiga, T. (1971). Subjective assessment of multichannel reproduction. *Journal of the Audio Engineering Society*, 19(9):744–751.

Nielsen, S. H. (1991). *Distance Perception in Hearing.* PhD thesis, Aalborg University Press, Aalborg, Denmark.

Nielsen, S. H. (1992). Auditory distance perception in different rooms. In *92nd Convention of the Audio Engineering Society, Preprint 3307*, Vienna.

Olive, S., Schuck, P., Sally, S., and Bonneville, M. (1994). The effect of loudspeaker placement on the listener preference ratings. *Journal of the Audio Engineering Society*, 42:651–669.

Pellegrini, R. S. (2000). Perception-based room rendering for auditory scenes. In *109th Convention of the Audio Engineering Society, Preprint 5229*, Los Angeles.

Pellegrini, R. S. (2001a). Quality assessment of auditory virtual environments. In *Proceedings of the 2001 International Conference on Auditory Display*, Espoo, Finland.

Pellegrini, R. S. (2001b). *A Virtual Reference Listening Room as an Application of Auditory Virtual Environments.* PhD thesis, Ruhr-Universität Bochum, Bochum, Germany.

Preibisch-Effenberger, R. (1966). *Die Schallokalisationsfähigkeit des Menschen und ihre audiometrische Verwendung zur klinischen Diagnostik [The human faculty of sound localization and its audiometric application to clinical diagnostics].* PhD thesis, Technische Universität, Dresden.

Puckette, M. and Zicarelli, D. (2001). MAX-MSP software, version 4.0.7.

Pulkki, V. (2001). *Spatial Sound Generation and Perception by Amplitude Panning Techniques.* PhD thesis, Helsinki University of Technology, Helsinki, Finland.

Quesnel, R. (1996). Timbral ear-trainer: Adaptive, interactive training of listening skills for evaluation of timbre. In *100th Convention of the Audio Engineering Society, Preprint 4241*, Copenhagen, Denmark.

Quesnel, R. (2001). *A Computer-Assisted Method for Training and Researching Timbre Memory and Evaluation Skills.* PhD thesis, McGill University, Montreal, Canada.

Ratliffe, P. (1974). Properties of hearing related to quadraphonic reproduction. BBC Research Department Report RD 1974/38, BBC.

Rayleigh, L. (1907). The theory of sound. *Philosophical Magazine*, 13:214–232.

Rife, D. D. (2000). *MLSSA: Maximum-Length Sequence System Analyzer, Reference Manual Version 10WA.* DRA Laboratories.

Rumsey, F. (2001). *Spatial Audio.* Music Technology Series. Focal Press, Oxford, UK.

Saberi, K., Dostal, L., Sadralodabai, T., Bull, V., and Perrott, D. R. (1991). Free-field release from masking. *Journal of the Acoustical Society of America*, 90:1355–1370.

Savioja, L., Backman, J., Järvinen, A., and Takala, T. (1995). Waveguide mesh method for low-frequency simulation of room acoustics. In *Proceedings of the 15th International Congress on Acoustics (ICA'95)*, volume 2, pages 637–640, Trondheim, Norway.

Schroeder, M. R. (1962). Natural sounding reverberation. *Journal of the Audio Engineering Society*, 10(3):219–223.

Schroeder, M. R. (1996). The "Schroeder frequency" revisited. *Journal of the Acoustical Society of America*, 99(5):3240–3241.

Sheeline, C. (1982). *An Investigation of the Effects of Direct and Reverberant Signal Interactions on Auditory Distance Perception.* PhD thesis, Stanford University, Stanford, California.

Shively, R. E. and House, W. N. (1996). Perceived boundary effects in an automotive vehicle interior. In *100th Convention of the Audio Engineering Society, Preprint 4245*, Copenhagen.

Silzle, A. (2002). Selection and tuning of HRTF's. In *112th Convention of the Audio Engineering Society, Preprint 5595*, Munich.

Smith, J. O. (1992). Physical modeling using digital waveguides. *Computer Music Journal*, 16(4):74–87.

Smith, J. O. (2002). *Digital Waveguide Modeling of Musical Instruments.* Center for Computer Research in Music and Acoustics (CCRMA), Stanford University. Web published at `http://www-ccrma.stanford.edu/~jos/waveguide/`.

Stautner, J. and Puckette, M. (1982). Designing multi-channel reverberators. *Computer Music Journal*, 6(1):52–65.

Stone, H. and Sidel, J. L. (1993). *Sensory Evaluation Practices.* Academic Press, San Diego, CA, 2nd edition.

Storms, R. L. (1998). *Auditory-Visual Cross-Modal Perception Phenomena.* PhD thesis, Naval Postgraduate School, Monterey, California.

Suokuisma, P., Zacharov, N., and Bech, S. (1998). Multichannel level alignment, Part I: Signals and methods. In *105th Convention of the Audio Engineering Society, Preprint 4815*, San Francisco.

Theile, G. (1980). *Über die Lokalisation im überlagerten Schallfeld.* PhD thesis, Der Technischen Universität Berlin.

Theile, G. (1990). On the performance of two-channel and multi-channel stereophony. In *88th Convention of the Audio Engineering Society, Preprint 2887*, Montreux.

Theile, G. (1993). The new sound format "3/2-stereo". In *94th Convention of the Audio Engineering Society, Preprint 3550a*, Berlin.

Theile, G. (2000). Multichannel natural recording based on psychoacoustic principles. In *108th Convention of the Audio Engineering Society, Preprint 5156*, Paris.

Theile, G. and Plenge, G. (1977). Localization of lateral phantom sources. *Journal of the Audio Engineering Society*, 25(4).

Toole, F. E. (1991). Binaural record/reproduction systems and their use in psychoacoustic investigations. In *91st Convention of the Audio Engineering Society, Preprint 3179*, New York.

Vorländer, M. (1989). Simulation of the transient and steady-state sound propagation in rooms using a new combined ray-tracing/image source algorithm. *Journal of the Acoustical Society of America*, 86:172–178.

Wagener, B. (1971). Räumliche Verteilungen der Hörrichtungen in synthetischen Schallfeldern [Spatial distributions of hearing-directions in synthetic sound fields]. *Acustica*, 25:203–219.

Wenzel, E. M., Arruda, M., Kistler, D. J., and Wightman, F. L. (1993). Localization using nonindividualized head-related transfer functions. *Journal of the Acoustical Society of America*, 94:111–123.

Woszczyk, W. R. (1993). Quality assessment of multichannel sound recordings. In *Proceeding of the 12th International Conference of the Audio Engineering Society on the Perception of Reproduced Sound*, Copenhagen. Audio Engineering Society.

Zacharov, N. and Bech, S. (2000). Multichannel level alignment, Part IV: The correlation between physical measures and subjective level calibration. In *109th Convention of the Audio Engineering Society, Preprint 5241*, Los Angeles. Audio Engineering Society.

Zacharov, N., Bech, S., and Suokuisma, P. (1998). Multichannel level alignment, Part II: The influence of signals and loudspeaker placement. In *105th Convention of the Audio Engineering Society, Preprint 4816*, San Francisco. Audio Engineering Society.

Zahoric, P. A. (1998). *Experiments in Auditory Distance Perception.* PhD thesis, University of Wisconsin - Madison.

Zahorik, P. and Wightman, F. (2001). Loudness constancy with varying sound source distance. *Nature Neuroscience*, 4:78–83.

Zwicker, E. and Fastl, H. (1999). *Psychoacoustics: facts and models.* Springer series in information sciences, 22. Springer, Berlin ; New York, 2nd updated edition.