Center for Computer Research in Music and Acoustics November 1982

Department of Music
Report No. STAN-M-13

AN INVESTIGATION OF THE EFFECTS OF DIRECT AND REVERBERANT SIGNAL INTERACTIONS ON AUDITORY DISTANCE PERCEPTION

by

Christopher W. Sheeline

Research sponsored by

National Science Foundation
and
System Development Foundation

CCRMA
DEPARTMENT OF MUSIC
Stanford University
Stanford, California 94305

Center for Computer Research in Music and Acoustics November 1982

Department of Music
Report No. STAN-M-13

AN INVESTIGATION OF THE EFFECTS OF DIRECT AND REVERBERANT SIGNAL INTERACTIONS ON AUDITORY DISTANCE PERCEPTION

by

Christopher W. Sheeline

The development of a complete model of auditory distance perception has been impeded by our poor understanding of the contributions made by room reverberation to the formation of distance judgements. Recent developments in computer technology for sound processing have made it possible to investigate in a rigorous manner the perceptual features of reverberation that affect distance perception. The thesis describes 1) an experiment that introduces the problem of working within real rooms for studies of this kind, 2) a system for simulating reverberation using a multi-channel computer driven sound system, 3) an experiment demonstrating the perceptual simularity between the room simulations and recordings of real rooms, and 4) a set of experiments investigating the relationship between intensity cues and reverberation cues in distance perception.

The results of these studies are significant at several levels. It has been demonstrated that it is possible to create compelling, natural sounding simulations of real rooms and that these simulations can be used to study distance perception. The relationship between intensity and reverberation cues has been clarified to the extent that it is now possible to suggest that a hierarchy of cues is utilized in the formation of distance inferences. The intensity of the direct sound serves as the primary determinant of perceived distance. However, in the absence of reverberation, the loudness cue is "impoverished" and listeners have great difficulty hearing the loudness differences in a "spatial" mode. Conversely, under high reverberation conditions, a similar effect is encountered, such that the reverberation masks loudness cues and erects a kind of "wall" beyond which the source will not recede. Thus, reverberation provides the "spatiality" that allows listeners to move from the domain of loudness inferences into the domain of distance inferences. Given the above two boundaries, an optimum degree of reverberance exists and it is within these boundaries that distance judgements and inferences are most successful.

This thesis was submitted to the Department of Hearing and Speech and the Committee on Graduate Studies of Stanford University in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

This research was supported by the National Science Foundation under Contract NSF BNS 77-22305-A2 and BNS 79-24646 and System Development Foundation under Grant SDF #345. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of Stanford University, any agency of the U. S. Government, or of sponsoring foundations.

Acknowledgements

The Center for Computer Research in Music and Acoustics, where I have been working for the last six years, is an extraordinary place. I have never encountered so many creative, industrious, and generous individuals under one roof and my work here has been made possible only through the cooperative support of these people. The gratitude that I feel is literally boundless.

My original interest in this research is due largely to my exposure to the work of John Chowning, the director of this center. Without his encouragement and insight and friendship, I doubt this work would have been started. To Chris Chafe, John Gordon, Bill Schottstaedt, Patte Wood, Dave Jaffe, Andy Schloss, Julius Smith, Mike McNabb, Jon Berger, Paul Wieneke, Jeff Borish, Loren Rush, Jan Mattox, John Strawn, Marc LeBrun, Tovar, and Steve McAdams, I owe thanks of all sorts. Without these supportive and talented friends, I doubt I could have accomplished more than a small part of this work. Their individual contributions of software, hardware, insight, experience and friendship would require pages to list, so I suspect it is best to just say "thank you". To the Chafe family, Chris, June and Zoe, I owe a special thanks for providing me with regular and healthy doses of domestic sanity and good red wine. Chris has been a continual source of friendship, support, encouragement and ideas. John Grey was an invaluable friend and mentor during my first few years here and without his early support I would have been lost.

Earl Schubert is incomparable, as a friend, adviser, teacher, and all-around great fellow. As my principal adviser, he devoted a great deal of his interest, time, and energy to my work.

I am indebted to Roger Shepard for the friendly and generous counsel that helped guide my work in the perceptual modeling of reverberant spaces. In addition, he was originally responsible for developing the multidimensional scaling techniques upon which much of my work rests. Jim Adams accepted me, a stranger, at his door, and was perhaps the person most responsible for my decision to actually finish this madness and graduate. I owe him immense thanks.

Finally, I would like to thank Darcy Prime: for her boundless confidence in me, for her love, support, encouragement, good humor, and for regularly reminding me that I have to take a break once in a while.

Table of Contents

Section	Page
1. Auditory Distance Perception	1
1.1 Introduction	1
1.2 Significant Contributing Research	2
1.2.1 Intensity:Loudness	4
1.2.2 Reverberation	5
1.2.3 Frequency Spectrum:Timbre	8
1.2.4 Interaural Time and Intensity Differences	9
1.2.5 Head Position	10
1.2.6 Visual Cues	11
1.2.7 Familiarity	12
1.3 Obstacles to the collection of experimental data	13
1.3.1 Complexity of parametric control of reverberation	13
1.3.2 The auditory system as an associative dimension to spatial orientation	14
2. A Preliminary Study in Auditory Distance Perception	15
2.1 Introduction	15
2.2 Experimental Design and Methods	16
2.3 Results and Discussion	18
2.4 Conclusion	23
3. Reverberation and Reverberation Modeling	24
3.1 Introduction	24
3.2 Subjective Preferences for Concert Halls	24
3.3 Perceptual Features of Reverberant Spaces	26
3.4 Modeling and Simulation of Reverberant Environments	27
3.4.1 Introduction	27
3.4.2 Convolution methods	27
3.4.3 Geometric modeling	28
3.4.5 Algorithmic integration of physical and perceptual features	29
3.5 CCRMA Methods	30
3.5.1 Introduction	30
3.5.2 Tap/Comb/LowPass/AllPass Reverberator	31

Section	Page

3.6 Verification of algorithmic reverberation modeling . . . 34
3.6.1 Introduction . . . 34
3.6.2 Stereo recordings of musical sounds in reverberant spaces . . . 35
3.6.3 Perceptual modeling by method of A:B comparison . . . 36
3.6.4 Multidimensional scaling of real and modeled rooms . . . 37
Two dimensional solution . . . 38
Three dimensional solution . . . 39
General conclusions concerning this study . . . 41

4. Varied Reverberation Conditions: Subjective Control of Loudness . . . 42

4.1 Introduction . . . 42

4.2 Experimental Design and Methods . . . 42
4.2.1 Listening Station: Hardware and Software Configuration . . . 43
4.2.2 Subjects . . . 44
4.2.3 Reverberator Conditions . . . 44
4.2.4 Procedure . . . 45

4.3 Results . . . 46
4.3.1 Introduction . . . 46
4.3.2 Depth of Source Penetration . . . 48
4.3.3 Amplitude:Distance Functions . . . 52
Slopes . . . 56
Loudness . . . 56

4.4 Just-Noticeable Differences . . . 57

4.5 Summary and Predictions . . . 61

5. Varied Reverberation Conditions: Magnitude Estimation of Source Penetration . . . 63

5.1 Introduction . . . 63

5.2 Experimental Design and Methods . . . 63

5.3 Listening Station, Procedure, and Subjects . . . 64

5.4 Results . . . 65
5.4.1 Introduction . . . 65
5.4.2 Optimum Range of Reverberation . . . 67
5.4.3 Reverberation as an Absolute Cue . . . 67
5.4.4 Individual Differences . . . 68

6. Conclusions and Suggestions for Further Research . . . 69

6.1 Reverberation Modeling . . . 69

6.2 Distance Perception . . . 70

6.3 General Conclusions . . . 70
7. Appendix I: Systems Concepts Digital Synthesizer Programming Specifications . . . 72
8. Bibliography . . . 75

List of Tables and Illustrations

Figure | Page

2.1 Half-matrix for multidimensional scaling study of intensity/distance pairs 16
2.2 Loudspeaker arrangement for intensity/distance study 17
2.3 KYST multidimensional scaling solution (numeric) 19
2.4 Graphical representation of one-dimensional KYST solution 19
2.5 Graphical representation of one-dimensional KYST solution grouped by diagonals 20
2.6 Graphical representation of one-dimensional KYST solution grouped by loudspeaker 21
2.7 Graphical representation of one-dimensional KYST solution: distance differences contrasted with intensity differences 27
3.1 Reverberation model used for implementing room simulations 32
3.2 Playback/recording configuration used in recording listening spaces for simulation 35
3.3 Half-matrix of stimulus set used in multidimensional scaling study of room simulations 37
3.4 KYST multidimensional scaling solution in 2 dimensions 38
3.5 KYST multidimensional scaling solution in 3 dimensions 39
3.6(a,b,c) KYST multidimensional scaling solution in 3 dimensions plotted in three 2-dimensional pairs 40
4.1 Comparison of the effect of various reverberation conditions on the range of movement of the source:all subjects, all conditions 48
4.2(a,b,c) Comparison of the effect of various reverberation conditions, grouped by reverberance 49
4.3 Reverberance vs. Mean depth of source penetration: all subjects, all conditions, by reverberation ratio 50
4.4 Reverberance vs. Mean depth of source penetration: all subjects, all conditions, by reverberation time 51
4.5(a,b,c) Intensity as a function of apparent distance: all subjects, all conditions, groups 1,2,3 53
4.5(d,e,f) Intensity as a function of apparent distance: all subjects, all conditions, groups 4,5,6 54
4.5(g,h,i) Intensity as a function of apparent distance: all subjects, all conditions, groups 7,8,9 55
4.6a Just-noticeable differences at 3 source positions for reverberant condition 4 58

Figure | Page

4.6b Just-noticeable differences at 3 source positions for reverberant condition 1 59
4.6c Just-noticeable differences at 3 source positions for reverberant condition 9 60
5.1 Range of apparent distance vs. reverberation ratio (linear ordinate): all subjects and mean 65
5.2 Range of apparent distance vs. reverberation ratio (log ordinate): all subjects and mean 66

Table | Page

4.1 Reverberant conditions of source placement study 44

1. Auditory Distance Perception

1.1 Introduction

Auditory spatial processing has been the subject of vigorous inquiry during the last forty years. However, most research has been oriented towards understanding the mechanisms by which the azimuth (angle) of a sound source is perceived, with less attention given to the perception of either elevation or distance. There appear to be two crucial issues that are not resolved in the available literature on auditory distance perception.

> **1)** There is considerable confusion among researchers concerning the ability of listeners to judge distance under different types of conditions, from anechoic to highly reverberant. Most researchers are willing to *conjecture* as to how listener responses to an experiment in one environment might change in a second environment, but there is no established body of data suggesting predictable changes. The essential questions remain: *how crude or precise is auditory distance perception in humans and under what conditions does this ability vary?*
>
> **2)** The major part of the previous work in spatial perception has been performed either with headphones or in anechoic environments. Few studies consider natural environmental factors such as reverberation. This is due mainly to the relative difficulty experimenters have had controlling these cues in the past. For any comprehensive understanding of auditory processing of spatial cues, it is imperative that investigations be carried out under conditions that include reverberation.

It may be possible to consider the perception of azimuth and elevation as being relatively unaffected by the reflections present in natural acoustic environments. However, there is good reason to believe that the perception of distance, the topic of this dissertation, is partially a function of cues provided by reverberation. Recent developments in digital signal processing techniques make it possible now to simulate and control natural sounding reverberation. Thus, it is possibile to explore in a more rigorous fashion the relationship between loudness and reverberation in the perception of auditory distance.

This dissertation deals primarily with issues **1** and **2**, above. It focuses in particular on the complex interaction between intensity cues and reverberation cues in exploring the acuity of auditory distance discrimination. A total of five studies are described, four of which use aritificial reverberation created with digital signal processing techniques.

Chapter 1 is a survey of the experimental and theoretical literature relating to distance perception. Chapter 2 describes an experiment from the early stages of this research that highlights the question of the interaction between the direct and reverberant signal in distance perception. Chapter 3 discusses perceptually significant features of room reverberation, and describes modeling methods and an experiment verifying the quality of the reverberation models. Chapters 4 and 5 describe a set of experiments in distance perception using artificial reverberation. Chapter 6 includes general conclusions and directions for further research.

1.2 Significant Contributing Research

1.2.1 Introduction

A number of studies are discussed in the following 6 sections. Each group is focused primarily on one of the potential cues introduced below. In each subsection, an outline of the basic physical properties of the cue is presented, followed by a discussion of its relationship to auditory distance perception as it appears in this incomplete experimental literature. Although there is an enormous body of work relating to spatial listening under diotic and dichotic conditions, this work is only marginally useful to the current discussion and is covered thoroughly elsewhere [e.g. Mills-68; Carterette and Friedman-79].

It is suggested that distance perception of auditory signals, under normally reverberant or natural listening conditions, is based upon four types of cues and is strongly *context dependent.*

1) *Sound Pressure Level*: In non-reflective environments, the intensity of the direct signal is inversely proportional to the square of the distance between the source and the listener. In reflective environments, the applicability of this square law decay function is modified by the amount of reverberant energy reaching the listener.

2) *Reverberation*: In a given acoustic environment, the ratio of reverberant to direct sound energy reaching the listener increases with distance. [Please note: Some authors invert the numerator and denominator of this ratio, referring to the ratio of direct to reverberant energy. In this thesis, the *reverberation ratio* refers to the ratio of reverberant to direct signal energy. Consequently, a *high* reverberation ratio is one in which the reverberant energy is high, relative to the direct signal energy.] It is also suggested that the reverberation *time* affects distance perception. These two factors are influenced by the design of the space, e.g. size, shape, and construction materials. Of particular importance in understanding how these parameters affect distance

perception is the fact that within a given room, although the reverberation time is *room dependent* and fairly constant, the reverberation ratio is *distance dependent* and increases at the position of the listener as the distance between the source position and the listener increases. The reverberation time will vary from room to room and the reverberation ratio will vary partially as a function of that change. A third reverberation parameter that changes with increased distance between source and listener is the temporal relationships between the direct signal and the prominent early reflections (occuring generally within the first 50-80 milleseconds of the room response). As distance increases, the interval between the arrival of direct sound and its first reflections will also change, as the early reflections are primarily a function of the proximity of the listener and source to room boundaries.

3) *Frequency Spectrum*: The absorptive properties of reflecting surfaces and obstacles in the sound path significantly modify the frequency spectrum of the reverberant energy over time. In addition, there is a drop off of high frequency energy in the direct signal with increased distance, as a result of the transmission properties of the air and as a function of temperature and humidity. This attenuation tends to be negligible for sounds in the audio domain over short distances. It has also been suggested [Bekesy-38; Coleman-68] that frequency spectrum plays a dual role in distance perception. For very short distances, an increase in the *emphasis* of low-frequency energy, relative to the high-frequency energy of the sound, should indicate an approaching source. At greater distances, this emphasis is not applicable and molecular absorption of high frequency energy accounts primarily for the signifcant *relative* spectral attenuation. This hypothesis is further explained in section **1.2.4**.

4) *Interaural Time and Intensity Differences/Head Position*: There are slight interaural differences in time and intensity when sounds originate off the median sagittal plane. These result from differences in path lengths as well as the shadow cast by the head. However, when considering distance perception under reverberant listening conditions, it is apparent that for sources on the median sagittal plane, there are no interaural differences in the incidence of the direct signal. When the reflections happen to be asymmetrical, as in the case of a listener closer to one wall than another, the resulting binaural difference tends to affect the spatial orientation with respect to the room, rather than the apparent distance of the source.

The various cues suggested above all tend to involve some mutual interdependency. However, their relative salience and the interactions among them under natural listening conditions (i.e. "in context") are not well defined. For example, differences in sound pressure and frequency spectrum of the sound are processed by the auditory system in very different ways, but it is suggested that both are dependent upon the listener's familiarity with the sound for any strong or absolute sense of distance. Likewise, it is apparent that in the absence of reverberation cues, listeners will tend to judge distance on the basis of loudness differences, and these judgments will be compressed over a range that bears little likeness to the range defined by the actual source distances. The literature reflects both the complexity of the problem and the absence of a system or model that encompasses all of the variables mentioned above. As there is little comprehension of the interaction effects of these cues for distance perception, it is understandably difficult to hypothesize a working model for further testing.

As distance perception is highly *context dependent*, the effects of *familiarity* with the source signal and the effect of the *visual environment* must be considered along with the cues to distance perception listed above. The *familiarity* of the signal interacts with the listener's accuracy of distance estimation, given the cues of sound pressure level and frequency spectrum. It has been asserted by Coleman [1962] that 'Accurate distance judgments should ... not be possible upon initial exposure to unfamiliar sounds. For each new stimulus, with its characteristics of frequency spectrum and intensity, the cues must be recorrelated with external space on the basis of experience.' In intra-trial comparisons of distance *change*, this dictum might need to be modified. Familiarity with the listening *environment* may be equally important in the absence of visual cues.

The source position suggested by *visual* cues has also been shown to influence auditory localization and distance estimates [e.g. Maxfield-30; Maxfield-31; Gardner-69a,b,c; Mershon et al-80]. In particular, Gardner's experiments suggest that listeners will select the nearest *rational* location of an apparent source of sound.

The consideration of these cues suggests that they carry unequal weights in the formation of distance judgments or inferences. A historical treatment of research relating to the above cues is presented in the following section.

1.2.2 Intensity:Loudness

Intensity has been studied under a number of different conditions and is the cue for distance perception whose *theoretical* principles are best understood. The inverse square law describes the attenuation of sound energy over distance. But though loudness as a variable in *free field* conditions correlates positively with the physical intensity of the signal reaching the eardrum, it becomes a more complex variable under conditions where *reflected* sound is being summed with the direct signal. To our knowledge, no investigator has attempted the systematic measurements of sound pressure levels for equal distances under both reverberant and free field conditions. It is a curious feature of this particular cue that even though it has been an independent variable of secondary importance in a number of the studies to be discussed [Bekesy-60; Coleman-62; Gardner-69; Molino-73; Mershon and King-75], no study rigorously investigates the relationship between loudness or intensity and the perception of distance. The conclusions of all of these authors reflect the fact that some familiarity with the source signal is crucial if listeners are to be capable of making loudness related distance inferences.

On the psychophysical level, *loudness* is the term most widely employed for subjective evaluations of sound intensity. An appreciation of the complexity of loudness as a perceptual dimension is crucial to an understanding of how other cues might interact with loudness in the formation of apparent distance judgments. Although several scales for loudness have been suggested [Stevens-55; Zwicker-58], the basic *sone* scale is the one with which most readers are familiar. This scale simply uses the loudness of a 1000 hz tone, 40 dB above threshold, as reference. There are problems with the sone scale, however. Power function biases, suggested by a number of researchers [Warren-58; Warren-77; Garner-58; Gibson-66], cast some doubt on the usefulness of the phon scale. An attempt to modify this scale [Stevens-55] met with limited success. The most comprehensive theory of loudness has been developed for an evaluation of the loudness of complex steady-state signals [Zwicker and Scharf-65]. This model attempts to deal with these various factors which influence loudness: 1) the sensitivity of the ear at different frequencies, 2) the locations of stimulation at different frequencies along the cochlea, 3) the effects of critical bandwidths and 4) the effects of masking patterns with respect to the above. What emerges is a complex predictor of auditory behavior that describes loudness as a function of frequency, bandwidth, critical band, and masking. The sone scale does not adequately describe these results. Warren's study on loudness [Warren-77] purports to show a bias in the Stevens type power function and suggests that a 6 dB reduction in intensity (reduction to 25 percent of stimulus intensity) is the most appropriate correlate to half loudness for sine tones. As may be seen below, no investigators have attempted to correlate perceived distance with loudness according to any one loudness scale. Warren, in fact, poses a loudness/distance model that minimizes the importance of reverberation.

1.2.3 Reverberation

Reverberation has been the most elusive and least understood member of the family of cues under discussion. It has generally been considered uncontrollable under experimental conditions and is notable in most studies for its absence. There are several straightforward reasons for this neglect. Reverberation is probably a multidimensional cue for localization and distance perception. The aspects of reflected sound that are important in spatial perception have never been particularly clear, although this subject is receiving increased attention. [See references below]. In the experiments that have considered reflected sound, the environments have generally been rooms that were neither too "live" nor too "dead", other criteria being unexplicit and difficult to articulate. Only very recently, chiefly as the result of work by M.R. Schroeder and his colleagues on isolating and identifying acoustic and psychoacoustic features of room reverberation, has it been possible to consider the nature of reverberation as an experimental variable. [Schroeder and Logan-61; Schroeder-61; Schroeder-62; Schroeder and Kuttruff-62; Schroeder-70; Schroeder, Atal, Sessler and West-70; Schroeder-73; Barron-74; Schroeder-75; Gottlob-75; Schroeder-79]. Chapter 3 will treat this subject in greater detail.

Bekesy investigated many attributes of spatial hearing. His experimental results tend to suggest not one, but many potential cues to distance perception and a strong dependence on the specific environmental or contextual conditions. His work in distance perception covered several different potential cues; loudness, direct signal spectrum, the ratio of reverberant to direct signal, and time delays between the incidence of the direct signal and the first reflections at the listener. He maintains, as have most investigators, that "Loudness has an effect upon the perceived distance only in the absence of more determinate physical cues." [Bekesy-38].

Bekesy's study of the ratio of reverberant to direct signal is particularly significant in light of recent studies that will be reported in chapters 4 and 5. Although he maintains that the "image corresponds merely to a mixture of close and remote sound images, both of which have already been determined by other cues", he does stress that the reverberation ratio is important for distance perception and that its complex nature makes study more problematic than for other cues. One less conclusive aspect of Bekesy's work involved the importance of time delay between the incidence of the direct sound and the initial high energy reflections. He was unable to successfully relate this feature of room acoustics to distance perception. A pilot study by this author reinforces Bekesy's conclusions with respect to early reflections.

The importance of the ratio of reverberant to direct signal for simulating a distance cue was supported by Chowning's work on the simulation of moving sound sources [Chowning-71]. His approach involved the use of two reverberation cues, *global* and *local*. *Global* reverberation refers to the even distribution of reverberant energy in a four loudspeaker environment. *Local* reverberation refers to the distribution of reverberant energy between the speaker pair that defines the location of the direct signal and is proportional to (1-(1/distance)). This method served to focus the reverberation with increased distance, enhancing the distance percept.

In an excellent study on intensity and reverberation cues in distance perception, Mershon and King clarify the distinction between what they describe as a relative cue (intensity) and an absolute cue (reverberation) [Mershon and King-75]. They conducted two studies, the first in a small 'hand made' reverberant tunnel, the second in an anechoic chamber. In both situations the listeners were blindfolded and faced loudspeakers at distances of roughly 2.75 meters and 5.5 meters (in the median sagittal plane). When presented with high or low intensity white noise (5 sec. dur.), subjects were asked to estimate the apparent distance of the sound. Responses were collected for each subject's first and second presentations. These estimates showed that although absolute accuracy was low under both conditions, the magnitude estimates under reverberant conditions were within the approximate range of the real sources. For the free field conditions, the estimates were too small by a factor of nearly 10! In addition, for the reverberant conditions, initial judgments of high intensity noise preserved the physical distance relationship of the speakers. The most interesting aspect of this study seems to be its demonstration of the clear contribution made by reverberation in suggesting appropriate boundaries (or potential boundaries) for a source or set of sources. Although only first and second presentations of the sounds were used in tabulating the results (thus creating conditions of 'unfamiliarity' for both the stimulus and the environment), it is quite evident that signal intensity alone is an inadequate cue to perceived distance.

In a study more fully described in the next chapter, Sheeline attempted to elicit a cognitive 'mapping' into auditory space of a set of intensity/distance combinations for stimuli presented under reverberant conditions in the median sagittal plane. A set of 5 loudspeakers faced the listener, at distances of 3', 6', 12', 24' and 48'. Pairs of stimuli were presented and listeners were asked to rate their apparent distance from each other on a dimensionless scale, yielding 'dissimilarity' judgments for each pair of stimuli. The most intense sounds from each loudspeaker had been equalized for loudness and, under anechoic conditions, there would presumably have been no perceived difference between sounds. However, it was expected that in the experimental reverberant environment, the listener would organize the various ranges of sounds according to a more complex scheme. Sheeline's findings support the premise that the cues provided by intensity differences between members of a set of stimuli in a reverberant environment though important, are not sufficient cues for the formation of reliable distance inferences. The results of the multidimensional scaling analysis demonstrated that the perceived relationships between the loudspeakers for each intensity level were preserved according to their real physical relationship, and the only plausible explanation for this effect is the presence of reverberation.

1.2.4 Frequency Spectrum:Timbre

The frequency spectrum of the source signal is affected by several factors: filtering as the result of obstacles in the sound path, filtering of the reverberation as a result of absorbtive properties of the reflecting surfaces, and by slight molecular absorption of energy in the higher frequencies for large distances (this attenuation is negligible in the audio frequency range for short distances). The combination of these effects creates a likely cue to distance perception, but presumably only for a familiar sound. If listeners are not already familiar with the timbre of a sound, they will not be perceptually sensitive to a relative attenuation of high frequency energy. Interaural differences in spectrum and stimulus bandwidth, known to provide essential information for localization tasks involving azimuth discrimination, have been little studied for their contributions to distance perception. Butler and Planert attempted diotic and monaural investigations of the effect of bandwidth on distance perception under free field conditions. In the binaural test, listeners' interaural time and intensity differences were minimized. In the monaural condition one ear was occluded. Their results indicate a *decreased ability to localize sounds in the median sagittal plane as bandwidth decreases.* Their stimuli consisted of narrow band noise, centered around eight kHz with bandwidths of one to six kHz, and broad band noise, thus covering most of the audio spectrum. They comment 'The spectral cues furnished by a bandwidth of only 2.0 or 3.0 kHz are impoverished, and the spectrum is not sufficiently modified when the sound source originates from widely different positions' [Butler and Planert-76].

Coleman suggests, in accordance with the findings of Bekesy, that frequency spectrum plays a dual role in distance perception. Within the range of roughly a meter, it is apparent that an emphasis of low-frequency energy, relative to the high-frequency energy of a sound, should indicate the approach of a source. This is due primarily to effects of particle velocity in a *spherical* sound field. Also, within this range, high-frequency attenuation as a result of molecular absorption is below the auditory threshold (on the order of .05dB for energy at 8000 Hz). However, beyond the range of a few feet, the nature of the sound field tends towards the planar, rather than the spherical, and this low-frequency emphasis for nearer sources no longer applies. However, as the distance between source and listener increases, the effect of molecular absorption increases as well. At a distance of 100 feet, the attenuation for energy at 8000 Hz is on the order of 3.3dB, well above the perceptual threshold. This attenuation would most likely serve as a relative indicator of increased distance. Thus, Coleman suggests that 'frequency spectrum may play a dual role in auditory depth perception, with relatively greater high-frequency content signaling a *closer* sound source at distances greater than a few feet, but signifying a more distant sound source when the source is close to the observer' [Coleman-68]. The spectrum of the source signal is, of course, a highly relative cue and its relevance appears to depend strongly on the listener's familiarity with the environment and with the signal.

As in the discussion of intensity and loudness, a brief distinction should be made between the acoustical or spectral features of the source signal and its *timbre*. Grey has provided a thorough and systematic series of experiments elucidating the perceived characteristics of musical timbre [Grey-75; Grey-77; Grey-79; Grey-78a; Grey and Gordon-78; Grey and Moorer-77]. The three perceived dimensions that emerged as most important for the distinction of various timbres were 1) spectral energy distribution 2) low amplitude, high frequency energy in the the initial attack segment of the tone, and 3) synchronicity in the attacks and decays of the higher harmonics. The first of these, the spectral energy distribution, relates directly to the issues treated in this dissertation. Consequently, *timbre* is used herein as a synonym for *perceived distribution of spectral energy*, keeping in mind the more complete meaning embodied in Grey's work. The source signal's physical energy spectrum can be thought of as one physical correlate of the source *timbre*.

1.2.5 Interaural Time and Intensity Differences

Although the effects of interaural time and intensity differences are well documented in the studies of azimuth perception (especially centering and lateralization), the studies relating these cues to distance perception have been primarily theoretical. Early researchers [Hartley and Fry-21; Wightman and Firestone-30] were concerned with measuring interaural differences in phase and intensity for source signals on and off the median sagittal plane at different distances. Wightman and Firestone determined that phase and intensity differences between the ears change significantly with distance up to a range of about four meters from the center of the listeners head. Beyond that point, for a given azimuth, intensity differences change more rapidly than phase differences. In a survey of cues to free field perception of distance, Coleman presents 5 possible cues, two of which are well established, namely loudness and frequency spectrum. The other three, pinna effects along with binaural intensity and phase differences for non 0 degree azimuth, are presented as theoretical possiblities rather than cues with demonstrated substantive effects for free field listening [Coleman-63]. In separate studies, Hirsch and Molino present theoretical models for the ability of human listeners to make distance judgments on the basis of interaural time and intensity differences [Hirsch-68; Molino-73]. Hirsch condsiders the differential path lengths between the source and the two ears as a function of the inverse square law, assuming a free field and an azimuth off the median sagittal plane. He collected no experimental data. Molino describes a modification to the Hirsch equation , including the report of a pilot experiment in which two listeners attempted to identify the source loudspeaker in free field tests using pure tones of 1kHz and 8 kHz. Five loudspeakers were located 3' - 48' from the listener at 0 degree azimuth. His findings, in agreement with those of Coleman and Gardner [Coleman-62; Gardner-68] indicate lack of ability to estimate distances on the basis of interaural intensity differences alone, but he fails to mention the specific results, the number of trials, or whether there was any improvement over time.

Recently Searle presented a model for auditory localization suggesting that interaural pinna disparities exist for many listeners. If this is the case, it may be an indication that some listeners use these disparities as supplementary aids in the processing of timing information and possibly for the formation of distance related inferences [Searle-73; Searle-75; Searle-76].

These binaural cues probably interact with nearly all of the other types of cues under consideration, especially reverberation. Given free field conditions, it seems unlikely that listeners will be able to make accurate judgments on the basis of time and intensity differences alone, particularly for sounds in the median sagittal plane. Molino's pilot study supports this conjecture. Under reverberant listening conditions, it is probable that interaural time and intensity differences become absorbed into a much more complex multidimensional processing scheme and their effect becomes difficult to isolate and negligible in that context.

1.2.6 Head Position

In one sense, the consideration of head position (azimuth) as a cue to distance perception is merely a restatement of the discussion of interaural time and intensity differences. However, as stated previously, very few experimental studies have been performed which relate distance perception specifically to interaural differences. Given our interest in reverberation as an integral part of any model for space perception, the precise measurement of interaural differences becomes highly impractical. It is possible to make theoretical predictions concerning the behavior of the direct signal, which reaches us before any reflections. However, as the Coleman, Hirsch, and Molino studies have shown, there is not yet convincing evidence for the consideration of interaural differences as having a primary influence on distance perception,

For purposes of experiment and discussion, however, azimuth *can* be considered as an independent variable. Under these conditions, the degree of rotation of the source off-axis is measurable, as is the effect of such a rotation.

Holt and Thurlow present some evidence that listeners under nonreverberant conditions are able to make distance judgments for sources off the median sagittal plane. In their study [Holt and Thurlow-69], head orientation was very important. Listeners facing 90 degrees from the source were much more successful in judging source distance than listeners facing directly towards the sound source. The theoretical explanation for this result is not entirely clear. The interaction of interaural intensity differences with a familiar source signal may be a strong distance cue under free-field conditions. Interaural time and intensity differences in the direct signal should not exist for sources in the median sagittal plane. It does seem possible, but not likely, that as the distance between the source and the listener increases off the median sagittal plane, central processing mechanisms are able to relate increased interaural disparities to distance differences.

1.2.7 Visual Cues

The visual cue or the 'expectation' of the listener is a factor that lends an uncertain weight to all aspects of distance perception and localization. In our personal experiences, we are rarely consciously aware of the perceptual and cognitive integration that takes place in our processing of the visual and auditory environments. If we found ourselves in an enormous stone cathedral, however, we would be surprised and disturbed if we heard only the direct signal from a trombone on the other side of the building. Likewise, if we found ourselves in a small acoustically treated room with a trombonist eight feet from us, it would seem unnatural to hear his music as though it came from the other end of Notre Dame in Paris. These examples may be extended to situations that we can more readily imagine. In our everyday experiences, we often hear sounds whose exact place of origin is difficult to specify. This lack of precision does not prevent us from making satisfactory inferences as to distance and angular location. Voices in the corridor passing an open door are usually perceived as just that. We are normally able to correlate information from the visual and 'known' environment with auditory information. If the cues from the two senses were overly disparate, either low intersensory correlation would result in confusion for the listener or one modality might dominate the other. One example of this problem occurs with motion picture sound effects. The inappropriate recording situation with respect to the location of the source in the film is often apparent to us, and is always apparent to the professional ear. It does, however, tend to disappear on continued exposure.

Gardner studied distance estimations for 0-degree and apparent 0-degree oriented speech sounds in free field listening conditions, for five loudspeakers in full view of the listener [Gardner-69]. The loudspeakers were located in full view, 10' - 30' from the listener in 5' increments. Intensity was varied for the 'active' loudspeakers (2 of the 5) and listeners were asked to identify the source loudspeaker. Gardner found that apparent sources for high level signals tend to be as near as the observer can rationally locate them. He refers in his conclusions to the well known 'proximity image effect' in which the selection of the nearest rational source seems to occur. In this experiment, it is clearly the visual field that provides the nearest rational source. It is not expected that these results would be duplicated for reverberant conditions, given the varying ratio of reverberant to direct signal energy over distance. In fact, the Sheeline study seems to indicate that even moderate amounts of reverberation can disturb this proximity image effect to some extent, as the 'rational' source suggested by visual cues is counterbalanced by reflective cues in the room that suggest other possible locations.

1.2.8 Familiarity

Next to reverberation, perhaps the most critical factor in the auditory domain for the perception of distance is the listener's familiarity with the source signal, or the class of source signal. Coleman demonstrated that listeners are unable to estimate distances between 9' and 27' for unfamiliar sounds presented in the median sagittal plane, when listening under free field conditions. The error rate for initial trials was approximately 65 percent. As trials progressed, however, most listeners were able to improve their distance estimates to an error rate of roughly 15 percent. He concluded that "Accurate distance judgments should, therefore, not be possible upon initial exposure to unfamiliar sounds. For each new stimulus, with its characteristics of frequency spectrum and intensity, the cues must be recorrelated with external space on the basis of experience" [Coleman-62]. This conclusion is supported by Molino [Molino-73].

The listener's familiarity with both the source signals and the acoustic environment is clearly a key component of any model for auditory distance perception. However, no studies have been performed that attempt to evaluate systematically the effects of familiarity with source material (e.g. musical instrument tones, speech, sounds from the environment) on distance perception. It is probable that the results of such a study would serve to support our belief that the most critical stimulus components for distance perception are changes in source loudness and changes in spectral content of the source, *relative to some known or familiar standards*, in the presence of reverberation cues.

1.3 Obstacles to the collection of experimental data

1.3.1 Complexity of parametric control of reverberation

The major deterrent to an analysis of the importance of reverberation cues in distance perception has been the inability of experimenters to control reverberation in any precise way. As stated in the last section, the extent of control has generally been confined to the inclusion or exclusion of reverberant cues. Chapter 3 presents a thorough examination of subjective features of room reverberation and of the specific model used in this research. As an introduction to the problem, however, a brief overview is included below.

It is possible to imagine a continuum of experimental methods for creating different, controlled reverberation cues. However, the examples at either end of this continuum call for enormous resources, and a cost-benefit analysis would be extremely difficult to perform. For instance, given sufficient money and time, it might be possible to find (or build) a building with a number of very different rooms close together. In this situation, if identical calibrated sound systems were set up in each room, an experiment could be run in which listeners made distance judgments within each different space. Ideally, context effects resulting from moving subjects among rooms could be minimized. This is not a particularly practical solution to the problem of controlling reverberation cues, however, as even the most skilled architectural acousticians have difficulty controlling reverberation with any precision.

The opposite end of this same continuum provides a method that requires equally extensive resources. This is the process of digital convolution of the selected source sound(s) and impulse responses recorded at different distances in different halls. The difficulty in this method is two-fold. The first, and most difficult problem is finding an impulse that is repeatable and generates adequate energy. This task has stymied acousticians for decades, although at least one method under development shows great promise. The second part of the problem is that the convolution of an adequate set of room and distance combinations with the desired source signal requires prodigious computer resources. However, both of these methods contain what might be considered a significant failing, and that is the lack of provision for the experimenter to *control* the environment.

At points along this methodological continuum running from the 'field' to the laboratory are different notions of what constitutes an appropriate simulation. The method selected for the studies described in chapters 3,4 and 5 takes advantage of the extensive experimental data regarding *perceptual* features of room reverberation. Given that listeners are insensitive to small differences in the size, shape or absorbency of real spaces, it is possible to create *generic* models of differing reverberant spaces. These simulations resemble no particular room or hall but do mimic a *class* of halls of similar volume and reverberance. This relatively recent development has provided a partial solution to the problem described in this section. Chapter 3 explains this system in some detail.

1.3.2 The auditory system as an associative dimension to spatial orientation

It is a fact that the auditory system provides less precise cues to spatial orientation than the visual system, for most people, most of the time. The 'proximity image effect' described by Gardner can be extended to a large number of more general cases in which the visual field provides a primary or more heavily weighted cue to source distance. For this reason, the mechanisms of distance perception have not received as much attention as they are due. In fact, the auditory system is quite capable of discriminating among different source distances and, under some conditions, is capable of fairly accurate distance judgments. That the system is an accessory to vision is incidental. For people who suffer considerable visual impairment, or for most people in dark environments or at night, when visual acuity is reduced, the auditory system plays a more significant role in spatial orientation.

The extent of concern with visual references in this dissertation has been minimized. Only trained listeners were used, and it was apparent that the visual field played little or no part in their judgments. Research that includes the study of untrained listener responses might well require a greater attention to the visual field on the part of the experimenter.

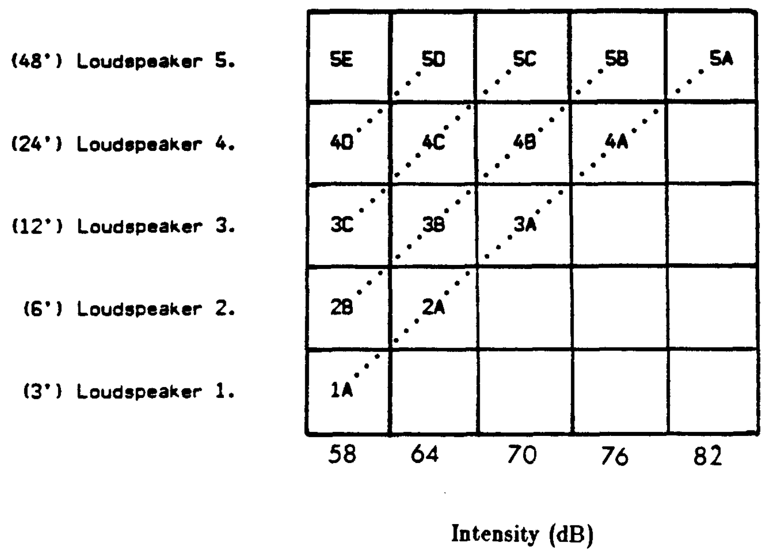
2. A Preliminary Study in Auditory Distance Perception

2.1 Introduction

A study was performed in an attempt to resolve the question, for this researcher, of the contribution made by reverberation in distance judgments. The inverse square law appears to provide an uncertain predictor of apparent distance for familiar and unfamiliar sounds under different types of environmental conditions. As suggested by Gardner's experiment [1968], *in a perfectly anechoic environment*, if the sound pressure of a signal is equalized at the listener's position with the same signal originating further away along the same line, most listeners will localize the sounds at, or near, the same source. A similar result could be inferred from both Coleman's and Molino's studies [Coleman-62; Molino-73]. A number of reasons were suggested in the last chapter to explain this imprecision in human ability to accurately resolve both 1) the distance between the source and the listener and 2) differences in distance between sources. The question explored in the following study was: *To what extent are the distance relationships suggested by the inverse square law modified by 'normal' room conditions, for a set of sources receding from the listener along the median sagittal plane?* Rather than reproduce the methods used in the anechoic studies mentioned above, a direct *cognitive* or *perceptual* representation was sought, using the technique of multidimensional scaling.

Given an anechoic environment and a set of sources covering a moderate (room size) range of distances, equalized in the manner suggested above, a listener should localize all of the equally loud sources at approximately the same position. For example, in Figure 2.1, if the amplitude-distance pairs along diagonal **A** are all equalized at the same sound pressure level, they would theoretically be localized at the same point on a line receding from the listener. A similar result would be expected for the sources corresponding to diagonals **B, C** and **D**. However, in an environment that provided additional cues, the loudness equalization would no longer be the only determining factor. How might listeners order the apparent physical proximities of the fifteen stimuli?

Figure 2.1

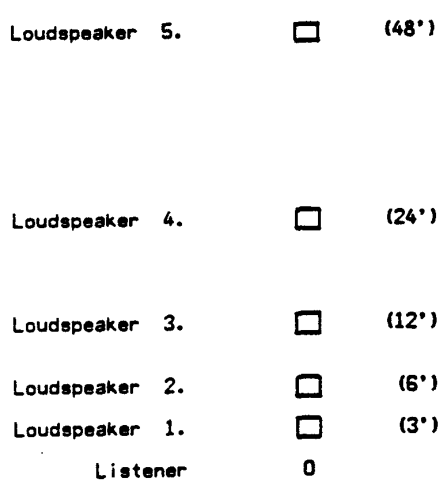


Under an anechoic listening condition, it is theorized that all source/distance combinations on a given diagonal would be localized at the same position, since the intensities on each diagonal would be equal at the position of the listener and few, if any, other cues would be available. The experiment was conducted under reverberant conditions in order to examine the deviation of the subjective data from the hypothetical anechoic data.

2.2 Experimental Design and Methods

Five speakers were arranged in the large conference room of the Stanford Artificial Intelligence Laboratory, an environment with a reverberation time of approximately 1 second, at distances of 3', 6', 12', 24' and 48' (Figure 2.2). Thus, the distance of each loudspeaker was twice the distance from the listener and the previous speaker. The speakers receded along the median plane of the listener. Each was raised slightly above the one in front of it so that no obvious filtering of the direct signal took place. A preamplifier circuit was built that allowed a signal at any one of five preset amplitudes to be routed to any of the five speakers.

Figure 2.2 Loudspeaker Arrangement



The loudspeakers spanned a range of 3' to 48' in a semi-reverberant conference room approximately 30'x75'x9'. The experimenter sat some distance away, manually selecting each intensity/distance combination.

Equalization was made with a sound level meter at a maximum level of 82 dB SPL, in the manner described in the introduction. The loudest sound emitted by the nearest loudspeaker, measured at 82 dB, corresponded to **1A** in Figure 2.1. This then became the point of reference for all the other stimuli. **2A,3A,4A** and **5A** were then equalized at the same sound pressure level, 82 dB, at the position of the listener. It is for this reason that only the half-matrix is complete in Figure 2.1, as each successive amplitude level was 6 dB greater than at the previous level. The total set of stimuli consisted of: 5 sounds of 82 dB SPL, one from each loudspeaker, 4 sounds at approximately 76 dB, each from one of the four distant speakers, 3 sounds at 70 dB, 2 at 64 dB and the last at roughly 58 dB. All sound pressure levels except for those at 82 dB are approximate, of course, as it was with reference to the inverse square law that the matrix was established, and by explicitly equalizing the intensity levels for the **B, C**, and **D** series, this assumption would have been untested. There was, unfortunately, a significant amount of band-limited ambient room noise from air circulation, including both low-frequency rumble and high-frequency hiss. This noise was averaged at approximately 62 dB but was focused primarily in the ranges below 128 hz and above 6000 hz. As the source signal was a 500 Hz square wave, masking proved not to be a problem. The loudspeakers were a closely matched set of 5 Phillips 544 motional feedback units, with built in amplifier circuitry.

The listener was asked, at the beginning of the experiment, to imagine a continuum of sound sources, rather than five discrete sources. We had assumed that the disorientation resulting from blindfolding the listener would be counterproductive to the generally relaxed nature of the experiment, although hindsight does not necessarily support that assumption. The total set of 15 stimuli formed a matrix of 225 possible pairs. Of these pairs, 105 were used (N(N-1)/2). This excluded the comparison of any matrix entry with itself, and represented each pair in only one of the two possible orders of presentation. Listeners were then asked to rate the *apparent distance between members of a pair*, on a dimensionless scale of 1 to 30. This rating condition was determined to be functionally equivalent to the more conventional request in multidimensional scaling studies for *similarity or dissimilarity ratings*. A total of 14 complete sets of data were collected, with several listeners running twice. Each subject was given as much practice as was felt necessary to establish the 'continuum'. This number was as low as 15 trials and as high as 50 (for several fairly novice listeners).

2.3 Results and Discussion

Two programs were used for the multidimensional scaling analysis, INDSCAL and KYST, both developed primarily at Bell Laboratories. INDSCAL is an algorithm designed to take advantage of individual differences in sets of response matrices. KYST is a program that does conventional non-metric multidimensional scaling and allows for a large number of different data conditions and constraints.

The INDSCAL analysis , which was performed first, utilized all 14 data sets and yielded solutions in 1, 2, and 3 dimensions, none of which lent themselves to a satisfactory interpretation. This appeared to be due primarily to individual differences between our experienced and inexperienced listeners. In general, for the subjects who were unused to listening critically, the task was difficult and confusing, and the results reflected that confusion.

The KYST analysis used only the 8 most consistent data sets. These represented all of the 'sophisticated' listeners, and the analysis yielded an illuminating 1-dimensional solution. The actual 1-dimensional KYST solution is shown in Figures 2.3 and 2.4.

Figure 2.3

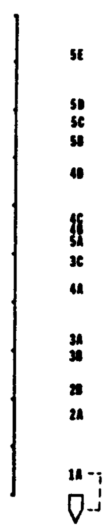
KYST Multidimensional Scaling Solution

Configuration of 15 points in 1 dimension has stress 0.278 formula 1.

Intensity/Distance Pair	Final Configuration
1A	1.854
2A	1.377
3A	0.718
4A	0.385
5A	-0.140
2B	1.125
3B	0.839
4B	-0.223
5B	-0.953
3C	0.085
4C	-0.272
5C	-1.158
4D	-0.646
5D	-1.234
5E	-1.629

The one-dimensional solution from KYST lends itself to reasonable interpretation, despite the fact that the stress level of .278 is quite high by Kruskal's standards [Kruskal-64a,b].

Figure 2.4

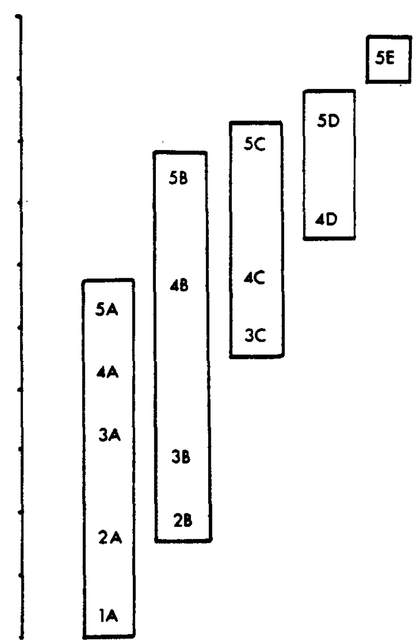


Graphical representation of the one-dimensional KYST solution. This is the perceptual mapping for the complete stimulus set, showing the perceived relationships among the different matrix entries. The listener position in the experiment was 3 ft. from the nearest loudspeaker, and that relationship is illustrated with reference to the nearest logical intensity/distance combination (1A).

In figures 2.5, 2.6, and 2.7, various sets of stimuli have been moved off the axis, in order to illustrate different relationships among them. The stress level was .278, not good at all by Kruskal's measure, but this is offset by the suggestion of an interesting and interpretable solution.

Two interesting observations may be made regarding the data as displayed in figure 2.5. 1) It can be seen that the relationships between speakers on each diagonal (i.e. at each loudness level), are preserved. Thus the *apparent* relationships between the 5 loudest equalized tones, the **A** diagonal, were preserved in the monotonic order corresponding to their real physical relationships, as were those on the **B, C** and **D** diagonals. On this level there was little confusion. As the listener heard signals of equal loudnesses originating at speakers of varying distances, the order of the speakers was preserved. 2) The perceived locations of sources on the **A** diagonal were spaced almost perfectly equally in the perceptual mapping. The perceived distances between 5A and 4A, 4A and 3A, 3A and 2A, and 2A and 1A were nearly identical, while the real distances were 24, 12, 6 and 3 feet, respectively. The smallest and largest differ by a factor of 8. Since the distance doubled while the SPL remained the same in each of these pairs, it appears that a very strong interaction occurred between the loudness cue and other factor(s), presumably reverberation.

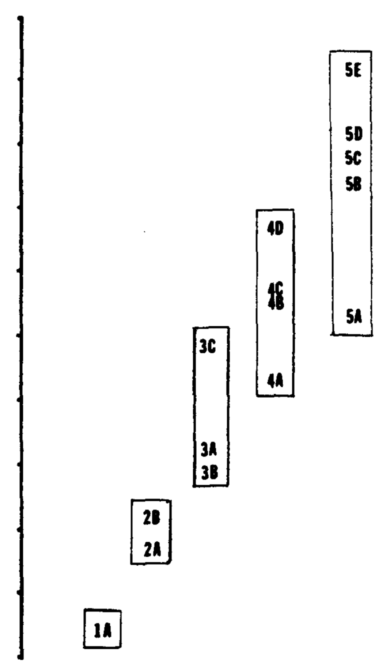
Figure 2.5



Graphical representation of the one-dimensional KYST solution with the intensity/distance combinations for each *diagonal* boxed separately and moved off axis for better viewing. In the original matrix of figure 2.1, each diagonal represents a set of intensity/distance combinations of equal intensity with respect to the listener. Each numerical increment on a given diagonal represents a doubling of the distance between the actual source and the listener. (e.g. 3A is 12 ft. from 2A which is 6 ft from 1A)

If the data is represented somewhat differently, as in Figure 2.6, it becomes clear that as the overall *intensity* increased at a specific loudspeaker, the sounds were perceived as originating closer to the listener. Thus, at position 5, as loudness levels E, D, C, B, and A were heard by the listener, the origin of the sound seemed to approach the listener, much as we would expect. The only exception to this is a confusion between 3A and 3B, which seemed to occupy the same source position. This could have been due to a degeneracy in the scaling solution, but it may well indicate that source resolution is more difficult in the middle ground. If that were the case, it would suggest that the interaction of loudness and reverberaion cues is most salient and *useful* when the energy ratio is very large or very small.

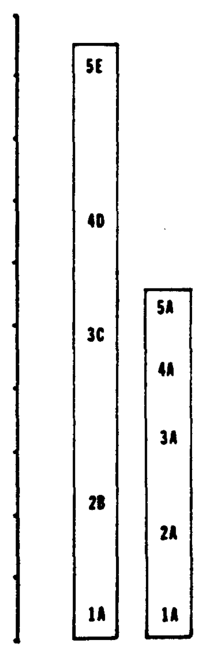
Figure 2.6



Graphical representation of the one-dimensional KYST solution with the intensity/distance combinations for each *loudspeaker* boxed separately and moved off axis for better viewing. Thus, for a given loudspeaker, each alphabetic increment represents a 6 dB reduction in intensity at that position.

Another compelling observation, viewed in Figure 2.7, is that the single tone that appeared at all 5 positions (1A, 2B, 3C, 4D, 5E) spanned nearly twice the apparent distance of that defined by the equal intensity tones on the **A** diagonal, and did so, again, in near-equal increments. The most plausible explanation for this occurence is that the interaction of loudness and reverberation can create a set of confounding relationships that can be mutually reinforcing or misleading.

Figure 2.7



Graphical representation of the one-dimensional KYST solution. The range of apparent distances spanned by a tone of constant intensity (58 dB) *at the five loudspeaker positions* (1A, 2B, 3C, 4D, 5E) is contrasted with the range of apparent distances spanned by the tone of constant intensity *at the listener position* (the **A** diagonal).

2.4 Conclusion

Two conclusions are suggested by the results of this study.

In a reverberant environment, intensity differences resulting from distance differences seem to provide most of the information necessary to resolve adequately a set of sources in the median sagittal plane [as in 1A, 2B, 3C, 4D, and 5E]. However, for sounds of equal intensity, originating at different distances, there is only sufficient information to resolve the *order* of the sources, not their real relationships [as in the A diagonal]. Thus, the reverberation appears to have interacted strongly with intensity in determining the relative acuity of distance perception.

Intensity differences without distance differences, in a reverberant environment, are a confounding factor in distance perception. This is clear when one considers that for each speaker position, the range of tones at that position created the illusion of a sound receding physically from that source. The several significant confusions in localization, notably 3C and 5A, seem to indicate the possible dimensions of this illusion. These two appeared quite close to each other, despite the fact that they were 36 feet apart. This suggests that in a reverberant environment, intensity differences without distance differences can generate a sense of changing distance.

The general conclusion that can be drawn supports the conclusion of Mershon and King, described in the last chapter [Mershon and King-75]. *In a reverberant environment, intensity differences alone are only a relative cue to distance perception, and it is clearly the contribution of the reverberant environment that determines, in any absolute sense, our ability to make distance discriminations. The reverberant cue does not necessarily increase acuity.*

3. Reverberation and Reverberation Modeling

3.1 Introduction

We have seen that the subject of room reverberation is highly complex and very poorly defined, from the perspective of someone attempting to model reverberation at a perceptual level for the purpose of psychophsyical research. This chapter is subdivided into several sections. The first two discuss attempts by several researchers to define the salient perceptual features of room reverberation and to understand subjective preferences for different concert halls. The third section discusses the appropriateness and relevance to this dissertation of several different methods of reverberation modeling. The fourth section describes the primary method used at CCRMA, and the fifth describes the results of a study in reverberation modeling.

3.2 Subjective Preferences for Concert Halls

Since the earliest quantitative work in concert hall acoustics [Sabine-22], acousticians and architects have been attempting to correlate the many and varied physical measurements that can be made in concert halls with the subjective assessments of listeners. This thesis is not the appropriate forum for a discussion of the many directions this research has taken, but it does seem reasonable to outline the cumulative findings of several researchers whose work is both comprehensive and mutually reinforcing.

Gottlob analyzed subjective responses to a number of different European concert halls, using multi-dimensional scaling and factor analysis, in an attempt to identify the salient perceptual dimensions of subjective preference. According to his data, a small number of measures are important in defining the subjective response of a listener[Gottlob-75].

a) Overall reverberation time is highly correlated with subjective preferences. A reverberation time of roughly 2 seconds appeared optimum, but this figure interacts with those below and serves only as a point of reference.

b) The ratio of early arriving energy to the total energy emerged as an important factor with optimum values in the range of .3-.35. This 'early' period covers approximately 50 milliseconds, although it varies from hall to hall.

c) Listeners preferred a high ratio of uncorrelated early lateral reflections to the total early reflections. This ratio is described by Gottlob as the interaural coherence.

d) Both (b) and (c) are closely related to the width of the concert hall, with narrower, rectangular halls consistently preferred to wider or fan-shaped halls.

e) The lower the interaural correlation, the greater the preference for the hall. Interaural correlation refers to the identity of the signals impinging on the two ears. Thus, a low interaural correlation implies that the early reflections be uncorrelated binaurally and the later reverberant energy be highly diffuse. Also notable is the fact that the early reflections for increasing the 'clarity' of the room reverberation arrive primarily from non-lateral directions.

e) The arrival time of the first prominent reflection is also significant. In most cases, the longer the delay, the lower the preference.

Barron's work [Barron-74] was an investigation of the effects of early reflections on subjective acoustical quality in concert halls. He supports most of Gottlob's findings, examining the effects of directionality, delay and frequency of the early reflections on subjective impression. He defines a term, "spatial impression" that relates closely to what we often call "presence", the closeness or intimacy or sense of being "surrounded" by the sound. He drew several major conclusions from his research.

a) The sense of spatial impression is primarily a function of the early lateral reflections and operates primarily for frequencies below about 1500 Hz.

b) Spatial impression varies little for delays of less than 8-10 milliseconds in the incidence of the first lateral reflection. Between 10 and about 60 milliseconds, there is little or no increase in spatial impression. Above 60 milliseconds, to a limit of 100 milliseconds, there is a slight increase in sense of spatial impresssion.

c) As sound intensity increases, the spatial impression increases monotonically.

The net result of the work by these two researchers is a much greater understanding of the general perceptual features of room reverberation, particularly for halls designed for musical listening. For the work described in this thesis, the above conclusions represent a solid foundation for a more generalized form of reverberation model or simulation.

3.3 Perceptual Features of Reverberant Spaces

One might suspect that in discussing the subjective preferences for different concert halls we would probably deal with the salient perceptual features of those halls, and in general this is true. However, one issue has received relatively little treatment in the research discussed in the last section, and that is the notion of the *liveness, brightness* or *brilliance* of a hall. This can be defined most simply as a stressing of the middle and high frequencies combined with a relatively slow decay in those regions. This term contrasts with *warmth* (adequate bass emphasis) in the traditional jargon of musicians and acousticians, but is an added factor in the consideration of subjective preferences. There are, in addition, a number of other subjective terms that are associated with architectural acoustics. A partial list of those has been treated in general terms by Beranek [Beranek-62] and includes, besides those above, *intimacy, loudness* and *clarity.* All of these can be redefined with relative success in terms of measures already defined by Gottlob and Barron. The measures which reappear in most cases are:

1) the initial delay gap of the first reflection (shorter delays tend to produce clearer and more intimate sound images)

2) reverberation time (affecting loudness, liveness, brilliance and warmth)

3) ratio of lateral to non-lateral early energy (affecting the clarity and intimacy of the hall)

From the perspective of one who is attempting to simulate these effects, this is useful information. We must remember, however, that measures of reverberation time are generally made across a wide range of frequencies, and this accounts for the variety of subjective dimensions that are at least partially defined by reverberation time.

It is appropriate to consider, at this point, the several different methods that might be used for the purpose of actually simulating a concert hall or large room using a computer-based sound system. As we have seen, the number of dimensions to consider is large and the interaction of those dimensions appears to be quite complex.

3.4 Modeling and Simulation of Reverberant Environments

3.4.1 Introduction

As was pointed out in the last chapter, there are several advantages to the simulation of room reverberation *in the laboratory.* Foremost are the resulting economy of time and effort and the elimination of variance due to the relocation of subjects and equipment from hall to hall. A second attractive feature of simulations is that they permit parametric control of relevant perceptual cues.

The disadvantage of a modeled system is that, when all is said and done, we still find ourselves wondering what the precise correspondence is between perception of the simulated, *realistic* conditions and the physical, *real* conditions.

There are three basic types of models that we can consider here, each of which has several subcategories or methods associated with it. The advantages and disadvantages of each, for the purposes of the distance research to be described, are outlined below. The models can be divided into two basic categories, 1) *black box* models in which the researcher manipulates recorded signals in a manner that requires *no knowledge* of the reverberant space, and 2) *knowledge based* models, in which the researcher applies acoustical and/or perceptual knowledge to the development of the simulation. The first of the methods described below, convolution, can be represented as a 'black box'. The second and third methods, geometric and algorithmic modeling, are knowledge-based systems.

3.4.2 Convolution methods

Convolution methods have marked advantages and disadvantages. They allow the researcher to perform very sophisticated analyses and experiments without ever really knowing anything about the reverberant space in question. The basic process involves the recording of an impulse response of a real hall and convolving this digitized impulse response with a digitized direct sound, recorded in a very dry or anechoic environment. If the impulse response is recorded using a mannekin or artificial head, it is possible to monitor the result of the convolution with headphones and obtain a *perfect physical and perceptual identity* with the response of the room to that particular direct sound, from the position of the mannekin. If the only objective of the study is to obtain this *perceptual identity,* then this process operates as a perfect black box. Absolutely no knowledge about the hall is required, and as long as the recordings and convolution are properly executed, the desired result is achieved. Given sufficient resources, it is certainly the preferred method, as it is possible to extract everything that one might want to know about a concert hall from the impulse response, or from a set of responses to the same hall from different positions and under different conditions. This is the primary method used by Schroeder, Gottlob and Barron in their studies of concert hall acoustics.

Unfortunately, the disadvantages to this method are substantial, and fall into several categories. The most severe is simply the resources that are required, financial, human, and computer. The stereophonic convolution of a two second impulse response with a monophonic direct signal of 1 second duration, at a sampling rate of 40 kHz, requires 6.4 billion multiplications. It is possible to use an FFT implementation of the convolution algorithm, in which the impulse response length is equal to the logarithm of the original length, but this still represents significant investment in computer time. Then, considering that comparisons are to be made, as in the work of the three researchers mentioned above, and considering that for musical listening the musical excerpts are more likely to be 30 seconds long than 1 second long, we face a formidable computing task. A second, and equally serious, problem is that of producing a true impulse of sufficient brevity (less than 25 microseconds) and sufficient energy (at least 60 dB SNR relative to the ambient noise level of the hall) to make the recording useful. Assuming that a suitable spark-producing mechanism (or some other device) can be contrived, we must then consider the fact that the low-frequency energy emitted by this device may be 30-50 dB below the energy in the higher frequencies and the consequent excitation of the space is unevenly biased towards the higher frequencies. [There is at least one method, originally suggested by Schroeder, that is under development at CCRMA and proves to be an exciting alternative. The technique envisioned for measuring the impulse response of the reverberant space is to excite the room with maximal length sequences (noise) and then crosscorrelate the noise with the measured response. It can be shown mathematically that this operation will provide the impulse response of the system. Because the excitation is sustained, unlike the spark-gap, it is possible to inject more energy into the room, providing an improved signal to noise ratio and increased low-frequency energy.[Borish-83]]

3.4.3 Geometric modeling

Geometric modeling methods provide a number of useful approaches, although they tend to be highly complex. In general, they involve programming the acoustical characteristics of the hall and then tracing the incidence and decay patterns of a large number of sound *rays* as they reflect off room boundaries. The more accurate the desired measure, the more complex and computationally intensive the process. Simple, two-dimensional *ray tracing* programs, with convex geometric shapes and a single absorption coefficient, can require minutes of computer time. As the designer increases the complexity of the model, the computation quickly becomes prohibitively expensive and time consuming. There is a general utility, however, in the application of this type of model to simple tasks, particularly for the purpose of determining patterns of early reflections. In addition, statistical averages can be made using a limited number of rays, and rough conclusions can be drawn concerning reverberation time, absorption and frequency response.

A second method, related to ray tracing but computationally simpler, is the use of an *image model.* This model takes advantage of the fact that the *theoretical manner of reflection of a sound wave at a boundary is specular,* and eliminates the problem of computing the actual path of each sound wave for a large number of reflections.

This method is most commonly used in the field of architectural acoustics. Real, physical room boundaries tend to be diffuse, however, and most features of room acoustics are very difficult to represent with a manageable number of coefficients and parameters. From the perspective of a researcher whose primary interest is simulating the *perceptual* characteristics of a general class of hall, these methods are not yet sufficiently precise. They are in the development stage and show great promise but are currently inadequate.

3.4.4 Algorithmic integration of physical and perceptual features

A third class of models takes into account a number of well understood physical and perceptual features of room reverberation and integrates that knowledge into a model that can be applied through relatively simple algorithmic methods to a digital signal. The particular model used for the research described in this thesis is presented in the next section under *CCRMA methods,* but the general process is outlined here.

If the researcher's goal is to create the *impresssion* of natural room reverberation, the entire realm of subjective features described in the first two sections of this chapter can be considered. If artificial means can be used to create those subjective impressions, coincidence with the physical properties of the reverberant space becomes an attractive byproduct of the simulation process, not a necessary condition of the model. For instance, if the primary impression that a listener has during the first 30 milliseconds of room response is created by 4 or 5 prominent early reflections, there is no purpose in including an additional half dozen that might occur within the same time window in a real room, but have insufficient energy to alter subjective impresssion. Similarly, if the later reverberant energy is perceptually equivalent to Gaussian noise decaying exponentially as a function of room volume, then it behooves the researcher to use this method and avoid the problems of more complex equations or geometric models for generating the later part of that decay, *even if the physical representation is closer to that of a real room in the more complex case.*

This type of data reduction proves extremely useful in processes that are as computationally intensive as room simulation. The obvious disadvantage is that it is nearly impossible to create a perceptual *identity.* In particular, this is true if the monitoring is done through headphones (where small binaural differences in phase and intensity can enhance or disturb the focused sound image of the direct signal). However, for the general purpose of creating and controlling a convincing and compelling impression of natural room reverberation, this method provides a fortunate alternative.

3.5 CCRMA Methods

3.5.1 Introduction

The development of useful digital reverberators had its origins at Bell Laboratories and the University of Gottingen under the direction of M.R. Schroeder. At Stanford University, under the direction of John Chowning and James A. Moorer, a number of further steps have been taken, utilizing principles developed by Schroeder. The work of Schroeder and his colleagues has involved the study of room acoustics and computer modeling of rooms and concert halls for the purpose of design and prediction of acoustical properties [Schroeder and Logan-61; Schroeder-61; Schroeder-62; Schroeder and Kuttruff-62; Schroeder-70; Schroeder, Atal, Sessler and West-70; Schroeder-73; Schroeder-75; Gottlob-75; Schroeder-79]. The idea of 'colorless' artificial reverberation originated with Schroeder and from this notion was developed the Schroeder 'all-pass' unit reverberator, a recirculating delay unit that simulates the reflections of the dense reverberation, without introducing modifications in the frequency domain.

The work at Stanford University (CCRMA) [Chowning-71; Chowning, Grey, and Moorer-77; Moorer-79] has been concentrated on the development of 'natural sounding' digital reverberation. The general thrust of this work has involved the use of recirculating and non-recirculating delay units, *unit reverberators*, in series and parallel configurations. Ideally, the salient perceptual features of room behavior are modeled at a fraction of the cost required by the modeling methods described earlier in this chapter. A general specification calls for straightforward control over the reverberation decay time, the overall density of the decay, the low-pass filtering due to absorptive materials on reflecting surfaces, the lateral and non-lateral dispersion of the reflections over time, and parameters that control the suggestion of room dimensions, shape and resonances.

Chowning developed a system for controlling the apparent movement of sources of sound in an artificially reverberant space. His work incorporated the use of loudness, selective weighting of reverberation for each channel, and doppler shift to enhance the distance cue with moving sound sources. This work represents, in many ways, the first major step in the development of the tools currently in use [Chowning-71].

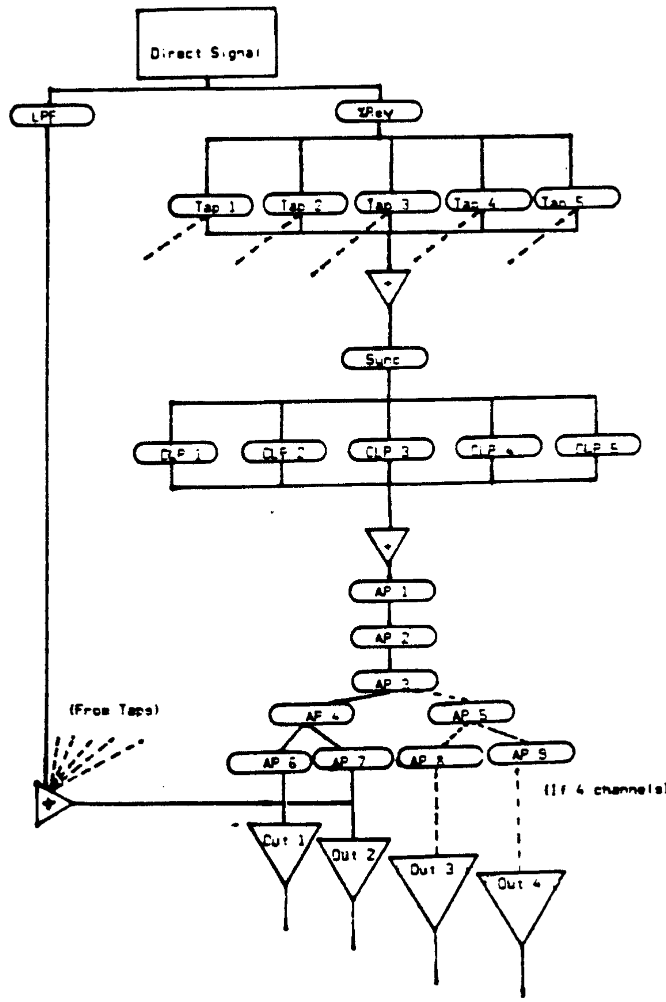
In work conducted both at CCRMA and at IRCAM (Institute for Research and the Co-ordination of Acoustics and Music), Moorer extended some of Schroeder's studies of reverberation in real rooms [Moorer-

79]. His work focused on phantom source studies such as the image model, on the development of a unit reverberator which had a controllable low-pass characteristic, and on complex configurations of recirculating and non-recirculating delays for good digital reverberation. The Moorer phantom source modeling gives, in many cases, a slightly more accurate figure for reverberation time in a real room than does the Schroeder model. The unit reverberator which emerged as the 'best value' was a comb filter, with a one-pole low-pass filter replacing the gain term in the comb. This has a low-pass characteristic dependent upon the delay length in the comb. As the delay length (corresponding to physical distance) increases, the high frequency attenuation increases and an increased low-pass filtering over distance is effected. When a number of these are combined in parallel, preceded by non-recirculating delays that simulate early reflections, and followed by a series of all-pass reverberators to increase the density and diffusion, a complex reverberator is created which roughly approximates the low-pass filtering effect of real rooms and provides extremely convincing reverberation. In Moorer's study, a nineteen tap section using the first 19 reflections taken from an impulse response of a two dimensional computer model of Boston Symphony Hall proved extremely successful and lifelike.

3.5.2 Tap/Comb/LowPass/AllPass Reverberator

It is now possible at CCRMA, using our Systems Concepts digital signal processor [described in Appendix 1], to control interactively most of the features listed above. In addition, source signals of any type: pure tones, complex synthetic tones, music, noise and speech may all be processed. The reverberation model used in this research is a slightly modified version of the Moorer system described above. Interactive control over a number of different parameters is implemented, and efficient non-interactive control is possible for the remainder. Figure 3.1 illustrates schematically the functional elements of this unit. It is implemented as a software 'patching' of hardwired modules on the digital signal processor and is described below. The actual system used to configure and drive the processor is far too complex to describe here [see Loy-77].

Figure 3.1



Where:

'LPF' is a weak, distance dependent, low pass filtering of the direct signal. A pair of cascaded 1 zero filters are used here.

'%Rev' is the percentage of the direct signal sent to the reverberator.

'Tap n' is an initial non recirculating delay designed to simulate the prominent early reflections found in real rooms.

'Sync' is a non recirculating delay that synchronizes the output of the last sample of the longest tap delay with the output of the first sample from the CLPs below.

'CLP n' is a comb reverberator with a 1-pole low-pass filter in the feedback loop. The delay lengths are chosen to reflect roughly the volume of the intended space.

'AP n' is an all pass reverberator that circulates the reflected, filtered and attenuated signal before output, increasing the density of the reverberation and decorrelating, in the time domain, the signal sent to the individual channels.

'Out n' is a channel being written to the digital to analog converters in the digital signal processor.

Examining figure 3.1 from the top down, it should be clear that two functions are performed in parallel, the processing of the *direct signal* and the processing of the *reverberation.*

Two modifications are made to the direct signal. *LPF* is a weak, distance dependent *low-pass filtering of the direct sound.* A pair of of cascaded 1-zero filters are used here. In the experiments described in the next section, this parameter was not varied, since all the comparison of reverberation quality used recordings made at equal distances. However, in general, this parameter is used to effect a very slight filtering that corresponds to the effect of molecular absorption in the air over very large distances. The filtered signal is then passed to an area in the processor that we will call *sum memory* in preparation for passage to the digital to analog converters (DACs) and the sound system. A set of dotted lines marked 'From Taps' indicates that while in sum memory, the direct signal is added to the output of the tap delays discussed below.

In a parallel path, a variable *percentage of the direct signal* amplitude, *%Rev,* is passed to the input to the reverberator. This percentage determines the ratio of reverberant to direct signal energy, the *reverberance* of the simulation, and interacts with a user specified gain term to control reverberation time.

This portion of the direct sound is then passed to a parallel set of *Tap* delay units. These are *non-recirculating delays* designed to simulate the prominent *early reflections* found in real rooms. They are of different lengths, generally falling in the range of 8-80 milliseconds. Under most circumstances, the range and separation of the delays depends roughly upon the desired size of the hall, with shorter delays suggesting closer room boundaries. The separation is primarily a function of trial and error, unless one is working from measured parameters of real rooms. Moorer suggests that one method of ordering the taps is in a 1:1.5 relationship between adjacent delays, but one quickly discovers that 'tuning' to avoid sharp echos and discontinuities must be performed as a matter of course. There appears to be no abstract method for determining their success ahead of time. The output of these delays is sent directly to the DAC sum memory where they are distributed to 2 or 4 channels with the direct signal. In addition, their output is summed and passed on through the reverberator.

The *Sync* unit is a non-recirculating delay that synchronizes the output of the last sample of the longest tap delay with the output of the first sample from the APs below. This is necessary for conditions in which a choice of a long last tap delay and a short first CLP create the anomaly of a dense and filtered reverberation reaching the listener before the last prominent early reflection.

The heart of the reverberator is the *CLP*, a *comb filter* with a 1-pole *low-pass filter* in the feedback loop. The delay lengths of the comb are chosen to reflect roughly the volume and reverberance of the intended space, but their main function is to simulate the low-pass characteristic of real rooms, as reflections are attenuated and filtered at the boundaries. Filter coefficients can be specified independently, or a single coefficient can be selected that will affect each delay separately. The output of these units is summed and passed on.

The next link, the *AP* units, are *all-pass* reverberators, in series and in parallel. Their function is to take the summed output of the CLP units and increase its density and diffusion. Simultaneously they *decorrelate* the signal from copies of itself, in the time domain, before passing them on to the digital to analog converters. This block of units represent the 'tail' of the reverberation. The delay lengths tend to be quite short, on the order of a millisecond, but they decay slowly, so the signal is constantly being reflected and attenuated in the manner of diffuse reflections in real rooms.

Out simply refers to the DAC sum memory, as described above, that passes the outputs of the direct signal, the taps and the all-pass units to the digital to analog converters. Two or four channels are normally selected.

3.6 Verification of algorithmic reverberation modeling

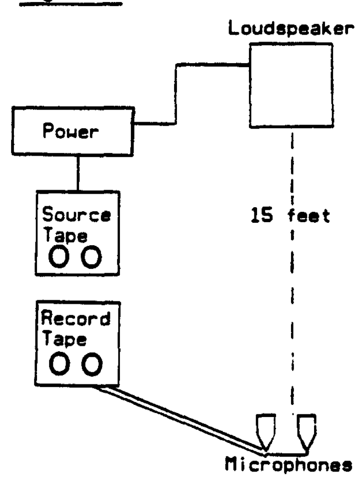
3.6.1 Introduction

An experiment was performed to verify that the algorithmic modeling described in the last section does, in fact, generate natural sounding artificial reverberation. The validity of the distance studies to be described rests on the assertion that the synthetic reverberation is perceived to be sufficiently similar to real room reverberation. The hypothesis was that if listeners do not distinguish among a mixed set of real and modeled halls on the basis of differences in "naturalness", but on the basis of analyzable reverberation parameters, then we can assume that the assertion is supported.

3.6.2 Stereo recordings of musical sounds in reverberant spaces

A number of trumpet sounds, covering the range of the instrument, were recorded in an essentially anechoic room. After suitable listening, we determined that a B flat arpeggio of two octaves, ascending and descending, would be least tiresome of all the recorded sounds for extended listening. This taped arpeggio was then used as the source signal for stereo recordings made in five different reverberant spaces in the Stanford community. In several halls, recordings were made at two distances. The playback/recording configuration used in those halls is illustrated in Figure 3.2. The monophonic source tape was played through a single loudspeaker at a sound pressure level of 84 dB measured 3 feet in front of the loudspeaker. A pair of microphones 6 inches apart, in a conventional stereo recording configuration, was placed exactly 15 feet from the loudspeaker and the arpeggio was recorded on a second tape recorder from that position. The halls selected included a small and minimally reverberant conference room, two medium sized lecture halls with low to moderate amounts of reverberation, a small, fairly reverberant chapel, and a large, extremely reverberant church. These stereo recordings were then digitized and edited on the computer. In addition, the unreverberated monophonic arpeggio was digitized, to be used as a source for generating the models.

Figure 3.2



The playback/recording configuration used for re-recording the dry trumpet tones in the listening spaces. In several halls, recordings were also made a distance of 30 ft.

3.6.3 Perceptual modeling by method of A:B comparison

The stereo recordings were then used in A-B comparison to generate an individual model for each real hall. It should be stressed that perfect perceptual identity was neither sought nor attained. The goal was to create a set of models that 'blended' perceptually with the real recordings in such a way that subjects differentiated among the whole set according to the same criteria that were used to differentiate among the recordings. This involved altering control parameters for most of the stages described in the reverberator of section 3.5.b, but principally the reverberation time, filter coefficients and tap delay lengths. Tape noise and some air conditioning noise that was preserved through the digitization process was also modeled. This solution was preferred to high and low pass filtering, as it was unclear how the low frequency rumble of the air conditioning might interact acoustically with the signals. It was also not desirable to eliminate this or the tape noise from the original recordings for fear of inadvertently removing relevant perceptual data.

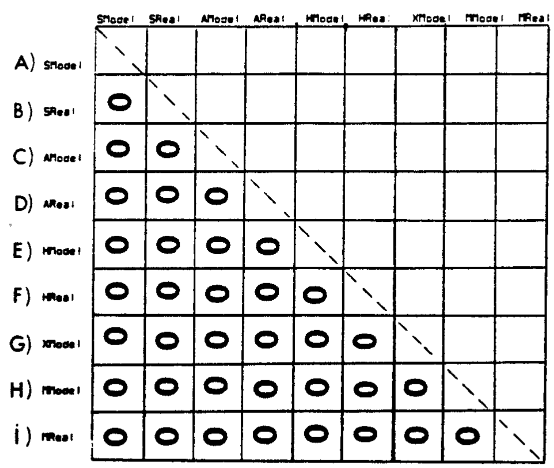
The range of reverberation times represented in the original recordings was approximately .25 seconds to 2.75 seconds. The frequency spectrum of the direct signal had audible energy to approximately 7.5 kHz. The default tap delays were suitable for three of the rooms and, after slight scaling, were used for the smallest and largest of the rooms, as well. The distribution in time of the six tap delays evolved over several months from an original set described by Moorer. For a sampling rate of 25641 samples/sec/channel the tap delay lengths range between 509 and 2039 samples (20 msec and 80 msec), corresponding to signal paths of lengths 22' and 90'. They were distributed between the two stereo channels to achieve bilateral asymmetry without creating the impression that the listener was much closer to one reflecting surface than another.

The actual modeling was done by a modified successive approximation, separately for each recorded example. The digitized recording was played, followed by the arpeggio processed through a basic version of the model room. If necessary a parametric adjustment was made and the process was repeated until the natural quality and perceptual similarity between the original and the model were satisfactory. In some cases, if it appeared that the recorded byproducts such as tape noise or air circulation noise were unusual, the models were 'backed off' from that point. In all other respects, each model was designed to belong to the same 'class' of rooms as its real counterpart.

3.6.4 Multidimensional scaling of real and modeled rooms

An experiment was performed using four out of the five recorded rooms and five perceptual models derived from those rooms, giving a total of nine stimuli. A half matrix of this set, as illustrated in Figure 3.3, was then presented in the form of paired comparisons. The listener was asked to evaluate the apparent similarity of every possible pair, on a dimensionless scale of 1 to 20. A group of six experienced listeners performed this study, several of whom had been subjects in the experiment described in Chapter 2. It was emphasized that they were to attend only to the source and the room, if possible. This was an attempt to steer them away from artifacts of the recording and digitizing process, such as the tape hiss and irrelevant room noise that was described earlier. In addition the subjects were questioned afterwards regarding the conscious basis upon which they distinguished differences among the various stimuli. A large number of trials was performed, with six repeated measures for each pair per subject. The total duration of the trials was 1 hour, including a 10 minute break half-way through.

Figure 3.3



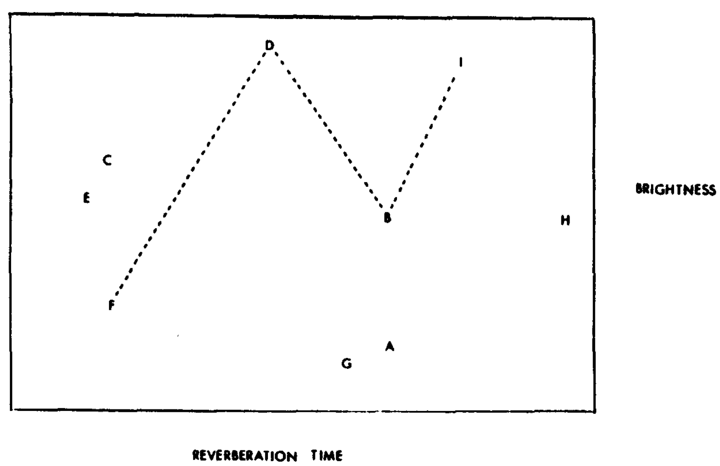
Half matrix of the stimulus set used in the multidimensional scaling study. The letter to the left of the vertical entries identifies the hall as it is represented in the multidimensional scaling solutions. Each pair was compared only once in each trial, and no hall was ever compared with itself. A total of 5 trials plus 1 practice trial were run on each subject. Each trial used a different random order for the paired comparisons. The real halls, which were all located in the Stanford community, and their corresponding simulations were as follows: *SReal, Smodel*: St. Ann's Chapel, Palo Alto; *AReal, AModel*: Annenberg Auditorium, Stanford; *HReal, HModel* Small conference room in the Department of Hearing and Speech Sciences, Stanford; *MReal, MModel*: Memorial Church, Stanford; *XModel*: a kind of "shot in the dark" that was modeled after no real hall, but sounded nice.

The results were tabulated and a multidimensional scaling analysis was performed, providing interpretable results in both two and three dimensions. Individual differences among subjects were also noted for the two dimensional solutions. Once again, KYST was the program used for the multidimensional scaling analysis.

Two dimensional solution

The group solution in two dimensions, Figure 3.4, can be well interpreted in terms of 1) reverberation time, and 2) degree of brightness or spectral modification of the reverberated signal. The reverberation times range from a perceived value of roughly .25 seconds to a duration of roughly 3 seconds. The bounds at both ends of this continuum were defined by models rather than recordings. In the case of spectral modification, the range of values was more narrowly defined and the brightest hall was represented by a real room while the dullest was represented by a model. Nevertheless, there was a distinct differentiation between those halls that tended to reflect the higher spectral components and those that did not. The KYST stress measure for this solution was .2.

Figure 3.4



KYST multidimensional scaling solution in 2 dimensions. The letters refer to the the individual halls as identified in Figure 3.3. Those connected by dots are the real halls.

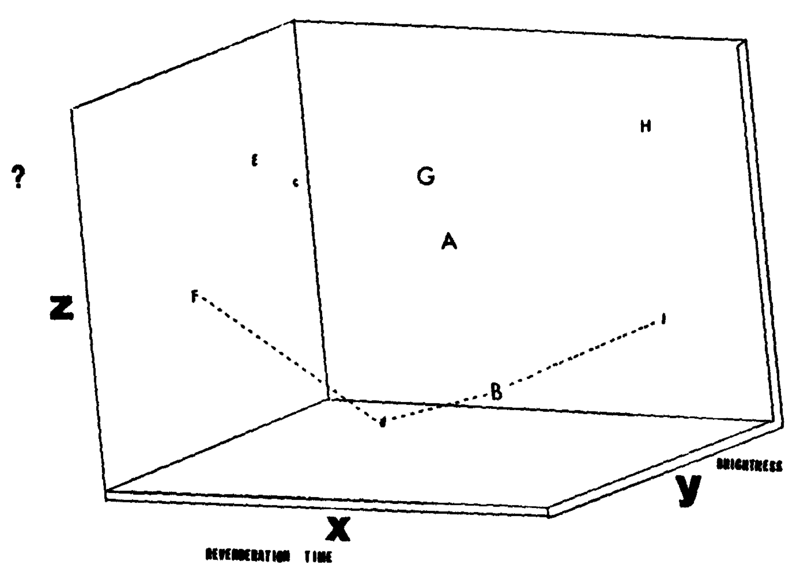
Most importantly, it can be seen that the real and modeled halls were separated primarily on the basis of these two dimensions. No parameter of 'naturalness' emerged in this phase of the analysis, supporting the contention of the adequacy of the models.

Three dimensional solution

In the three dimensional solution, illustrated in Figure 3.5, the two dimensions from the previous solution were relatively stable, and the third dimension represented a clear division between the real halls and the models. Our impression, however, as a result of discussions with listeners after the experiment, was that they tended to be responding more to differences among secondary characteristics such as noise from air circulation versus our simulated air conditioner noise and real versus simulated tape hiss. This is not conclusive, of course, but when asked after the experiment to listen to selected pairs and identify each member of the pair as a real or modeled room, only one of the six subjects who performed the complete 216 comparisons was *consistently* able to distinguish among them. The KYST stress measure for this solution was .148.

Figure 3.5 uses a size factor to simulate the visual distance cue for the viewer, the larger letters suggesting greater proximity to the viewer along the Y axis. Figure 3.6 illustrates the three dimensional solution as it is mapped out in two dimensional pairs, to clarify the perceptual relationships of that solution.

Figure 3.5



KYST multidimensional scaling solution in 3 dimensions. The Z axis has been labeled with a question mark, as it is uncertain whether or not this dimension reflects some aspect of the "natural" quality of the hall. The real halls are again connected by dots in this figure. The letters are of different sizes in order to provide a crude visual distance cue, with the smaller letters receding from the viewer.

Figure 3.6 (a)

Figure 3.6 (b)

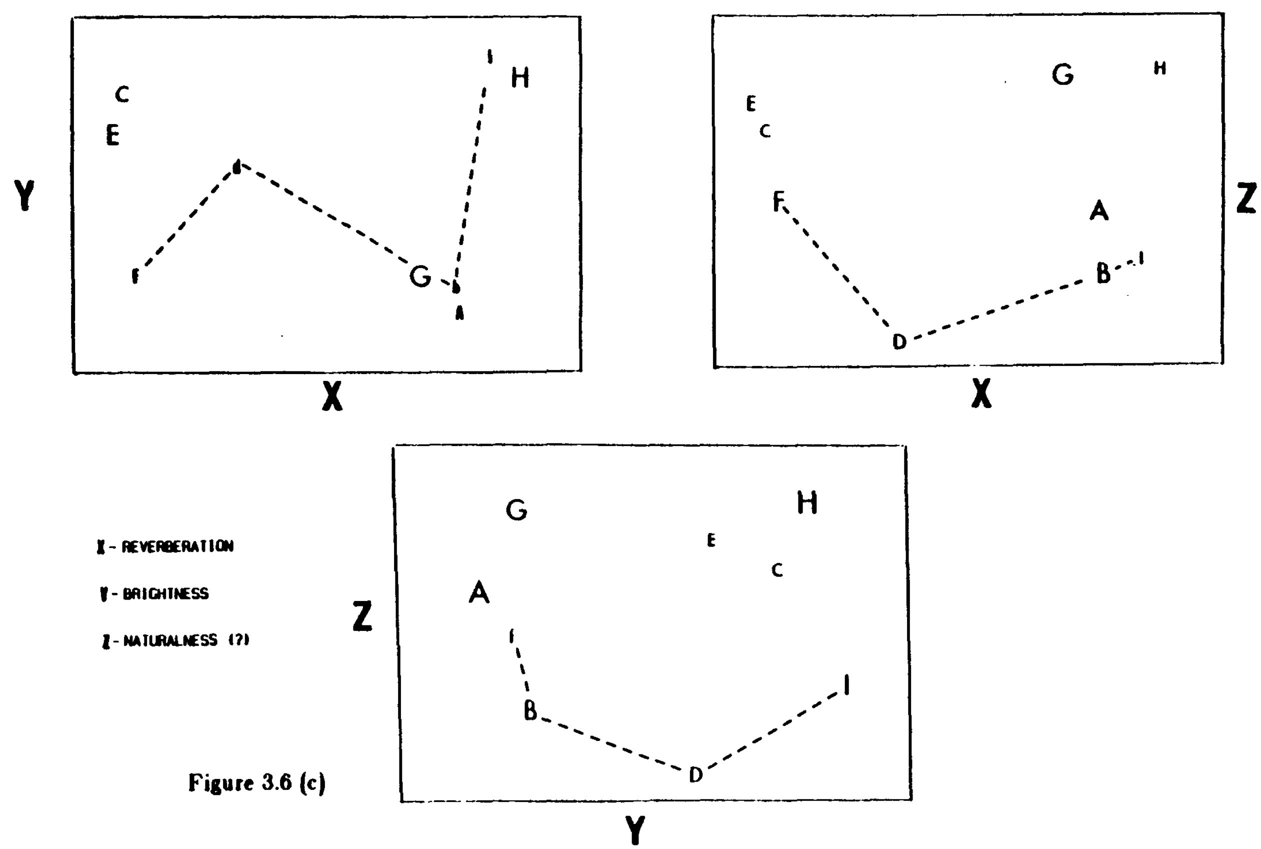


Figure 3.6 (c)

KYST multidimensional scaling solution in 3 dimensions, plotted in three 2-dimensional pairs. The third dimension is again suggested by the size of the identifying letter and the real halls are again connected by the dots.

General conclusions concerning this study

1) The most salient characteristic for all listeners, when asked to differentiate among listening spaces, is *reverberation time*. The variance among different halls that we commonly encounter is enormous, and a difference of one-half second or more is difficult to miss. In addition, as a general measure of room absorption, reverberation time relates closely to the ratio of reverberant to direct signal energy.

2) Perceptual distinctions between large halls on the basis of spectral energy distributions are more difficult to define, but it is now apparent that the assessments are made on the basis of two criteria. One is the *number of resonant modes* in the hall. By this we mean the ability of the room to reflect many different frequencies simultaneously. This is best observed in halls with longer reverberation times. The other criterion appears to be the *distribution of resonant modes*. If the hall tends to emphasize certain frequency ranges over others, this may be perceived as a hall that "rings" or has an "uneven" response to different types of sound sources.

3) The adequacy of our models for further study in the area of distance perception has been demonstrated. Although listeners did appear to distinguish the real halls from the models, it is apparent that the distinctions were tertiary and were based in large part on secondary cues, not relating to the quality of the synthetic reverberation.

4. Varied Reverberation Conditions: Subjective Control of Loundess

4.1 Introduction

The next phase of the research reported here involved the use of simulated reverberation in studying the effects of different reverberation conditions upon perceived distance. The *primary* focus at this time was to demonstrate that *the degree of reverberation associated with the direct sound will affect a listener's ability to discriminate distance differences.* A second goal was to determine the *range of potential source distances* that could be generated under different reverberant conditions. An experiment was designed to explore both these issues.

The subject's task was to adjust the direct signal amplitude in order to place the apparent source of the sound at twice the distance of the previous source from the listener. This amplitude change caused a commensurate change in the ratio of reverberant to direct signal and the reverberance of the hall affected this rate of change. The reverberance also affected the overall range of the ratios occurring in any single condition. There was no direct estimation of magnitude, nor was the listener required to meet any criteria as to the range or the number of positions that he should be able to define. Thus, if the listener was unconvinced by the illusion, or if he was not able to generate a source at the next distance, he was not encouraged to continue with that trial.

4.2 Experimental Design and Methods

In experimental design, it is critical that the task be based on assumptions that do not create response biases in the subjects, unless those response biases are specifically under study. This is particularly true of studies that use methods of adjustment, as the range of possible responses allows the subject an inordinate amount of control over the experimental results. However, this potential problem has an advantageous side as well, as the subjects are not so constrained by the method and the experimenter can evaluate individual differences among subjects more effectively. As these methods permit a more precise subjective control over the experimental variable in question, they are likely to yield correspondingly better data.

This study was designed to test several hypotheses.

1) Distance judgments differ under different reverberant conditions.

2) Differences in these conditions will affect both the potential range of apparent distances that listeners can generate and the acuity of apparent distance judgments.

3) *Loudness differences provide the strongest relative cue to distance differences.*

4.2.1 Listening Station: Hardware and Software Configuration

The physical listening station for this study was designed to meet several pre-established criteria.

a) It allowed interactive control of the direct signal amplitude and immediate audition of the adjusted stimulus.

b) It suppressed room cues from the listening environment. Headphones were not a satisfactory solution for reasons already described.

c) It gave the subject the freedom to pace himself and to develop his own experimental strategy.

The physical station consisted of a pair of monitor loudspeakers that effectively eliminate external room cues when used in near-field listening (MDM-4 loudspeakers from Calibration Standard Instruments). They were mounted approximately 4 feet apart and directed towards the listener, whose head was centered 2.5 feet from each loudspeaker. Directly in front of the listener, between the loudspeakers, were a CRT and keyboard connected to the CCRMA mainframe computer. This computer is interfaced to the digital signal processor and, during the course of an experimental trial, had exclusive access to it. The output of the signal processor returned to the listening station and was controlled by a high-quality mixing console, so that listeners could adjust the overall gain of the playback to a comfortable level. The digital output of the signal processor was 14 bits (= 84 dB SNR) before conversion to analog. (Although this was adequate, 16 bits (= 96 dB SNR) or more would have been preferable, as it was still possible to hear digital noise resulting from round-off error in the low-order bit of the signal processor. At 96 dB, for our listening conditions, this noise would have been below threshold).

After the inital 'master' output level was set, all listener control devolved to the use of the computer keyboard. An interactive program with a control structure familiar to all of the subjects was used in the experiment. When one of the nine reverberation conditions was selected, the signal processor was 'patched' immediately in the correct configuration for those conditions. The listener then indicated that the computer should process the source signal through that room simulation, reapeating the process approximately every

6 seconds. At this point, control of the direct signal amplitude was passed to the listener, and he was free to adjust it in order to satisfy the experimental task.

4.2.2 Subjects

The subjects used in this study were all normal hearing males with years of experience working with digital sound systems. This emerged as a necessary condition for several reasons.

a) Although naive listeners could be trained to work in the modeled environment, this type of task would require more practice than was feasible, given the time-sharing nature of the computer sound system.

b) The interactive controls were designed for people familiar with the CCRMA music system. Again, training would have required more time than was available.

c) The task was expected to prove difficult for some subjects. Given the untried and speculative nature of this research, it seemed probable that a known and homogeneous pool of subjects would be more useful than the cross-section provided by the uncontrolled solicitation of normal hearing adults.

4.2.3 Reverberator Conditions

A set of 9 reverberation models was created. They differed from each other in only two physically describable ways, reverberation time and ratio of reverberant to direct signal. Low, medium and high values were chosen for each of these two parameters, and the combinatorial pairs formed the 9 models. Table 4.1 illustrates these conditions, listing them in approximate order of reverberance.

Table 4.1

Group	Condition #	Reverberance	Ratio : Reverberation Time
	(4)	→ .0425	Low : Low
1. Low	(1)	→ .0450	Low : Medium
	(5)	→ .0475	Low : High
	(6)	→ .1275	Medium : Low
2. Medium	(2)	→ .1350	Medium : Medium
	(7)	→ .1425	Medium : High
	(8)	→ .2125	High : Low
3. High	(3)	→ .2250	High : Medium
	(9)	→ .2375	High : High

The nine reverberant conditions for the adjustment study are classified above in 3 groups, depending upon the reverberance associated with that condition. Reverberance refers to the product of the reverberation ratio (low = .05, medium = .15, or high = .25) and the gain term that determines reverberation time (low = .85, medium = .9, or high = .95). In the last column, the level of reverberation ratio (low, medium, or high) is associated with the level of reverberation time (low, medium, or high) for that condition.

The specific value associated with each condition is the product of the actual ratio used and the dimensionless gain term associated with reverberation time. It can be seen that they fall into three groups, low, medium, and high ratio conditions, with a range of reverberation times represented in each group. The net effect of these combinations was to create the impression of three classes of rooms, and for each class there existed a range of decay times. It is important to remember that because these two parameters interact, a given reverberation time parameter is relative and will represent a much shorter decay in the 5% condition than in the 25% condition. All nine sounded realistic, and the general impression that one had when comparing them was differences in reverberance, the subjective sense of the "fullness" and "resonance" of the space.

4.2.4 Procedure

The experimental task, as outlined previously, was quite straightforward. An actual trial took the following form.

1) The subject set the master gain to a comfortable level, by listening to the direct signal at its highest intensity as it was processed through the current room simulation. The source image appeared directly in between the two loudspeakers and very close to the listener. This was the loudest sound that the listener would encounter, and he was instructed to find a 'comfortable' and 'realistic' setting. The amplitude of the direct signal had been normalized, internally, to a maximum of .99 and this value was displayed on the CRT. Each change in amplitude made by the listener after this point was immediately reflected in the display, and the direct signal was scaled accordingly. It was unfortunate that this was necessary, but the interactive system required the use of the monitor. Listeners were instructed that doubling the distance by numerical division would not help them, as amplitude adjustments were being made internally (this was not the case). It was later determined through data analysis that listeners' results were not biased towards numerical division for the distance doublings.

2) The subject familiarized himself with the keyboard controls.

3) The stimulus began playing, repeating every six seconds.

4) At the first, loudest level, the subject localized the apparent source of the sound in front of him and then decreased the amplitude of the direct sound by an arbitrary amount. This decrement required at most two keystrokes and could be done in either a step-wise manner or a 'random-access' manner.

5) The subject 'marked' a current amplitude level and then flipped back and forth between the original level and the current level, adjusting the current level until the source appeared to be twice as far away.

6) Once this point was established, its amplitude level was written to the subject's response file on the computer.

7) The same process was then repeated. The subject decremented the amplitude again, attempting to double the previous distance between the apparent source and himself.

7) This process was repeated until the listener felt that he could not make the source recede any further or until the illusion of a focused source broke down.

Thus, the desired relationships between sources defined a geometric progression, defined as 2^{n-1}, **n=1,2,3..**, where n is an integer representing the number of positions at which the listener was able to place a source before the illusion broke down. The function was assumed continuous, but measures were taken only at integer points. There was no specific magnitude associated with the increment, however. For each person, the relationship depended upon the apparent location of the nearest source. Since there is little evidence that people associate auditory distance information with physical units of distance, it made more sense to ask for judgments based on distance relationships.

4.3 Results

4.3.1 Introduction

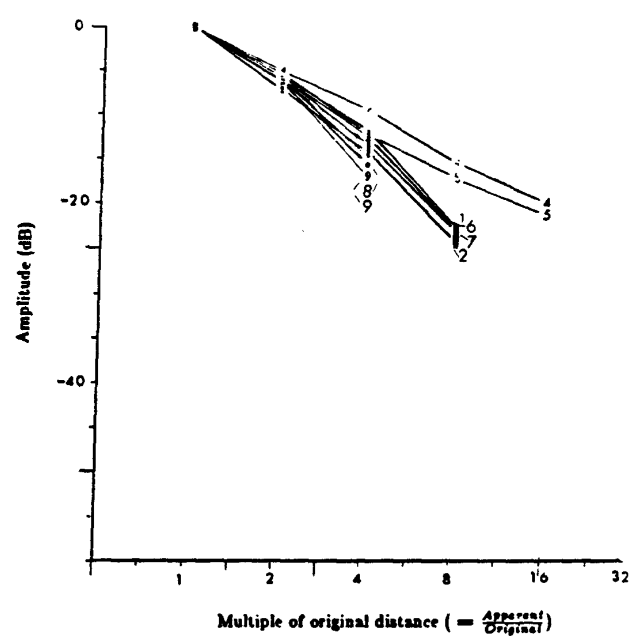
A total of 8 subjects performed the experiment. The mean duration of the experimental session was just under an hour, with several subjects finishing in about $\frac{3}{4}$ hour and several requiring $1\frac{1}{4}$ hours. The 9 experimental conditions yielded quite different results, supporting the contention that reverberant conditions can affect the *potential effective range* of apparent distance judgments. In general, the *depth* to which the sources penetrated varied from condition to condition, with a minimum of twice the original distance ($= 2^1$) represented and a maximum of 32 ($= 2^5$ multiples of the original distance.

In the following subsections, the data is considered in terms of the primary differences exhibited across the nine reverberation conditions. These were the *depth of source penetration* into the illusory space, the *form of the amplitude:distance function*, and the *individual differences among subjects*.

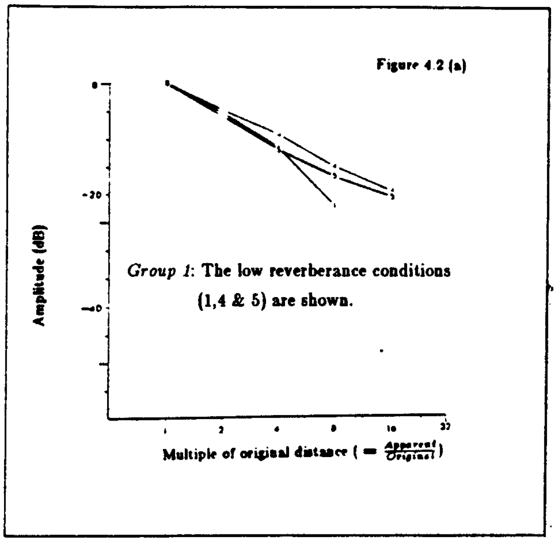
4.3.2 Depth of Source Penetration

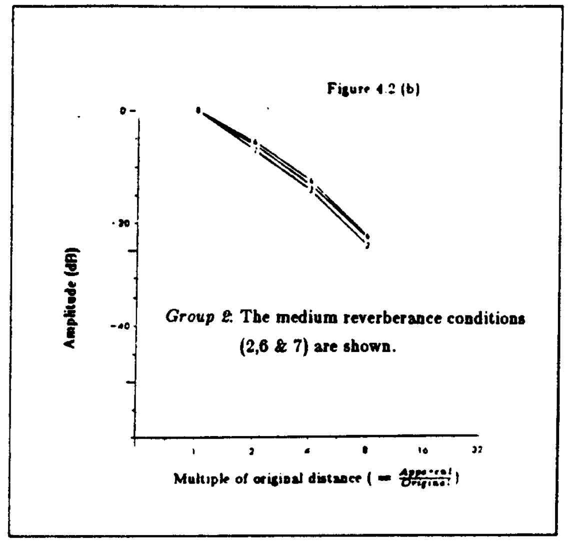
The number of source positions that the subjects were able to define in each reverberant space is the first and most obvious point to consider. The average penetration, across all subjects, for each condition, is charted in figure 4.1. In addition, the 9 conditions are divided into 3 groups of 3. The groups are formed as a function of the *product* of the reverberation *ratio* and a dimensionless gain term that determines the reverberation *time*. This product is useful as a measure of the overall *reverberance* of the condition. It should be remembered that the reverberation ratio refers to the ratio of reverberant to direct signal energy, while the reverberation time refers to the time it takes for the reverberant sound energy to decay 60 dB. The data for the individual groups of 3 are plotted in figures 4.2a, 4.2b, and 4.2c. The mean penetrations for all subjects are shown in figure 4.3, separated into these 3 groups. Several patterns are immediately apparent.

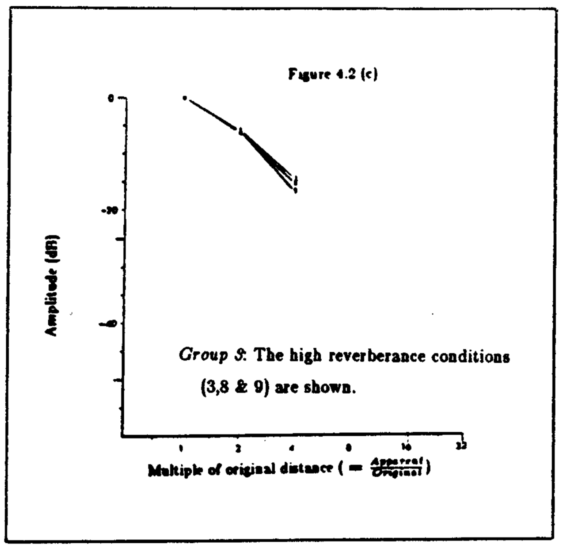
Figure 4.1



Comparison of the effect of various reverberation conditions on the range of movement of the source. The means of all subjects for all nine conditions are shown. Each line represents the mean for all subjects on the condition associated with that line. If 3 or more subjects on a given condition were able to place a source further than the others, then the mean was taken for those subjects and that data point was used. Thus, in condition number 5, five of the six subjects placed the source at a distance multiple of 16; the sixth did not. In that case, the mean was taken for the five who did, and that data point was included.



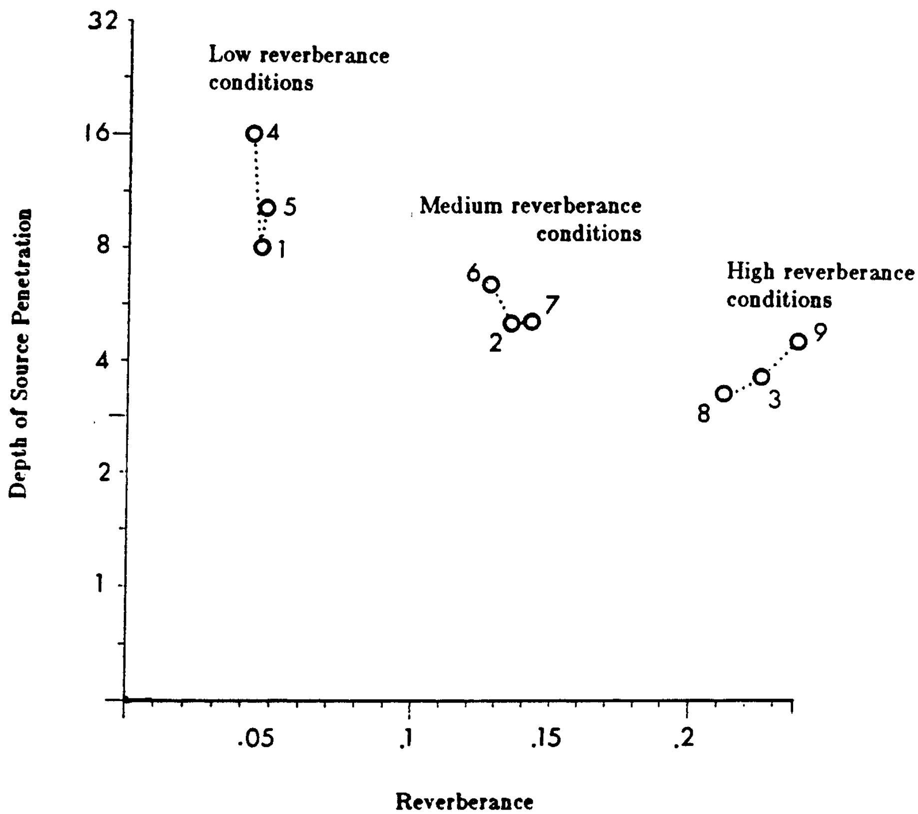




A comparison of the effect of various reverberation conditions on the range of movement of the source. The data is separated into the low, medium and high reverberance groups. Each line represents the mean for all subjects on the experimental condition associated with that line. If 3 or more subjects on a given condition were able to place a source further than the others, then the mean was taken for those subjects and that data point was used. Thus, in condition number 5, five of the six subjects placed the source at a distance multiple of 16; the sixth did not. In that case, the mean was taken for the five who did, and that data point was included.

Group 1, the *low ratio* group, had an average penetration approximately $2^{1.5}$ ($= 2.83$) times greater than that of the *high ratio* group. The middle ratio group mean penetration was roughly $\frac{1}{2}$ that of the low ratio group and about 1.4 times that of the high ratio group. The extent of these differences contrasts strongly with the relative similarity displayed by the means of groups classed by reverberation time, as shown in figure 4.4.

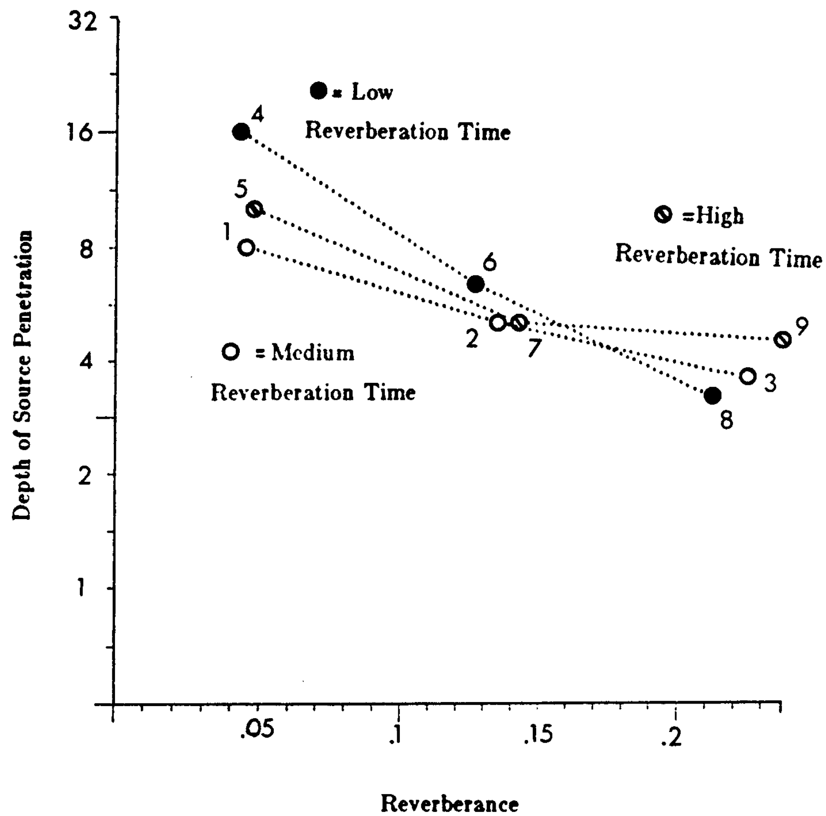
Figure 4.3



The reverberance (= the product of the gain term that determines reverberation time and the revererberation ratio) is plotted against the mean source penetration for all subjects on each condition. Individual conditions and their reverberance are shown in Table 4.1. The ordinate is again represented in terms of multiples of the original distance ($= \frac{Apparent}{Original}$)

It is apparent that reverberation *time* had less effect than the reverberation ratio on the depth of penetration. In fact, examining from (Table 4.1) the *Low:Low, Low:High* pairs and the *High:Low, High:High* pairs in figure 4.4, we see that the effect increasing reverberation time was to increase the penetration in the high ratio condition and decrease the penetration in the low ratio condition.

Figure 4.4



The reverberance is plotted against the mean source penetration for all subjects on each condition, as in Figure 4.3. However, the individual conditions are now associated by reverberation time (see Table 4.1).

The most reasonable explanation for this effect is that the conditions of a relatively high reverberation ratio appear more realistic when associated with a longer reverberation time, and a similar realism would occur for the low ratio, shorter reverberation conditions. This suggests that the realism of the simulation will affect the potential depth of source penetration. The greater the degree of realism associated with a given simulation, the more compelling the illusion and the greater the relative depth of source penetration.

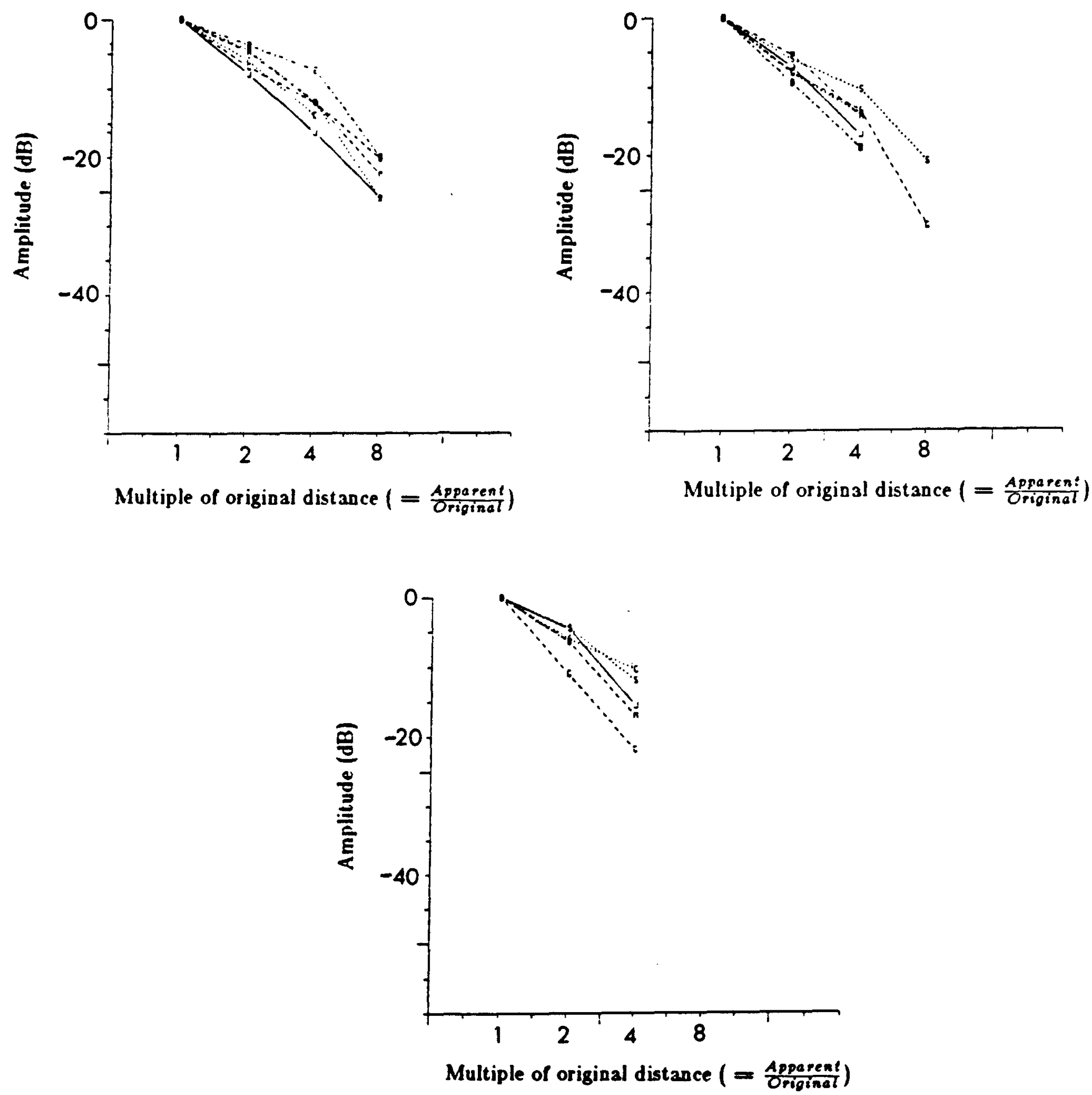
Another point that should be noted relates to the individual subjective differences, which were moderate relative to the depth of source penetration. In no condition was there a difference greater than 2, and only 3 differences of that magnitude occurred in the 9 conditions. It is interesting to note that those 3 consisted entirely of the conditions in group 1 above, the low ratio conditions. These also had the greatest source penetration, and the subjective differences may well be due to the apparent enlargement of the potential space, thus creating more room for these differences to occur.

4.3.3 Amplitude:Distance Functions

Figures 4.5a through 4.5i show the data for all subjects on all conditions, represented as a logarithmic scale of distance relationships versus *logarithmic* amplitude (dB). However, the most noteworthy features of these plots are embodied in figures 4.2a, 4.2b, and 4.2c. These represent the means for all subjects on all conditions and are separated respectively into the low, medium and high reverberation ratio groups.

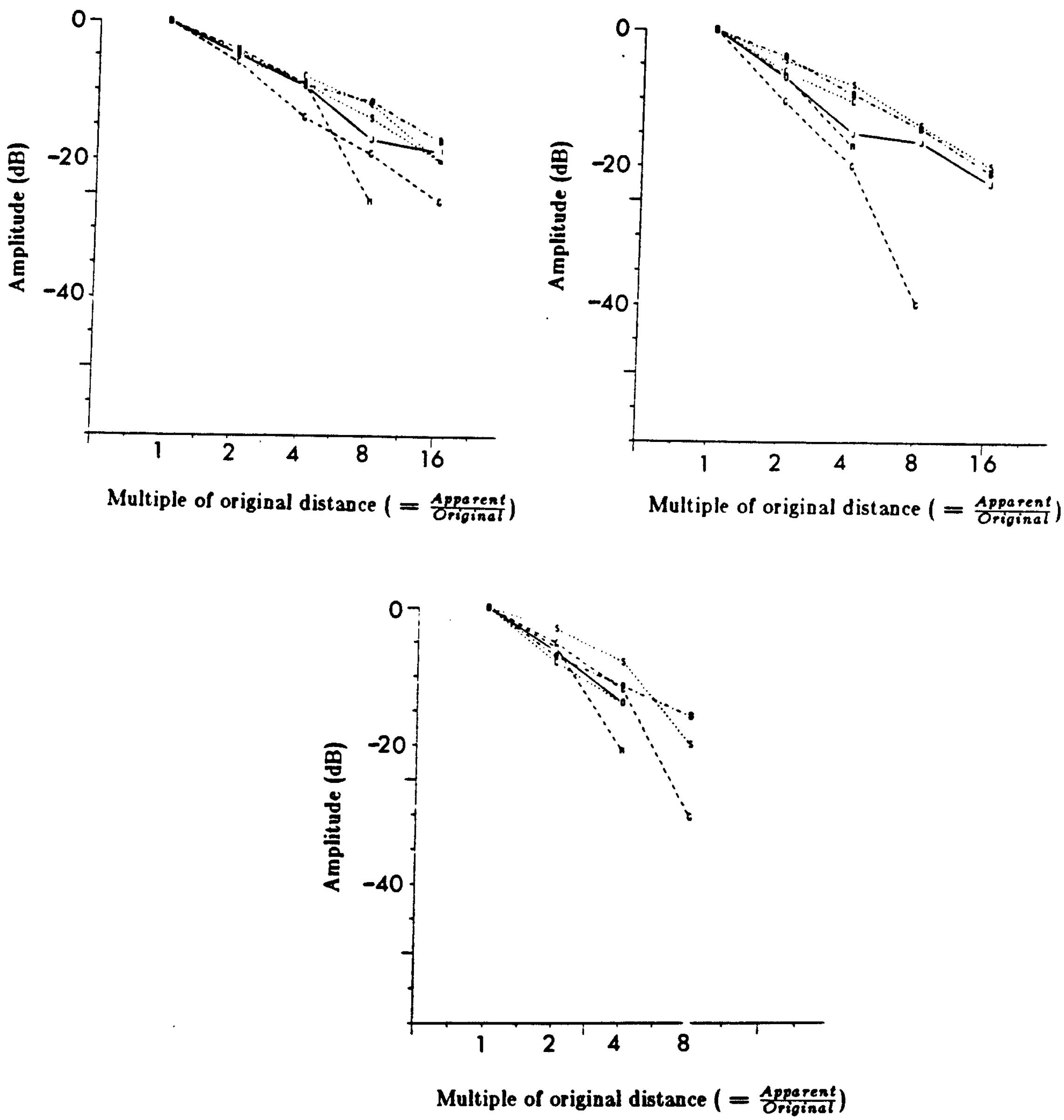
The most obvious feature is the difference in degree of source penetration across the three groups, as mentioned in the last section. As the reverberation ratio increased from 5% to 15%, the source depth was halved, and was halved again by an increase from 15% to 25%. In addition, the slopes of the three sets of group means vary in a manner suggesting that reverberation interacts very strongly with loudness to enhance or disrupt the distance cue.

Figure 4.5 (a,b,c)



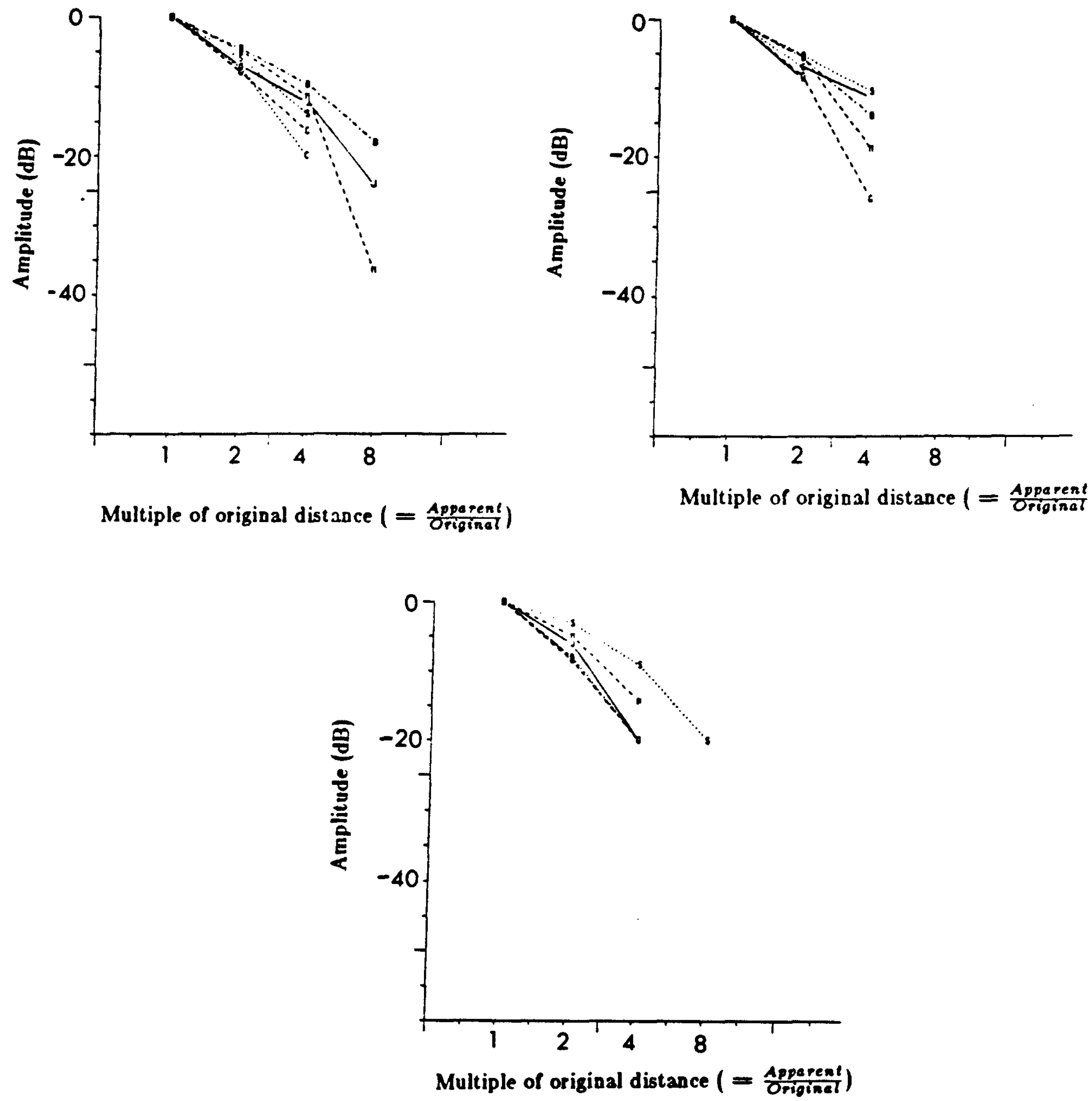
Intensity is plotted as a function of the apparent distance of the listener from the source, for the conditions specified (1,2, & 3 corresponding to low ratio:medium reverberation time, medium ratio:medium reverberation time, high ratio:medium reverberation time).

Figure 4.5 (d,e,f)



Intensity is plotted as a function of the apparent distance of the listener from the source, for the conditions specified (4,5, & 6 corresponding to low ratio:low reverberation time, low ratio:high reverberation time, and medium ratio:low reverberation time).

Figure 4.5 (g,h,i)



Intensity is plotted as a function of the apparent distance of the listener from the source, for the conditions specified (7,8, & 9 corresponding to medium ratio:high reverberation time, high ratio:low reverberation time, and high ratio:high reverberation time).

Slopes

It is interesting to note the manner in which the slopes differ for these three groups. In general, the greater the reverberation ratio, the greater the requisite decrease in intensity to double the source distance. In *every* condition, the mean first doubling occurred at -6dB, relative to the original. At this point, however, there is a divergence that apparently depends upon reverberation ratio. For the lowest ratio, the second source distance also required a 6dB decrement. After this point it appears that a smaller decrement was sufficient to perform the distance doubling. A slope of 3 or 4 dB/doubling was adequate. For the middle and high reverberation ratios, after the first doubling, the slopes increased to about 8dB/doubling and 10dB/doubling, respectively. These shifts appear to support and complement the suggestion from the last section that *high levels of reverberance interfere with the formation of distance judgments.* Under conditions of high reverberance, direct signal attenuation for a given doubling must be increased proportionally. It *may* be that this proportional increase is compensating for the lack of change in secondary factors such as spectral attenuation of the direct signal.

Loudness

The loudness of the direct signal, as demonstrated by the amplitude attenuation functions just described, is clearly not an absolute cue to distance perception. In each room condition, once the source had receded from the 'near field' (past the second distance), a different degree of attenuation was required for doubling the distance. As a consequence, is not easy to define the precise relationship between the *intensity* of the direct sound for a given condition and its *loudness.* Taking into account that scalings of -6dB to -10dB have been proposed for half-loudness under free-field conditions, several conclusions may be drawn.

> **1)** Under conditions of moderate to heavy reverberation, source intensity decrements in the range of 6dB to 10dB appear to be crudely related to doubled source distance. These figures correspond to those suggested by Warren and Stevens respectively for half-loudness judgments. The implication is that as reverberance increases, the listener is obliged to attend more closely to the direct portion of the sound in making the corresponding double distance judgment. A second possible explanation, mentioned previously, is that the increased attenuation is simply compensation for the fact that other cues, and spectral cues in particular, are not available.

2) Under conditions of low reverberation, a different rule seems to apply. Source intensity decrements in the range of 3dB to 4dB provide a more compelling impression of a source at twice the distance of the original. In the case of a real source receding from the listener, the actual physical energy decrement of the direct sound is roughly 6dB. Thus, it appears that reverberation acts as a mediating factor and is an *additive* cue. This *interaction* of loudness and reverberation presumably promotes the use of both cues in a way that is not possible when the reverberation is at a higher level and is effectively masking some direct cues.

3) Loudness differences, in all cases, are demonstrated to be the cues that most effectively suggest the distance differences. The *magnitude* of the loudness difference required to generate a certain relationship between two sources is not at all predictable, unless the room conditions are known in advance.

4.4 Just-Noticeable Differences

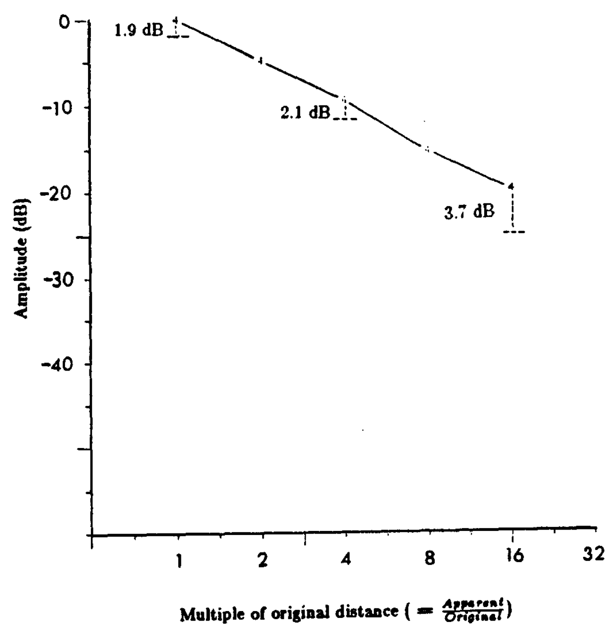
A separate study was performed that explored the degree of perceptual acuity for distance differences. This was necessary in order 1) to determine the just-noticeable differences associated with distance perception in these modeled environments and 2) to demonstrate that the furthest source under each condition was not simply the result of a boundary effect but had been placed inside a region that permitted distance discrimination.

Three conditions were tested, each condition representing one of the three reverberance groups described previously. These three corresponded to the *low:low, medium:medium, high:high* conditions of table 4.1. The mean amplitudes of all subjects for each condition in the earlier experiment were selected as a reference amplitudes. Thus, for the low reverberation condition, as all subjects were able to double the distance 4 times, 5 amplitudes were selected (.99, .56, .33, .14, .09). For the medium and high reverberation conditions, subjects were consistently capable of doubling the distance twice, yielding 3 mean amplitudes for each condition (.99, .44, .19, and .99, .49, and .09 respectively).

Listeners were seated as in the earlier studies in this section, but the experimenter controlled the amplitudes, alternating back an forth between the reference amplitude and one slightly lower, gradually decreasing the second until the listener claimed to have heard a difference in distance, as opposed to amplitude. Repeated measures were taken, but no attempt was made to determine the JND for the reverse condition, in which the amplitude would be *increased* relative to the reference level.

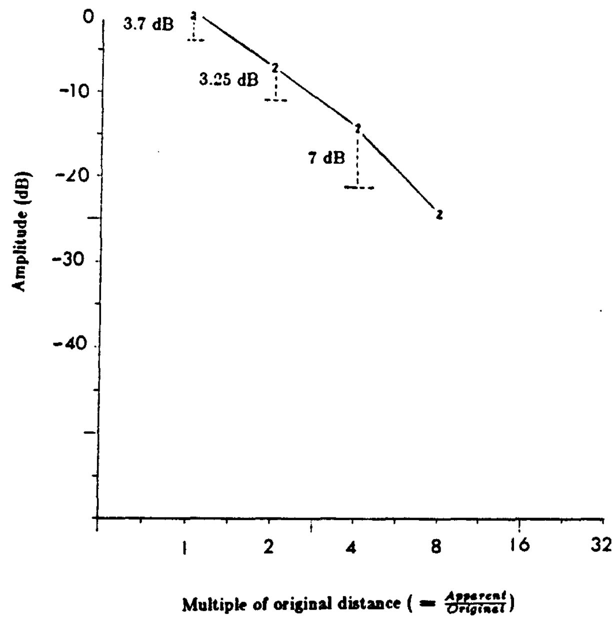
The results are displayed in figures 4.6a, 4.6b, 4.6c. In each figure, the JNDs in dB are plotted at the multiples of the original distance that were used in that reverberant condition. At least 2 conclusions are suggested by the data.

Figure 4.6 (a)



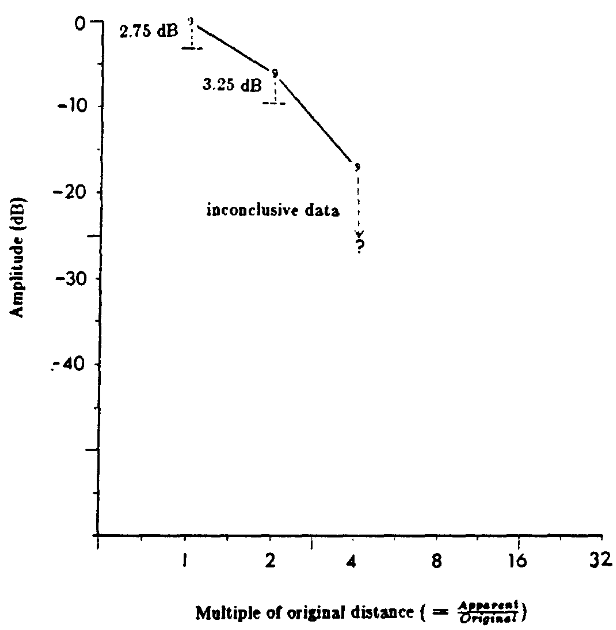
Just Noticeable Differences at 3 source positions, the first, third, and fifth depths selected by subjects in condition *no. 4*. The JNDs are given in dB. The reference intensity used at each depth was the mean of all intensities associated with that apparent distance in the subjective adjustment study described in sections 4.2 and 4.3.

Figure 4.6 (b)



Just Noticeable Differences at 3 source positions, the first, second and third depths selected by subjects in condition *no. 2*. The JNDs are given in dB. The reference intensity used at each depth was the mean of all intensities associated with that apparent distance in the subjective adjustment study described in sections 4.2 and 4.3.

Figure 4.6 (c)



Just Noticeable Differences at 3 source positions, the first, second and third depths selected by subjects in condition *no. 0*. The JNDs are given in dB. The reference intensity used at each depth was the mean of all intensities associated with that apparent distance in the subjective adjustment study described in sections 4.2 and 4.3. At the third depth, there was significant variance among subjects. As a result, this data point was not included.

1) Approximately 3 JNDs exist between the double distance positions of the earlier studies, but only for the first 4 or 8 multiples of apparent distance. As the number of apparent distance multiples increased beyond 8, the JND appeared to increase and the number of JNDs between double distance positions is less than 3.

2) If we accept the suggestion that discriminable differences in distance occur roughly 3 times before a doubling of distance is heard, it is possible to make an educated guess as to the degree of precision that listeners bring to distance inferences, relative to other perceptual modes. In general, it appears that distance discriminability is inferior to that of intensity, pitch, and azimuth discriminability, as expected. It may also be concluded that, in contrast to the visual domain, auditory distance discrimination is greatly inferior, serving in human localization to augment and enhance the spatial percept that is defined primarily by the visual system.

4.5 Summary and Predictions

It has been demonstrated that reverberant conditions do affect a listener's ability to discriminate distance differences. Increased reverberation influences the depth to which listeners feel capable of localizing a source. In addition, the direct signal attenuation necessary to achieve the placement of a source at a given position will be greater as reverberance increases.

The question was raised during the course of these studies as to the robustness of experimental results obtained with these very new modeling techniques. It was simply not possible to know whether the results could be generalized to 'real-life' listening, and that verification is left to another researcher with a great deal of time on his or her hands. It was possible, however, to attempt a corroboration of the study just described. Two hypotheses were made on the basis of the experimental data described in this chapter, and an experiment (described in the next chapter) was designed to test these hypotheses.

The first hypothesis, stated simply, is that there is an *optimum range of reverberance* within which listeners are able to take advantage of reverberation as a distance cue.

The second hypothesis, harkening back to the Coleman [1962] and Mershon and King[1975] studies, was that *reverberation has an absolute effect on perceived distance, independent of individual subjective differences or response biases.*

5. Varied Reverberation Conditions: Magnitude Estimation of Source Penetration

5.1 Introduction

In the last chapter, two hypotheses were proposed for testing, as a corroboration of the robustness of both the methods and the results. The study described below served this purpose, providing a complementary set of data in support of the earlier experiments.

The hypotheses to be tested were that **1)** reverberation serves as a distance cue within an optimum range of reverberance, and **2)** that reverberation has an absolute effect on perceived distance, independent of individual subjective differences or response biases.

5.2 Experimental Design and Methods

The method of *magnitude estimation* was selected as a complement to the subjective adjustment of the earlier study. However, rather than direct estimations of source distance, a more appropriate measure seemed to be the range defined by a set of sounds receding from the listener.

Three source intensities were selected from the means of the low reverberation conditions (conditions 1, 4, and 5) in the last experiment. The intensities selected were those used in the placement of sources at the first, third, and fifth positions (corresponding to source *amplitudes* of 1, .25, and .0625). The intensity relationships correspond roughly to the *average* -6dB/distance doubling discussed in the last chapter. Consequently, since the distance had quadrupled in each case, the second had $\frac{1}{16}$ the power of the first and the third had $\frac{1}{16}$ the power of the second.

Nine reverberation conditions were selected, slightly exceeding the range of percentages used in the earlier study, but varied by the experimenter only in percentage. Increments of 3% were used, yielding reverberation ratios of 0%, 3%, 6%, 9%, 12%, 15%, 18%, 21%, and 24%.

A single trumpet tone was selected and digitally edited to include the attack and initial steady state portions of an E above middle C. The cutoff was a sharp transient, smoothed so that the resultant 400 millisecond tone sounded like a very short articulated trumpet tone, played mezzoforte. This tone, with an amplitude of 1.0 (A1) was then copied and scaled to amplitudes of .25 (A2) and .0625 (A3), and the three copies were concatenated so that they could be played in sequence, each separated from its neighbors by 1.6 seconds of silence. The set of three sounds was then processed by each of the nine reverberator conditions and the resulting combination of direct and processed sounds were copied to digital storage in nine *sound files* for later playing.

The nine sound files, representing the nine reverberator conditions, were randomized and presented to listeners. A four second pause occurred after each condition, to allow time for magnitude judgments. Eight measures were made for each condition, producing a total of **72** judgments, not including 18 earlier practice judgments. Listeners were asked to rate, on a scale of 1 to 50 the magnitude of the apparent distance defined by the first and third sources in each condition. If there was no *depth* defined by the three, then a **1** was to be selected. If the range was *very great*, a **50** was appropriate. For anything in between, the listeners were allowed to select their own range, scale, increment, etc. Essentially, no constraints were placed on the listeners' responses, other than an insistence that they consider the *distance or range* of the sources, not 'size' or 'reverberance' of the hall, etc. It was explained that response strategies would not work. They had to listen for depth, and respond accordingly. This scale and method were chosen for several reasons.

a) A forced choice scheme would have reduced the individual differences to a more limited range than would most likely be represented. The large scale allowed listeners a great deal of flexibility.

b) Real units of magnitude, such as feet or meters, might have proved useful, but in the experimenter's experience, listeners are much more comfortable in dealing with *proximity* relationships than real units of distance.

c) A range of nine reverberation ratios, falling between *0* and *24%*, provided sufficient precision to allow evaluation of reverberation both as an absolute cue and as a relative cue with an optimum range of effectiveness.

5.3 Listening Station, Procedure, and Subjects

As in the earlier studies, a pair of 'near-field' monitor loudspeakers were used, at a distance of about 2.5 feet from the listener's head. The stimuli were played via the computer/signal processor patch described earlier, although the experimenter was in the room, initializing each set of twelve judgments and allowing a minute of rest at the end of each set. This corresponded to a total elapsed time for the experiment of 18 minutes, not including the practice and instructions.

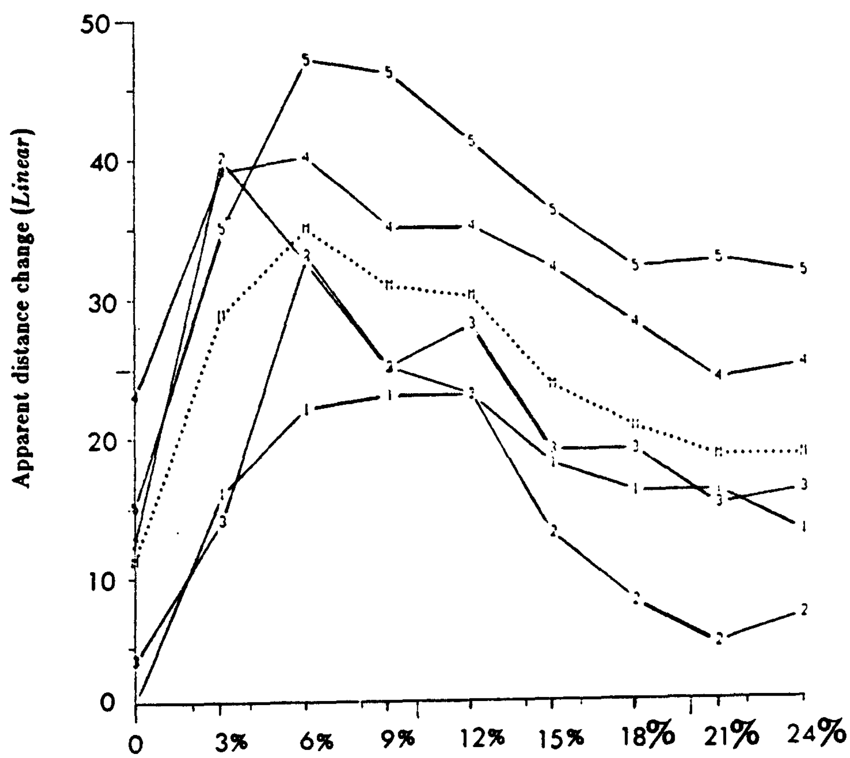
The subjects were essentially the same as in the two earlier studies, although two people who had not participated before replaced two who were unavailable.

5.4 Results

5.4.1 Introduction

The results for individual subjects can be seen in figures 5.1 and 5.2. The first is plotted with a linear ordinate, the second with a logarithmic (base 2) ordinate, since we have no way of knowing whether subjects tend to hear linear distance relationships or ratio relationships when asked to perform such a task. The function for each subject represents the mean of that subject's eight measures for each condition.

Figure 5.1

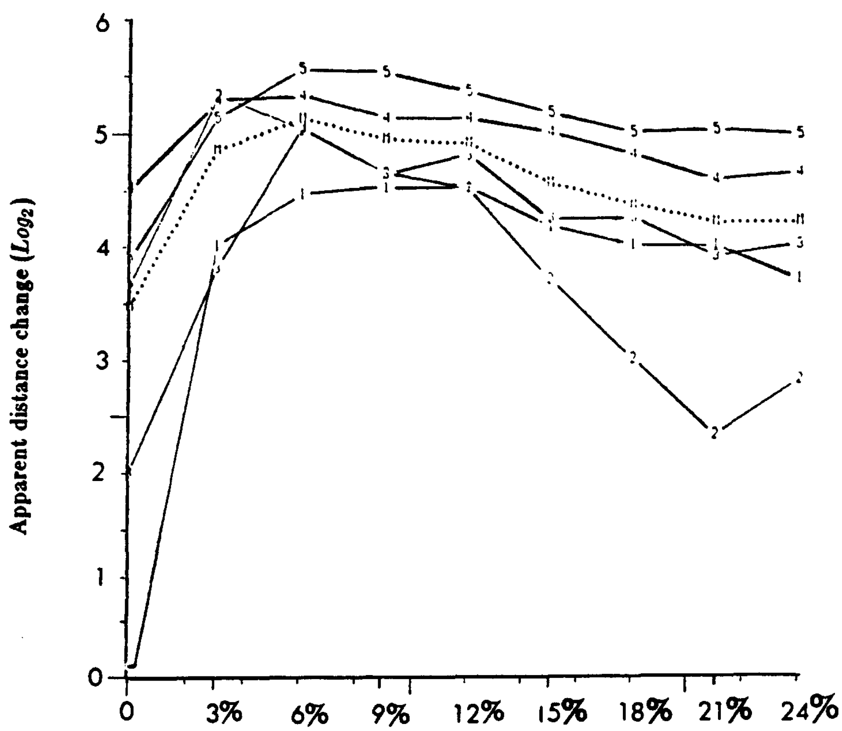


Reverberation Ratio (Percent = $\frac{Reverberant}{Direct}$ signal energy)

Range of apparent distance spanned by 2 successive 6dB decrements as a function of reverberation ratio (see text for description). The function for each subject represents the mean of that subject's eight measures for each of the 9 ratio conditions with the mean for all subjects represented by the dotted line. The ordinate is not a measure of specific units, but is a measure of relative magnitude. Each subject established his own scale.

Immediately obvious is the contrast between 1) the individual *differences among subjects* in terms of the numeric range defining the apparent distance change and 2) the *similarity of the form of the function shared by all subjects*. This is particularly clear in the figure with the linear ordinate (5.1). The hypotheses proposed at the beginning of this chapter are supported by this data and will be dealt with individually below.

Figure 5.2



Reverberation Ratio (Percent = $\frac{Reverberant}{Direct}$ signal energy)

Range of apparent distance spanned by 2 successive 6dB decrements as a function of reverberation ratio (see text for description). The function for each subject represents the mean of that subject's eight measures for each of the 9 ratio conditions with the mean for all subjects represented by the dotted line. As in Figure 5.1, the ordinate is not a measure of specific units, but is a measure of relative magnitude. This measure is represented as the logarithm$_2$ of the assigned value. Each subject established his own scale.

5.4.2 Optimum Range of Reverberation

Earlier in this thesis, it was suggested that reverberation could serve as either an impediment or as a boon to auditory distance perception. The last experiment was designed to elicit, by a totally different method, evidence corroborating the suggestion from the earlier study that the *depth of penetration of the source sound* is a useful indicator of the degree to which reverberation is aiding or hindering the formation of distance judgments.

The form of the graphical data in figures 5.1 and 5.2 make it quite straightforward to assess the general characteristics of an optimum range of reverberation. The means for all subjects are represented by the dotted lines in both plots. For all subjects, the greatest area of penetration occurred in the reverberation range **3%** to **12%**, with an apparent peak at **6%**. Below **3%**, the next value was **0**. For this anomalous condition, not only does the data suggest a very low level of penetration, but *all* the subjects mentioned afterwards that their depth ratings for the non-reverberant condition lacked conviction. Most said it required significant imagination to assign a depth rating, but it was possible. Above about **12%**, it appears that the depth decreased gradually with increased reverberation, reaching a *local boundary* at **24%** of roughly $\frac{1}{2}$ the depth defined by the **6%** condition.

Within the optimum range, several things appear to happen. A balance exists between direct signal energy and reverberant energy, so that there is sufficient reverberance to create a realistic *sense* of spaciousness but not so much that the reflective properties of the modeled hall mask the listener's ability to differentiate the direct signal from the reverberation. At the near and middle positions simulated in these studies, this was not a problem. However, the softest direct sound, with an amplitude $\frac{1}{16}$ that of the loudest sound, was easily masked in the higher reverberation conditions. It appears that this masking effect is the greatest contributing factor to the limitation of effectiveness of reverberation as a distance cue.

It is appropriate to raise the question as to the general application of these results in considering listeners' perceptions in real spaces. Although, as mentioned earlier, the experimental data is very scarce, from our own experiences we know that a variance in perceptual acuity is quite common. Granted that there are usually visual and contextual cues to aid us in our inferences, it is still true that in a highly reverberant room or church, we have a greatly reduced sense of the breadth and locus of the sound source.

5.4.3 Reverberation as an Absolute Cue

In the case of all but one of the subjects, the mean depth of source penetration was no greater than $\frac{1}{2}$ the depth of the source at the high reverberation boundary. When considered with the results of the last section and the listeners' comments concerning the difficulty in assigning any depth at all to the non-reverberant condition, this evidence suggests that reverberation serves as both an *absolute cue and a relative cue* to perceived distance. It has been demonstrated by several researchers [Coleman-62; Gardner-69a; Mershon and King-75] that distance inferences in free-field listening are very difficult. The postulation that it is merely the presence or absence of reflected sound that affects perceived distance is not sufficient.

5.4.4 Individual Differences

The lack of a consistent range accross subjects in the selection of depths proves useful in the formation of a global theory of auditory distance perception. Clearly, subjects approached these tasks with different mental sets and different notions of the appropriate 'scale' or 'magnitude'. *Despite these differences,* however, there is compelling data to suggest that *distance judgments are made according to an independent criterion that supercedes these differences.* This criterion is the *the ratio of reverberant to direct signal energy at a given moment, relative to the recent past.* On the basis of a recent set of events, an auditory *context,* it appears that listeners have very similar experiences, although they may associate that common experience in radically different ways. Hence the variety of subjective ranges and the limited variance in the overall perceptual experience.

6. Conclusions

6.1 Reverberation Modeling

Reverberation modeling has been treated here as a subset of the much more general field of *acoustical modeling.* An extraordinarily complex set of physical characteristics has been reduced to a minimum number of significant perceptual features, to the extent that it has been possible given our available hardware and software resources. The comparative study discussed in Chapter 3 demonstrated the natural quality of the simulations and the perceptual similarity existing between these models and a set of recordings made in a number of different real halls.

The achievement of perceptual *similarity* represents only a first step, however, in the pursuit of perceptual *identity* that is feasible, given a computer driven sound system of sufficient sophistication. For the acoustical designer or engineer, it should be possible in the near future to program a special purpose computer with the precise *physical* details of an entire listening space. Conversely, a device could be designed that processed sound according to *psychophysical* criteria. Audition would take place in either a specially constructed room utilizing a multichannel loudspeaker configuration or over headphones, depending upon the modeling methods used. A number of authors have suggested a variety of possible methods directed towards these goals, with methods tending to vary a great deal, depending upon the interest and general orientation of the author. However, one can generalize and say that all are interested in somehow fooling the ears through *perceptual data reduction* [Gerzon-73,77; Moorer-79; Stautner-82], or in eliminating the effect of the transmission line transfer function in order to create a wavefront eqivalent to that of a virtual source or set of virtual sources [Schroeder-61,62].

A macroscopic view of all these efforts suggests that as the technological resources become available, the trend in modeling will be towards systems that attempt time varying simulations of the actual wavefronts of virtual sources within illusory spaces. The key word is *simulations*, however. The method by which one could actually *create* the wavefront for a virtual source is not known at this time, and is theoretically and conceptually far removed from the current state of affairs. However, continued attempts towards perceptual modeling promise to yield simulation methods of increasing accuracy and sophistication, incorporating characteristics derived from both the acoustic and psychoacoustic media. It is at the intersection of these two converging trends that the greatest number of applications are suggested.

6.2 Distance Perception

The treatment of auditory distance perception has been restricted in this thesis to stationary sources in the median sagittal plane. Nevertheless, several critically important questions concerning the relationship between the direct signal and the reverberant signal have been answered. In addition, there now exists a body of evidence supporting a preliminary general model for auditory distance perception and suggesting a straightforward set of studies to evaluate this model. The results of the work described in this thesis are summarized below.

1) Reverberant conditions have an absolute affect on the subjective ability to discriminate distance differences. Under anechoic conditions, intensity and spectral differences are the only cues available to the listener. The elimination of reflected sound tends to compress the apparent range of the listening space, reducing potential depth and shifting the assignment of discriminable differences from a "spatial" mode to a "loudness" mode. Under conditions of excessive reverberation, the effect is of a masking "wall", beyond which the sound source is unable to penetrate. The net effect is similar to that of the anechoic condition.

2) There is an *optimum range of reverberance* within which listeners are able to take advantage of reverberation as a distance cue. Although ideal ratios of reverberant to direct signal energy will vary from condition to condition, the data support the assertion that auditory *depth* perception and relative distance discriminability are best under conditions of low to moderate reverberation.

3) Conclusions 1 and 2 above are true, independent of individual subjective differences. In general, listeners will bring unique contextual associations to the auditory experience. This supports the available data in suggesting that individual differences will be large and unpredictable, in any absolute sense. Despite these differences, however, the *form* of the listener response is invariably the same.

6.3 General Conclusions

As mentioned above, it is possible to formulate a general theory of distance perception that is well supported by the data and that lends itself to experimental study. This model is outlined below.

It will be remembered that the initial proposition of this thesis was that distance perception in the median saggital plane operates as a function of three physical factors. These are 1) the intensity of the direct signal, 2) reverberation, and 3) spectral cues. In addition, it was suggested that the auditory context and the listener's familiarity with the source signal and environment were critical to accurate determination of proximity. As the interaural cues were not considered in this work, it is not currently possible to say what role they might play. Two pilot studies performed by this author evaluated the ability of listeners to make distance related inferences on the basis of 1) changing reverberation ratio with all other factors held constant and 2) changing high frequency spectral cues with all other factors held constant. The net result of these studies was the conclusion that, in the absence of change in direct signal intensity, listeners did not hear *distance* changes but heard qualitative changes in the *room* or the *source signal*.

If the above conclusion is applied in the formulation of a general theory, a hierarchy of cues is suggested. As suggested by numerous researchers loudness serves as the primary determinant of perceived distance. However, in the absence of reverberation, the loudness cue is "impoverished" and listeners have great difficulty hearing the loudness differences in a "spatial" mode. Conversely, under high reverberation conditions, a similar effect is encountered, such that the reverberation masks loudness cues and erects a kind of "wall" beyond which the source will not recede. Thus, *reverberation provides the "spatiality" that allows listeners to move from the domain of loudness inferences into the domain of distance inferences.* Given the above two limits, an *optimum degree of reverberance exists, within which spectral cues operate. The ability to discriminate distance differences, as constrained by the ideal operating limits of the local reverberance, is certainly secondary to the visual and contextual cues provided by other sensory and cognitive systems.*

Appendix I

by Peter Samson, Systems Concepts Inc.

SYSTEMS CONCEPTS DIGITAL SYNTHESIZER PROGRAMMING SPECIFICATION

INTRODUCTION 780714

Generators and Modifiers

The synthesizer has two kinds of processing elements: generators and modifiers. An additional type of element, termed a delay unit, is optional.

Generators produce sine, square, and sawtooth waves, pulse trains, and equal-amplitude sum-of-cosines (band-limited pulse trains); apply linear and exponential envelopes; perform frequency modulation; can automatically sweep frequency linearly; read data from computer memory; and write data into computer memory or to digital-to-analog converters. Up to 256 generators can be active at one time.

Modifiers simulate a resonance or antiresonance; perform amplitude modulation, four-quadrant multiplication, mixing, clipping, and memory (sample and hold) functions; can generate uniform noise; and pass data to and from the optional delay units. Up to 128 modifiers can be active at the same time.

Delay units have two uses: as delay lines for signals; and to hold precomputed tables, such as time-domain waveforms. Up to 32 delay units can be active at the same time.

Passes and Ticks; Sum Memory

The processing performed on a per-sample basis comprises one pass. A pass is a series of ticks, of three types: processing ticks, overhead ticks, and update ticks. Processing ticks perform the calculations corresponding to generators and modifiers, and update ticks permit performance of commands to load new parameters. Within a pass, all processing ticks are performed first, then all overhead ticks, then all update ticks. A tick of any type takes 195 nsec. The number of processing ticks per pass is the maximum of: the number of generators used; twice the number of modifiers used. For delay units, divide the number of processing ticks minus six by four to get the number of delay memory cycles possible per pass. The number of delay units that can be used is this number less however many delay memory cycles the computer may make during the processing ticks. There are eight overhead ticks per pass. The number of update ticks per pass should be chosen according to the number of processing and overhead ticks to give the desired overall sample rate.

Information is passed among generators and modifiers through a scratchpad area called sum memory, which is divided into four 64-word quadrants. In one quadrant, sums are accumulated of generator outputs during a given pass; another quadrant holds the accumulated generator sums from the previous pass. The other two quadrants act likewise for modifier outputs. Any generator or modifier can read data from either previous-pass quadrant, and any modifier can read from the current-pass modifier quadrant also.

Computer Interface

Information is passed to and from the computer in two ways: I/O instructions, and direct memory access. With the delay memory option, a low-bandwidth bidirectional 20-bit path permits read- and write-accesses by the computer.

Computer I/O instructions perform general control, status sensing, and diagnostic functions. The direct memory access path is provided for data transfer in real time. There are three types of such data transfer: commands (to the device), read data (per sample) (to the device), and write data (per sample) (from the device). Each of these three has its own word count (WC) and core address (CA) registers in the device; they are set up by I/O instructions. Commands are always 32 bits; read data may be either 16 or 32 bits, giving a choice between packed data and full precision (the left 20 bits are significant in 32-bit mode; in 16-bit mode, the left 16-bit data item precedes the right one); write data is the left 20 of 32 bits. The device has buffering for 28 commands, 4 read-data items, and 1 write-data item.

The synthesizer can be conditioned to interrupt the computer in various circumstances. One class of them can be termed data errors: arithmetic overflow during processing, and command overrun. Command overrun occurs when a Linger command is performed which specifies a pass at least 1, but no more than 4096, before the current pass. The other class of interrupt conditions relates to direct memory access. Separate indications are provided for read data, write data, and command WCs being exhausted, and also for underrun conditions. Command underrun occurs when on an update tick there is no command to be performed (normally when there is no update activity due, a Linger command is being performed). The read data and write data underrun states occur when the device must stop its clock momentarily to wait for memory access; this means the device is not operating in real time.

SYSTEMS CONCEPTS DIGITAL SYNTHESIZER ANALOG OUTPUT SPECIFICATION

The signal path for one analog output involves the following sections:

- Channel selection logic (addressing)
- Digital hold register
- Digital to analog converter
- Sample-and-hold
- Program-controlled low-pass filter
- Buffer amplifier.

Each section is specified at 25 degrees C as follows.

Channel selection logic: 4 bits (1 of 16)

Digital hold register: 14 bits

Digital to analog converter: 14 bits
Linearity: 0.005%

Sample-and-hold: full power bandwidth 0 to 40 kHz

Filter: two modes
Mode 0: 1-pole RC at 200 kHz
Mode 1: 6-pole Butterworth, 4 programmable frequencies subject to the relationships f0=A, f1=A+B, f2=A+C, f3=A+B+C; full power bandwidth 0 to 18.5 kHz max.

Buffer amplifier: output +/- 5 V max., unbalanced
Output current: 4 mA max.
Short circuit protection: to ground only
Full power bandwidth: 0 to 18.5 kHz for 10 V swing
Output source impedance: 100 ohms
Output connector: BNC jack

The following are overall figures with Mode 0 filtering:

Gain error: 2.5%

Offset error: 20 mV

Noise at sampling rate and its harmonics: 10 mV max. (RMS)

Other noise 10 Hz to 50 kHz: 1 mV max. (RMS)

Bibliography

[Arabie and Shepard-73] Arabie P., and Shepard, R.N., "Representations of similarities as additive combinations of discrete overlapping properties", Presented at Math. Psych. Meeting in Montreal (1973).

[Baker and Young-75] Baker, R.F., and Young, F.M., "A note on an empirical evaluation of the ISIS procedure", *Psycometrika*, 40 (1975), 413

[Basset and Eastmond-64] Basset,I.G., Eastmond,E.H., "Echolocation: Measurement of pitch versus distance for sounds reflected from a flat surface," *Journal of the Acoustical Society of America*, 36, (1964) 911

[Banister-26] Banister, H. "Three Experiments on the localization of tones," *British Journal of Psychology*, 16, (1926)

[Bekesy-38] Bekesy, G.Von, "Uber die Entstchung der Entfernungsempfindung beim Horen"*Akustiche Zeitschrift*,3, (1938) 21-31 also found in Bekesy, G.Von, *Experiments in Hearing*, McGraw Hill, New York (1960) 301-313.

[Blauert-69] Blauert, J. "Sound localization in the median plane",*Acustica*,22, (1969)

[Borish-83] Borish, Jeffrey, and Angell, James B., "An efficient algorithm for measuring the impulse response using pseudorandom noise," Journal of the Audio Engineering Society [Forthcoming]

[Butler-69] Butler, R.A., "Monaural and binaural localization of noise bursts vertically in the median sagittal plane" *Journal of Auditory Research*, 3, (1969)

[Butler and Naunton-62] Butler R.A. and Naunton, R.F. "Some effects of unilateral auditory masking upon the localization of sound in space". *Journal of the Acoustical Society of America*,34, (1962)

[Butler-72] Butler, R.A. "Influence of spatial separation of sound sources on auditory evoked response." *Neuropsychologica*,10, (1972), 824.

[Butler and Naunton-64] Butler R.A., Naunton, R.F. "Role of stimulus frequency and duration in the phenomenon of localization shifts". *Journal of the Acoustical Society of America*, 36 (1964)

[Butler ad Planert-76] Butler R.A., Planert, N., "The influence of stimulus bandwidth on localization of sound in space",*Perception and Psychophysics*,19,1 (1976)

[Carroll and Chang-70] Carroll, J.D., and Chang, J.J., "Analysis of individual differences in multidimensional scaling via an N-way generalization of 'Eckart-Young' decompositions",*Psychometrika*, 35,(1970), 283

[Carroll and Wish-73] Carroll, J.D., and Wish, M., "Models and methods for three-way multidimensional scaling", in *Contemporary Developments in Mathematical Psychology*, R.C. Atkinson, D.H. Krantz, R.D. Luce, And P. Suppes, eds. W.H. Freeman, San Francisco (1973).

[Chowning, Grey, Moorer, and Rush-74] Chowning, J.M., Grey, J.M., Moorer, J.A., and Rush, L., "Computer simulation of music instrument tones in reverberant environments", Stanford University Music Department Publication STAN-M-1,1974.

[Chowning-71] Chowning, J.M., "The simulation of moving sound sources", Journal of the Audio Engineering Society, 2-6, 1971. Reprinted in Computer Music Journal, Vol. 1, no. 3, 1977.

[Coleman-62] Coleman, P.D., "Failure to localize the source distance of an unfamiliar sound." *Journal of the Acoustical Society of America*, 34, (1962), 345.

[Coleman-63] Coleman, P.D., "An analysic of cues to auditory depth perception in free space", *Psychological Bulletin*, 60, (1963)

[Coleman-68] Coleman, P.D., "Dual role of frequency spectrum in determination of auditory distance", *Journal of the Acoustical Society of America*, 44, (1968)

[Cooper and Shepard-73] Cooper, L.A., Shepard, R.N., "Chronometric studies of the rotation of mental images", in *Visual Information Processing*, W.G. Chase ed., Academic Press, New York. 1973.

[Cooper and Shepard-76] Cooper, L.A., Shepard, R.N., "Transformations on representations of objects in space", *Center for Human Information Processing, report no. 59,* (April,1976)

[Deatherage and Hirsch-59] Deatherage, B.H. and Hirsch, I.J. "Auditory localization of clicks", *Journal of the Acoustical Society of America*, 31, (1959)

[Dunn and Parfitt-67] Dunn, B.E., and Parfitt, S. "Localization with complete masking",*Perception and Psychophysics*, 2, (1967)

[Ebata, Sone, and Nimura-68] Ebata, M., Sone, T., and Nimura, T., "On the perception of direction of echo", *Journal of the Acoustical Society of America*, 44, (1968)

[Feddersen, Sandel, Teas, and Jeffress-57] Feddersen, W.E., Sandel, T.T., Teas, D.C., Jeffress, L.A., "Localization of high frequency tones", *Journal of the Acoustical Society of America*, 29, (1957) 988

[Garner-58] Garner, W.R., "Advantages of a discriminability criterion for loudness scaling", *Journal of the Acoustical Society of America*, 30 (1958) 1005-1012

[Gardner-68b] Gardner M.B. "Lateral localization of 0 degree and near 0 degree oriented speech signals in anechoic space" *Journal of the Acoustical Society of America*, 44(3), (1968)

[Gardner-68a] Gardner, M.B. "Historical backround of Hass and/or precedence effect", *Journal of the Acoustical Society of America*, 43(6), (1968)

[Gardner-69a] Gardner M.B. "Distance estimation of 0 degree or apparent 0 degree oriented speech signals in anechoic space",*Journal of the Acoustical Society of America*, 45(1), (1969)

[Gardner-69b] Gardner, M.B. "Image Fusion, Broadening, and Displacement in Sound Location", *Journal of the Acoustical Society of America*, 42, (1969).

[Gardner-69c] Gardner, M.B. "Proximity image effect in sound localization." *Journal of the Acoustical Society of America*, 43(1), (1969), 163.

[Gerzon-73] Gerzon, M.A., "Periphony: With-height sound reproduction", *Journal of the Audio Engineering Society*, 21, (1973) 2-10

[Gerzon-77] Gerzon, M.A., "The optimum choice of surround sound encoding specifications", *56th conference of the Audio Engineering Society*, Preprint no. 1199 (A-5), Paris, France, March 1-4, 1977

[Gordon and Grey-78] J. W. Gordon and J. M. Grey, "Perception of spectral modifications on orchestral instrument tones," *Computer Music Journal*, July, 1978.

[Grey-77] J. M. Grey, "Multidimensional perceptual scaling of musical timbres," *Journal of the Acoustical Society of America*, May, 1977.

[Grey-78a] J. M. Grey, "Timbre discrimination in musical patterns," *Journal of the Acoustical Society of America*, August, 1978.

[Grey-78b] J. M. Grey, "Experiments in the perception of instrumental timbre," invited paper for the *Bulletin of the Council for Research in Music Education*, 1978.

[Grey and Gordon-78] J. M. Grey and J. W. Gordon, "Perceptual effects of spectral modifications on Musical Timbres," *Journal of the Acoustical Society of America*, May, 1978.

[Grey and Moorer-77] J. M. Grey and J. A. Moorer, "Perceptual evaluation of synthesized musical instrument tones," *Journal of the Acoustical Society of America*, August, 1977.

[Grey-75] Grey, J.M., "An exploration of musical timbre", Stanford University Music Department Publication STAN-M-2, 133pp, 1975.

[Grey-79] J. M. Grey, "Categorical versus continuous perception of musical timbre," For the *Journal of the Acoustical Society of America*, (in press).

[Gravetter-70] Gravetter, F.J. Lockhead, G.R. "Criterial dispersions and the stimulus range", *Proceedings of the Annual Convention of the American Psychological Association*, 5(pt. 1), (1970).

[Guttman-62] Guttman, N., "A mapping of binaural click lateralizations", *Journal of the Acoustical Society of America*, 34, (1962), 87.

[Hanson and Kock-57] Hanson,R.L., Kock,W.E., "Interesting effect of 2 loudspeakers under free space conditions", *Journal of the Acoustical Society of America* 29, (1957) 145.

[Harris-72] Harris, J.D., "A florilegium of experiments on directional hearing",*Acta Oto-Laryngologica, Supplement. No. 298,* (1972)

[Harris-60] Harris, G.G., "Binaural interactions of impulsive stimuli and pure tones", *Journal of the Acoustical Society of America*, 32, (1960), 685..

[Hartley and Fry-21]? Hartley, Fry,., "Binaural localization of pure tones", *Phys. Review.* 18, ser.2,p431 (1921)

[Hirsch-68] Hirsch, H.R., "Perception of the range of a sound source of unknown strength", *Journal of the Acoustical Society of America*, 43, (1968), 373

[Holt and Thurlow-69] Holt, R.E. and Thrulow, W.R., "Subject orientation and judgment of distance of a sound source", *Journal of the Acoustical Society of America*, 46, (1969), 1584

[Huppert and Ainfwe-68] Huppert, F., Ainfwe, G. "An aftereffect in judgement of auditory duration" *Perception and Psychophysics*, 2(11), (1968).

[Jeffress, Blodgett and Deatherage-62] Jeffress, L.A., Blodgett, H.C., and Deatherage, B.H. "Effect of interaural correlation on the precision of centering a noise." *Journal of the Acoustical Society of America*, 34, (1962)

[Johnson-67] Johnson, S, "Hierarchic clustering schemes", *Psychometrika*, 32, (1976) 241

[Jordan-70] Jordan, V.L. "Acoustical criteria for auditoriums and their relation to model techniques", *Journal of the Acoustical Society of America*, 47(2), (1970)

[Kanda, Oguro and Ohara-70] Kanda, H., Oguro, H., Ohara, T. "Fundamental frequency of signal sound and audible distance", *Journal of Science and Labour*, 46(2), (1970).

[Knudson-70] Knudson, V.O. "Model testing of auditoriums", *Journal of the Acoustical Society of America*, 47(2), (1970)

[Kruskal-64a] Kruskal, J.B., "Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis", *Psychometrika*, 29, (1964) 1

[Kruskal-64b] Kruskal, J.B., "Nonmetric multidimensional scaling: A numerical method", *Psychometrika*, 29, (1964) 115

[Lackner-72] Lackner, J.R.,"An auditory illusion of depth", *Cognition*, 1(2-3),(1972), 201.

[Leakey-57] Leakey, D.M. Further effects produced by two loudspeakers in echo-free conditions", *Journal of the Acoustical Society of America*, 29, (1957)

[Leakey-59] Leaky, D.M., "Some measurements of effects of interchannel intensity and time differences in two channel sound", *Journal of the Acoustical Society of America* 31,7 (1959) 977

[Lochner and Van Keet-60] Lochner,J.P.A., Van Keet, "Stereo and quasistereo reproduction" *Journal of the Acoustical Society of America* 32 (1960)

[Loy-77] Loy, D.G., "Reference manual and tutorial for the Systems Concepts Digital Synthesizer", Stanford University Music Department Publication STAN-M-6, 95pp, 1977.

[Mershon and King-75] Mershon, D.H., King, L.E. "Intensity and reverberation as factors in auditory perception of egocentric distance", *Perception and Psychophysics*,18,6 (1975)

[Mershon et al-80] Mershon, D.H., Desaulniers, D.H., Amerson, T.L., and Kiefer, S.A., "Visual capture in auditory distance perception: Proximity image effect reconsidered", *Journal of Auditory Research*, 20, (1980) 129-136

[Meyer and Kuttruf-62] Meyer & Kuttruf "Progress in architectural acoustics",*Technical Aspects of Sound*, Richson (ed), Elsevian N.Y. (1962)

[Mills-59] Mills, A.W., "On the minimum audible angle", *Journal of the Acoustical Society of America*, 30, (1959) 237

[Mills-60] Mills, A.W., "Lateralization of high frequency tones", *Journal of the Acoustical Society of America*, 32, (1960) 132.

[Mills-72] Mills, A.W., "Auditory Localization", in *Foundations of Modern Auditory Theory*, Vol II., Academic Press, New York (1972)

[Molino-73] Molino, J., "Perceiving the range of a sound source when the direction is known". *Journal of the Acoustical Society of America*, 53(5), (1973) 1301.

[Molino-74] Molino, J., "Psychophysical verification of predicted interaural differences in locating distant sound sources", *Journal of the Acoustical Society of America*, 55, (1974) 139

[Moore-77] Moore, F.R., "Real time interactive computer music synthesis", Stanford University Music Department Publication STAN-M-7, 109pp, 1977.

[Moorer-75] Moorer, J.A., "On the loudness of complex, time-variant tones", Stanford University Music Department Publication STAN-M-4, 18pp, 1975.

[Moorer-77] J. A. Moorer, "Signal processing aspects of computer music - A survey", Invited Paper, *Proceedings of the IEEE*, July, 1977. Reprinted in *Computer Music Journal*, Vol. 1, No. 1, (1977).

[Moorer-79] Moorer, J.A., "About this reverberation business", *Computer Music Journal*, Vol. 3, No 2., June (1979)

[Moushegian and Jeffress-59] Moushegian, G., and Jeffress,L.A., "Role of interaural time and intensity differences in the lateralization of low frequency tones." *Journal of the Acoustical Society of America*, 31, (1959) 1441.

[Nordlund-62] Nordlund, B., "Physical factors in angular localization",*Acta Oto-Laryngologica*, 54, (1962)

[Perrott and Elfner-68] Perrott, D.R., and Elfner, L.F. "Monaural localization", *Journal of Auditory Research*, 8, (1968)

[Perrot-69] Perrot, D.R. "Role of signal onset in sound localization", *Journal of the Acoustical Society of America*, 45, (1969), 436.

[Pollack and Rose-67] Pollack,I., and Rose, M. "Effect of head movement on the localization of sounds in the equatorial plane", *Perception and Psychophysics*, 2, (1967)

[Pollack-68] Pollack, I. "Auditory pulsed doppler discrimination", *Journal of the Acoustical Society of America*, 44, (1968) 592.

[Robins and Shepard-77]. Robins,C., Shepard,R.N., "Spatio-temporal probing of apparent rotational movement", *Perception and Psychophysics*, 22(1) (1977) 12

[Roffler and Butler-68a] Roffler, S.K., and Butler, R.A. "Localization of tonal stimuli in the vertical plane" *Journal of the Acoustical Society of America*, 43, (1968)

[Roffler and Butler-68b] Roffler, S.K., and Butler, R.A. "Factors that influence the localization of sound in the vertical plane" *Journal of the Acoustical Society of America*, 43, (1968)

[Ruff and Perret-76] Ruff, R.M., Perret, E. "Auditory spatial pattern perception aided by visual choices", *Psychological Research*, 38, (1976) 369

[Sabine-22] Sabine, W,C., "Theatre acoustics", *Collected Papers on Acoustics*, Cambridge, Mass. Harvard University Press (1922)

[Sandel,Teas,Feddersen and Jeffress-55] Sandel,Teas,Feddersen and Jeffress. "Localization of sound from single and paired sources", *Journal of the Acoustical Society of America*, 27, (1955), 842.

[Schroeder and Logan-61] Schroeder, M.R., Logan, B.F. "'Colorless' artificial reverberation", *Journal of the Audio Engineering Society*, 9, (1961), 192.

[Schroeder-61] Schroeder, M.R. "Improved quasi-sterophony and colorless artificial reverberation", *Journal of the Acoustical Society of America*, 33, (1961), 1061.

[Schroeder-62] Schroeder, M.R., "Natural sounding artificial reverberation", *Journal of the Audio Engineering Society*, 10, (1962), 219.

[Schroeder and Kuttruff-62] Schroeder, M.R., Kuttruff, K.H., "On Frequency response curves in rooms", *Journal of the Acoustical Society of America*, 34,(1962), 76.

[Schroeder-70] Schroeder M.R. "Digital simulation of sound transmission in reverberant spaces", *Journal of the Acoustical Society of America*, 47(2), (1970)

[Schroeder, Atal, Sessler and West-70] Schroeder, M.R., Atal, G.M., Sessler, G.M., West, J.E. "Acoustic measurements of Philharmonic Hall (NY)", *Journal of the Acoustical Society of America*, 40, (1970), 431.

[Schubert-63] Schubert, E.D., "Interpretation of the Butler-Naunton localization shifts", *Journal of the Acoustical Society of America*, 35, (1963)

[Searle-73] Searle, C.L., "Cues required for externalization and vertical localization", *Journal of the Acoustical Society of America*, 54, (1973)

[Searle-75] Searle, C.L. et al, "Binaural pinna disparity: another auditory localization cue", *Journal of the Acoustical Society of America*, 57, (1975)

[Searle-76] Searle, C.L. et al, "Model for auditory localization", *Journal of the Acoustical Society of America*, 60, (1975), 1164

[Shepard-62a] Shepard, R.N., "The analysis of proximities: Multidimensional scaling with an unknown distance function. I", *Psychometrika*, 27, (1962) 125

[Shepard-62b] Shepard, R.N., "The analysis of proximities: Multidimensional scaling with an unknown distance function. II", *Psychometrika*, 27, (1962) 125

[Shepard-72] Shepard, R.N., "Psychological representations of speech sounds", in *Human Communication*, E.E. Davis and P.B. Denes, eds. McGraw-Hill, New York (1972)

[Shepard-74] Shepard, R.N., "Representations of structure in similarity data: problems and prospects", *Psychometrika*, 39, (1974) 373

[Shepard and Judd-76] Shepard, R.N., Judd,S., "Perceptual illusion of rotation of three-dimensional objects", *Science*, 191, (1976) 952

[Simpson, Cochran and Throop-69] Simpson, W.E., Cochran, P., Throop, J., "Magnitude estimates of the distance of speech stimuli", *Psychonomic Bulletin*, 1(2), 1969) 1

[Stevens and Newman-36] Stevens,S.S, Newman,E.B., "Localization of actual sources of sound", *American Journal of Psychology*, 48, (1936)

[Teas-62] Teas,D.C., "Lateralization of acoustic transients", *Journal of the Acoustical Society of America*, 34, (1962), 1460.

[Tobias and Schubert-59] Tobias,J.V., Schubert, E.D., "Effective onset duration of auditory stimuli", *Journal of the Acoustical Society of America*, 31, (1959), 1595.

[Tobias and Zerlin-59] Tobias,J.V., Zerlin,S., "Lateralization threshold as a function of stimulus duration", *Journal of the Acoustical Society of America*, 31,(1959), 1591.

[Toole-67] Toole, F.E., "Free-field localization of acoustic images", *Journal of the Acoustical Society of America*, 42, (1967), 1179.

[Thurlow, Marten and Balchandra-65] Thurlow,W.R., Marten, and Balchandra, "Localization of sound from two plane-wave sources", *Journal of the Acoustical Society of America*, 37, (1965) 837

[Veneklasen-70] Veneklasen P.S. "Model techniques in architectural acoustics", *Journal of the Acoustical Society of America*, 47(2), (1970)

[Vermuelen-60] Vermuelen, "Stereo Reverberation", *Journal of the Audio Engineering Society*, 6 (1958)

[Wallach, Newman and Rosezweig-49] Wallach,H., Newman,E.B., Rosenzweig,M.R., "The Precedence Effect in Sound Localization", *American Journal of Pschology*, 62 (1949) 315

[Wallach-39] Wallach,H., "On sound localization", *Journal of the Acoustical Society of America*, 10, (1939), 270

[Wallach-40] Wallach,H., "Role of head movements and vestibular and visual cues in sound localization", *Journal of Experimental Psychology*, 27 (1940), 339

[Walters-70] Walters B.G. "Instrumentation for acoustic modeling", *Journal of the Acoustical Society of America*, 47(2), (1970)

[Warren-58] Warren, R.M., "A basis for judgments of sensory intensity", *American Journal of Psychology*, 71 (1958) 675-687

[Warren-77] Warren, R.M., "Subjective loudness and its physical correlate", *Acustica* 37, (1977) 334-346

[Whitworth and Jeffress-61] Whitworth,R.H., Jeffress,L.A., "Time vs Intensity in the localization of tones", *Journal of the Acoustical Society of America*, 33,(1961) 143

[Young and Cliff-72] Young, F.M. and Cliff, N., "Interactive scaling with individual subjects", *Psychometrika*, 37, (1972) 385

[Zwicker and Scharf-65] Zwicker, E. and Scharf, B.A. "A model of loudness summation", *Psychological Review*, 72(1) (1965) 3

[Zwislocki and Feldman-56] Zwislocki, J., and Feldman, R.S., "Just- noticeable differences in dichotic phase", *Journal of the Acoustical Society of America*, 28, (1956) 860