Center for Computer Research in Music and Acoustics November 1982

Department of Music
Report No. STAN-M-12

AUDITORY DISTANCE PERCEPTION UNDER NATURAL SOUNDING CONDITIONS

Final Report

by

John Chowning and Christopher Sheeline

Research sponsored by

National Science 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-12

AUDITORY DISTANCE PERCEPTION UNDER NATURAL SOUNDING CONDITIONS

Final Report

by

John Chowning and Christopher Sheeline

The primary objective of this project was the development of a practical method for generating perceptual conditions of a realistic and room-like nature, for the purpose of testing the ability of humans to judge the source distance of sound. Most distance studies are executed in laboratories in which control of reverberation and other room cues is negligible, if it exists at all. This research project optimized and extended our ability to bring "natural sounding conditions" into the laboratory.

Although it is primarily through the use of digital signal processing applied to the physical signal that this is possible, our objective has been biased towards a set of perceptual criteria that determine the final success of our simulations. A special purpose digital computer is used to simulate those characteristics of room reverberation that are most salient to listeners, such as the reverberation time and the apparent density of the reverberation. The studies described in this report describe the specific details of our reverberation system and the methods used verify the quality of the rooms simulations.

Sounds were recorded in four different auditoria on the Stanford University campus. These were then modeled on our system, using the same source sounds, and psychological testing was performed to determine perceptual criteria that emerged were the reverberation time of the hall and the apparent "brightness", and the simulations were judged to be very similar to the real rooms.

This research was supported by the National Science Foundation under Contract NSF BNS 79-24646. 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.

Introduction

The original proposal was for funding over a period of two years and included two distinct sets of experiments. The first was concerned with the perceptual modeling of room reverberation. The second involved experiments in distance perception that took advantage of those modeled reverberant spaces. In fact, we were funded only for the first year of the research and the increased emphasis of the first set of studies in the following report reflects that limitation of resources.

Overview: Simulation of natural sounding environments

The greater part of our work to date has been concentrated in the area of modeling real reverberant spaces using computer driven sound systems of two and four audio channels. The primary emphasis has been the creation of reverberation algorithms that can be utilized by our digital signal processor. Our ideal is to create acoustical impressions similar in "naturalness" to the impressions made by recordings of real rooms, using identical source material for the recordings and the digital processing.

Real rooms display features that are analyzable both acoustically and perceptually, and it is important to distinguish the application for one model or type of model from another. It might be possible to model reverberant spaces perceptually without consideration for any but the most elementary acoustical details, but the acoustical dimensions of such a model are not obvious. One facet of our research is directed towards determining the limits of this physical data reduction.

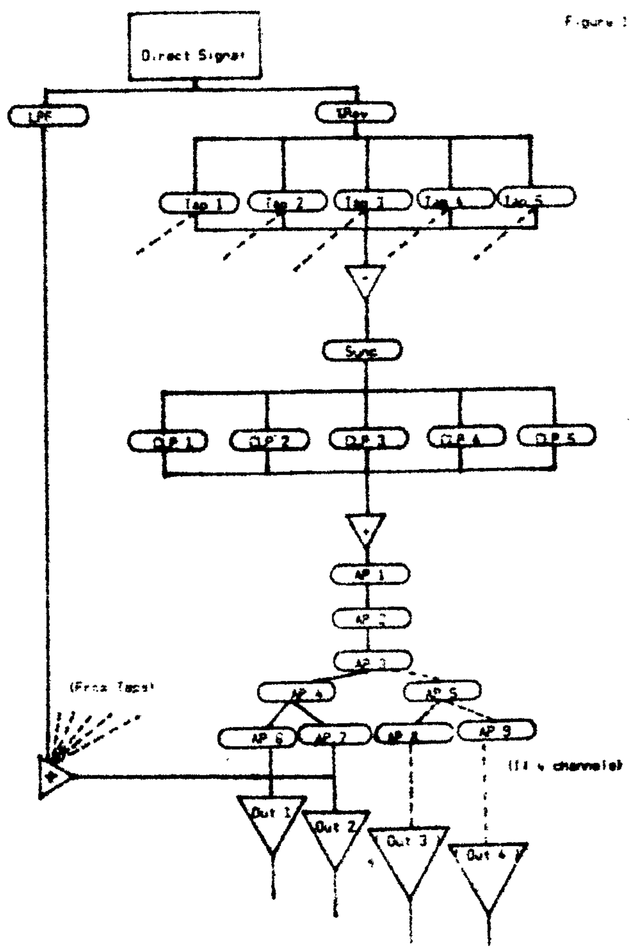
The possibility does exist, however, to achieve a perfect perceptual identity through precise acoustical modeling, using digital convolution of a real impulse response with a sound source. The unusual characteristic of this method is that it can be performed with absolutely no knowledge concerning the acoustical or perceptual nature of the space. The convolution process functions as a "black box" through which the impulse response and the source signal pass, emerging as a signal identical to that which would have been obtained if the original source had been recorded in the space, rather than an impulse.

Our goal has been to find a middle ground that achieves *perceptual similarity* through application of knowledge derived from both physical and perceptual acoustics. Our strategy has been to implement a design incorporating the control of relevant acoustical characteristics such as early reflections, spectral envelope and reverberation time with efficient algorithms that 1) are easily implemented on our signal processor and 2) allow interactive control of the relevant parameters such as filter coefficients and delay lengths, reverberation time and percentage of reverberant to direct signal.

Improved digital reverberation algorithms

The basic reverberation algorithm that we continue to use is described extensively in the original proposal and in Moorer [1979]. Figure 1 illustrates this basic design. It has a number of properties that make it ideal for our purposes. In particular it allows for the definition of features in both the acoustical and perceptual domains.

Figure 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.

Early Reflections

"Tap" delays, in our terminology, are non-recirculating delays used primarily for the simulation of early reflections of the direct signal. The important characteristic of these delays is that they must reflect, in a gross sense, the size of the space being modeled. For example, if we are sampling our sounds at a rate of 25000 samples per second, and the room we are attempting to model has dimensions on the order of 15' x 25' x 40', if we use a reverberator with 10 early reflections ranging in size from 500 samples to 2000 samples, we are implying a range of times varying between 20 msec and 80 msec and commensurate distances between 23' and 90'. These delay times, then, suggest a space somewhat larger than we really intend. The simple solution to this problem is to find a set of tap delays that work well, perceptually, for a given generic shape and volume, and then scale those delay lengths by a value proportional to the desired change in volume.

A second factor relating to tap delays is the distribution of these delays over the early part of the signal. We have tried three approaches and find that their success is as much a function of intuition as of structured analysis. The first method used was that selected by Moorer, the use of measured delay times and amplitudes of early reflections for several well-known concert halls. From those figures, he also made adjustments to suit his needs for smaller reverberators. The second method uses an image model for determining the first N reflections of a space, given the positions of the listener and source, volume, dimensions and a single absorption coefficient for that room. The third method is based primarily on trial and error, beginning with a set of values determined by measurement(or an image model) and optimized perceptually. No single method demonstrates clear superiority, although the third really just represents a modification of the first two, based on perceptual criteria. As such, it is the preferred method.

Direct/Reverberant Signal Ratio

The percentage of direct to reverberant signal is somewhat of a red herring, if viewed as an acoustical dimension, but perceptually it is extremely important. Since physical rooms are stable, the decay of reverberation will proceed according to well known rules and the properties of the space. Under these conditions, the ratio has no meaning. However, in digital reverberation systems, the portion of the signal that is sent to the reverberator is the most general and useful way of controlling the apparent resonance of the modeled hall. It should be noted that this value bears no clear relationship to any measurable physical quantity in real space. In our simulations, however, it does interact with the coefficient that is the primary determinant of reverberation time and consequently bears a perceptual relationship to the apparent size and reverberance of the room.

To determine an appropriate ratio of direct to reverberant signal, one must know the absorbent properties that have been designed into the system. The resultant choice of value will relate to the resonance of the hall, the apparent distance of the source from the listener, and the reverberation time. Values ranging between 1 percent and 100 percent, over a range of distances, are not uncommon.

Room-dependant Spectral Resonances

Another feature of our reverberators is their property of modeling, roughly, the spectral resonances characteristic of most rooms, as well as the low-pass effects of molecular absorption in the air and the absorptive qualities of most construction and furnishing materials. Our primary concern in modeling is matching the room resonances at a perceptual level, and this can usually be accomplished by the adjustment of a small number of filter parameters. Although Moorer suggests that a minimum of six low pass reverberation units (LPRs in Fig 1) are required, we have found that five are adequate if the parameters are optimized.

Binaural Decorrelation

Lastly, our design makes it possible to take a single source signal and reverberate it, presenting an individually delayed and attenuated version of the final reverberation to each loudspeaker channel. In effect this decorrelates the signal binaurally in time, introducing a desirable asymmetry for two and four channel listening. This ability to distribute unique outputs for each channel also permits us to distribute the outputs from the tap delays in an asynchronous and "roomlike" manner.

Impulse response modeling for dimensional analysis

Our original plan was to record impulse responses in a number of different halls in the Stanford area. We intended to use those recordings in two ways. The first was to perform a dimensional analysis of the impulse response, extracting precise acoustical measurements of reverberation time, dynamic spectral response, and early reflections. We also hoped to have an impulse generator of sufficient brevity, such as a spark gap, to allow convolutions of the digitized impulse response with the same unreverberated sounds we were using to play through the reverberator models. Several problems occurred that made it virtually impossible to achieve this end.

We were unable to locate an appropriate instrument for generating a true impulse of sufficient energy to allow the kind of dynamic range we were hoping to achieve. After some searching, we decided that the best alternative would be to generate an impulse on the computer, record that impulse and use it, played through a loudspeaker, as our source signal. The net result of our efforts was that we were unable to generate an impulsive sound with adequate energy for our purposes.

As a consequence of these problems, we decided to postpone the acoustic dimensional analysis of reverberant spaces until a more robust method of room excitation was fully developed and we had the availability of a portable digital recording medium. The technique envisioned for measuring the impulse response of the reverberant space is to excite the room with 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 impulse source, it is possible to inject more energy into the room thereby providing an improved signal to noise ratio. An implementation of this method is currently under developedment at CCRMA.

Physical/psychophysical properties of reverberation based on stochastic and deterministic criteria

Another set of experiments currently underway is focused on the modeling of reverberation based purely on stochastic and deterministic statistical rules. Impulse responses are generated according to either deterministic or stochastic temporal rules and deterministic or stochastic rules for amplitude decay. These are then convolved with musical instrument tones and auditioned for naturalness. We are in the first stages of this work and results appear promising, although it is too early to predict what particular modifications to these rules will be necessary.

Stereo recordings of musical sounds in reverberant spaces

In order to continue with the studies in distance perception without pursuing the originally intended dimensional analysis of real rooms, a method of demonstrating the naturalness of our current models was devised. The focus of this experiment was the demonstration of perceptual similarity of experience between listening to our models and listening in real rooms.

A number of trumpet sounds, covering the range of the instrument, were recorded in a "dead" 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. These recordings were then digitized and edited on the computer. In addition, the original unreverberated recording was digitized.

Perceptual modeling by method of AB comparison

The stereo recordings were then used individually in A-B comparison with models, while we attempted to create a similar sounding model for each recorded room. This involved altering most of the parameters discussed at the beginning of this report, but primarily the reverberation time, filter coefficients and tap delay lengths. Tape noise and some air conditioning noise which was preserved through the digitization process was also modeled. This solution was preferred to high and low pass filtering, as we were uncertain how the low frequency rumble of the air conditioning might interact acoustically with the signals, and we were unwilling 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 degree of spectral attenuation of the direct signal in the originals ranged between apparent perceptual thresholds of 7 kHz and 9 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 trial and error, separately for each recorded example. The digitized recording was played, a basic version of the model was played after it, adjusted, played again, and so on until we were satisfied with the natural quality and perceptual similarity between the model and the original. 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, the models were programmed to resemble the real rooms as closely as possible.

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. The study required the listener to evaluate the apparent dissimilarity between pairs of rooms from the total of nine. It was emphasized that they were to attend, if possible, to the room and its reverberance and interaction with the direct sound. This was an attempt to steer them away from artifacts of the recording and digitizing process, such as the tape hiss and uncorrelated room noise that was described earlier. In addition the subjects were questioned regarding their conscious ability to distinguish differences between the various stimuli. A large number of trials was performed, with six repeated measures for each pair per subject. These results were tabulated and a multidimensional scaling analysis was performed, providing analyzable results in both two and three dimensions. Individual differences among subjects were also noted for the two dimensional solutions.

Two dimensional solution

The group solution 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.

Three dimensional solution

In the three dimensional solution, 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.

General conclusions concerning this study

1) The most salient characteristic for all listeners, when asked to differentiate among listening spaces, is that of 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 direct to reverberant signal energy. The relationship becomes

clear when we note that the product of this ratio and the reverberation time is approximately one (1.0) in all of the halls we modeled. This suggests a strong negative correlation between the two.

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 based in large part on secondary cues.

Distance perception in the median sagittal plane

We are currently completing a set of experiments that investigate the salience and interaction of the major cues to auditory distance perception. Each experimental condition requires the listener to vary a single parameter relating to disance cue, such as the amplitude of the direct signal, the ratio of direct to reverberant signal, or the coefficients controlling the low-pass filtering of the direct signal. Several "generic" reverberator models, based on those used in the earlier modeling studies, are being used. These are characterized by short, medium, and long reverberation times (.25 - 2.5 sec), and by differing spectral envelopes. The results of these studies will be published in Sheeline [1982].

References

[Federkow 1978] Federkow, G. et al., "Computer-controlled sound distribution system for the performance of electroacoustic music", Computer Music Journal, 2(3) p33

[Carterette and Friedman 1978] Handbook of Perception, Volume 4:Hearing, Carterette E.C., and Friedman, M.P., (Eds.), Academic Press, New York (1978)

[Chowning 1971], Chowning, J.M., "The Simulation of Moving Sound Sources," Journal of the Audio Engineering Society, 19,(1971), pp 2-6; also in Computer Music Journal 1,3 (1977) pp48--52 and pp. XX--XX in this book.

[Durlach and Colburn 1978] Durlach, N.I., and Colburn, H.S., "Binaural Phenomena", in E.C.Carterette and M.P. Friedman (Eds.) Handbook of Perception, Volume 4:Hearing, Academic Press, New York 1978

[Fletcher and Munson 1933] Fletcher, H.F., and Munson, W.A., "Loudness, its definition, measurement, and calculation", Journal of the Acoustical Society of America, 5,(1933), pp82-108

[Gardner 1968] Gardner, M.B., "Historical background of Haas and/or precedence effect", Journal of the Acoustical Society of America, 43, 1968 p1243

[Gottlob 1975] Gottlob, D. "Comparisons of objective acoustic parameters in concert halls with results of subjective experiments", Ph.D thesis, Library Translation 1895, German Government Publication by the "Building Research Establishment", Department of the Environment (1975)

[Holt and Thurlow 1969] 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), p1584

[Marks 1974] Marks, L.E., "On scales of sensation: Prolegomena to any future psychophysics that will be able to come forth as science", Perception and Psychophysics, 4,(1968), pp358-376 [Mills 1970] Mills, A.W., "Auditory Localization" in J.V. Tobias (Ed.) Foundations of Modern Auditory Theory, Volume 2, Academic Press, New York 1970

[Moorer 1979] Moorer, James A. "About this reverberation business", Computer Music Journal, 3,2 (1979)

[Reynolds 1977] Reynolds, R., "Explorations in sound/space manipulation", available from the Center for Music Experiment, Dept. of Music, University of California in San Diego.

[Robinson 1957] Robinson, D.W., "The subjective loudness scale", Acustica,7,(1957), pp217-233

[Robinson and Dadson 1958] Robinson, D.W., and Dadson, R.S., "A re-determination of the equal-loudness relations for pure tones", British Journal of Applied Physics, 7,(1956), pp166-181

[Scharf 1978] Scharf, B., "Loudness", in E.C. Carterette and M.P. Friedman (Eds.) Handbook of Perception, Volume 4:Hearing, Academic Press, New York, 1978

[Schroeder 1962] Schroeder, M.R., "Natural sounding artificial reverberation", Journal of the Audio Engineering Society, 10, (1962) p219

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

[Schroeder 1973] Schroeder, M.R., "Computer models for concert hall acoustics" American Journal of Physics, 41, (1973) p461

[Searle 1976] Searle, C.L. et al. "Model for Auditory Localization", Journal of the Acoustical Society of America, 60, (1976, p1164

[Sheeline 1982], Sheeline, C.W., "Auditory Distance Perception Under Natural Sounding Conditions", PhD Dissertation in progress. Expected date of completion, 10/82

[Stevens 1955] Stevens, S.S., "The measurement of loudness", Journal of the Acoustical Society of America, 27,(1955), p815

[Stevens 1959] Stevens, S.S., "On the validity of the loudness scale", Journal of the Acoustical Society of America, 31,(1959),pp995-1003

[Tobias 1970] Foundations of Modern Auditory Theory, Jerry V. Tobias (Ed.), Academic Press, New York 1970

[Warren 1977] Warren, R.M., "Subjective loudness and it's physical correlate", Acustica, 33,(1977)

[Zwicker, Flottorp, and Stevens 1957] Zwicker, E., Flottorp, G., and Stevens, S.S., "Critical bandwidth in loudness summation", Journal of the Acoustical Society of America, 35,(1957), pp548-557

[Zwislocki 1978] Zwislocki, J.J., "Masking: Experimental and theoretical", in E.C. Carterette and M.P. Friedman (Eds.) Handbook of Perception, Volume 4:Hearing, Academic Press, New York 1978