Center for Computer Research in Music and Acoustics June 1974

Department of Music
Report No. STAN-M-1

COMPUTER SIMULATION OF MUSIC INSTRUMENT TONES IN REVERBERANT ENVIRONMENTS

by

John M. Chowning
John M. Grey
James A. Moorer
Loren Rush

CCRMA
Artificial Intelligence Laboratory
DEPARTMENT OF MUSIC
Stanford University

Center for Computer Research in Music and Acoustics June 1974

Department of Music
Report No. STAN-M-1

COMPUTER SIMULATION OF MUSIC INSTRUMENT TONES IN REVERBERANT ENVIRONMENTS

by

John M. Chowning
John M. Grey
James A. Moorer
Loren Rush

ABSTRACT

This is a reprint of selected portions of the NSF proposal which resulted in a grant to the computer music group for research over a two-year period.

Novel and powerful computer simulation techniques have been developed which produce realistic music instrument tones that can be dynamically moved to arbitrary positions within a simulated reverberant space of arbitrary size by means of computer control of four loudspeakers. Research support for the simulation of complex auditory signals and environments will allow the further development and application of computer techniques for digital signal processing, graphics, and computer based subjective scaling, toward the analysis, data reduction, and synthesis of music instrument tones and reverberant spaces. Main areas of inquiry are: 1) those physical characteristics of a tone which have perceptual significance, 2) the simplest data base for perceptual representation of a tone, 3) the effect of reverberation and location on the perception of a tone, and 4) optimum artificial reverberation techniques and position and number of loudspeakers for producing a full illusion of azimuth, distance, and altitude.

COMPUTER SIMULATION OF MUSIC INSTRUMENT TONES IN REVERBERANT SPACES

COMPUTER SIMULATION OF MUSIC INSTRUMENT TONES IN REVERBERANT SPACES

Leland C. Smith & John M. Chowning, Department of Music
John M. Grey, Psychology - James A. Moorer, Computer Science - Loren Rush, Music
Program Coordinator - Albert Cohen, Chairman of the Department of Music
Consultant - John R. Pierce, California Institute of Technology

Faculty Advisory Group
James B. Angell, Department of Electrical Engineering
John McCarthy, Department of Computer Science
Earl D. Schubert, Hearing & Speech Sciences, School of Medicine
Roger N. Shepard, Department of Psychology
Head of External Advisory Group - Max V. Mathews, Bell Telephone Laboratories
Stanford University, June 1974

ABSTRACT

Novel and powerful computer simulation techniques have been developed which produce realistic music instrument tones that can be dynamically moved to arbitrary positions within a simulated reverberant space of arbitrary size by means of computer control of four loudspeakers. Research support for the simulation of complex auditory signals and environments will allow the further development and application of computer techniques for digital signal processing, graphics, and computer based subjective scaling, toward the analysis, data reduction, and synthesis of music instrument tones and reverberant spaces. Main areas of inquiry are: 1) those physical characteristics of a tone which have perceptual significance, 2) the simplest data base for perceptual representation of a tone, 3) the effect of reverberation and location on the perception of a tone, and 4) optimum artificial reverberation techniques and position and number of loudspeakers for producing a full illusion of azimuth, distance, and altitude. These areas have been scantily investigated, if at all, and they bear on a larger more profound problem of intense cross-disciplinary interest: the cognitive processing and organization of auditory stimuli. The advanced state of computer technology now makes possible the realization of a small computer system for the purpose of real-time simulation. The proposed research includes the specification of, and program development for, a small special purpose computing system for real-time, interactive acoustical signal processing. The research in simulation and system development has significant applications in a variety of areas including psychology, education, architectural acoustics, audio engineering, and music.

READER'S GUIDE

We have included this guide to describe the contents of this proposal and to point the reader to those sections which might be most appropriate for his review. We have broken our presentation into several levels of detail concerning our current and proposed research. A general description of the nature of our work appears in the following section I, Introduction.

A considerably more detailed account of our research is presented in section II, Research Proposal, which itself is broken into two levels of detail (see note to the reader in section II). Here, we would like to point out the two main areas of interest in the Research Proposal. The first of these is is in engineering, and more specifically in digital signal processing. There are three topics in the proposal which relate to this field. The first topic is concerned with frequency modulation, a concept from radio broadcasting technology, is used in a new and unusual manner to synthesize musical tones. Unlike broadcasting, the waveform is never demodulated. We use the modulation as a method of generating a controlled spectrum, harmonic or inharmonic, which can change with time. Section II-A-2 discusses frequency modulation synthesis in great detail. It should be mentioned that this discovery has far-reaching implications for music synthesis and psychoacoustics. The frequency modulation method offers an extremely efficient method of producing a very rich range of timbres. The second topic related to engineering is the heterodyne filter. Appendix A derives this analysis technique which is useful in analysing harmonic spectra such as are exhibited by many music instruments. The last topic in engineering is in artificial reverberation by digital methods. Section II-B-1 introduces digital reverberators, including a new form of colorless reverberation. The derivation for the new form as well as the old form is given in Appendix D. The digital reverberators are derived from analog all-pass transfer functions by the techniques of digital signal processing.

The second main area of interest in the Research Proposal is that of psychoacoustics. Although not the main thrust of the proposed research, psychoacoustics is a componant of every phase of our research. The topics which would be of greatest concern in this area include the perception of complex tones in our development of techniques for the synthesis of musical timbres, and the coordination of localization cues in our simulation of moving sound sources using four speaker-channels. One aspect of our work bears on the perceptually salient aspects of music instrument tones, and deals with the systematic simplification of the analyzed instrument waveform in order to determine which features of the initially complex wave are important in its perception. For example, our success in synthesizing many instrument tones on the basis of a tremendously simplified representation of the waveform is a significant step towards finding which aspects of the tones are perceptually important. The section on Current Research in part II-A-1, Additive Synthesis Based on the Analysis of Real Tones, goes into this line of research. A second, related, topic dealing with the perceptual cues for natural tones has evolved from our development of the frequency modulation synthesis technique. Here, the time-varying parameter of the bandwidth of the signal has been found to be a most important cue for the simulation of many instrumental tones. This line of research is detailed in the section on Current Research in part II-A-2, Frequency Modulation Synthesis. The Proposed Research sections of parts II-A-1 and II-A-2 are also of interest to the psychoacoustician, as is much of the material in the next part, II-A-3, Towards a General Model for Simulation, which deals with interactions between the above lines of research in the pursuit of a more general model. Under Proposed Research in part II-A-3 are two topics of particular psychoacoustic significance, the application of multidimensional scaling techniques to the study of timbre perception, and the investigation of categorical perception for timbre. The second main topic of interest to the psychoacoustician is our development of techniques for the simulation of localized and moving sound sources using four speaker-channels. The Current Research section of part II-B-2 discuss the technique, based on the coordination of several localization cues. We point out research that we would like to pursue with respect to these cues under the Proposed Research section, and this should be of interest to the psychoacoustician. A related topic of research deals with the perceptual scaling of our digital reverberation techniques and the simulation of real rooms, appearing in the Proposed Research section of part II-B-1, Simulation of Reverberant Spaces.

I. INTRODUCTION

The program of research presented here has as its ultimate goal the production of acoustical waves by means of computer control over loudspeakers which can provide for a listener the impression of any music instrument tone at any location within any reverberant space. Between the user who specifies the tone in a space and the listener who perceives it, there must be a small but powerful real-time computing system, a small and highly optimized data base, and efficient algorithms which compute the control signals for the loudspeakers on the basis of the physical correlates to the perceptual cues of the tone and space.

The computer production of acoustical waves which contain the cues required for the *perception* of a music instrument tone in a reverberant space is a problem that is fundamentally different from that of producing waves which contain all of the information of an original instrument source in a real room. The first we define to be a problem of simulation, or the process of providing the perceptual impression of the natural source, whereas the second is that of reproduction, or the process of achieving an exact or close imitation of the natural waves. While it is true that an accurate reproduction of the original acoustical waves will necessarily contain the perceptual cues of an instrument tone and the space, it is not necessarily true that a simulation of a tone will contain all of the information in the original waves. We draw attention to this distinction between reproduction and simulation, because what can be learned from the two processes is significantly different. The broadcast, tele-communications, and recording industries have for the most part solved the problems of reproduction and the accumulated knowledge is vast, having to do with bandwidth, signal to noise ratios, equalization, and encoding. An equivalent research effort in simulation has only just begun, where the goal is the production of those features *alone* of a complex wave to which the human perceptual mechanisms respond (Risset & Mathews, 1969). Research in the computer simulation of complex auditory signals will produce knowledge in the general area of perceptual representation by isolating those physical features of complex signals which are required to give the appearance of naturalness.

It has become clear that for the purpose of simulation, digital computers provide the most effective control of loudspeakers. The loudspeaker is a device of extraordinary richness and potential in that it can be used to reproduce nearly any perceivable sound, perhaps not perfectly, but certainly with more than adequate fidelity. The computer is programmed to generate a sequence of numbers or samples, which are a numerical representation of the instantaneous amplitude of a desired waveform. The accuracy of the representation increases as the time interval between successive samples decreases and as the numerical precision of each sample increases. The samples are passed in sequence to digital to analog converters, whose voltage outputs are amplified and applied to loudspeakers. The precision and flexibility of this method is very great and has allowed the development of analysis and synthesis techniques which are uniquely suitable to digital processing.

A. SUMMARY OF CURRENT AND PROPOSED RESEARCH

CURRENT RESEARCH

At the Stanford Artificial Intelligence Laboratory, pilot programs and techniques have been developed for the analysis and synthesis of complex signals and for the simulation of moving sound sources in reverberant spaces, some of which are startlingly simple in implementation and novel in conception.

Analytical programs have been written which digitize the acoustical wave of a music instrument by means of an analog to digital converter. The program then reduces the data to time-variant frequency and amplitude functions and the wave is reformed from these functions through additive synthesis and played through the digital to analog converter in order to determine the goodness of fit of the analysis. Further reductions are made to the data with the aim of discovering the optimal physical representation of the tone with regard to its perceptual features. These procedures have been successfully applied to tones from several instruments, including the violin, one of the most complex of all instruments, where the data reduction ratio is *greater than 250 to 1!* without any disturbance of the perceptual images of these tones. Extension of this research to a wide base of signals will have major implications for auditory theory which is concerned with the perceived features of complex sounds, and will further bear rich rewards in the approach to the most efficient computer-based simulation algorithms for natural tones. Our development of a novel algorithm for the interpolation in the multidimensional physical timbre space between two so simplified tones presents tools for the psychological evaluation of the perception of timbre, for example, the possible existence of categorical perception.

An altogether new technique for generating complex acoustical waves using the computer was discovered in our lab several years ago. The technique is based upon a special application of simple frequency modulation, where with two parameters and two time domain functions a large number of highly differentiated tones can be produced which have a strong resemblance to natural instrument tones. The technique does not have the generality of the additive synthesis mentioned above, however the simplicity of control has provocative implications. For many tones the data reduction ratio is a factor of ten greater than the ratio for additive synthesis. It is to a large degree the extraordinarily simple physical correlations to the perceptually complex images resulting from this technique that has generated a far-reaching research interest in the significant perceptual cues for such images. Since its initial discovery here, this technique has received wide interest from researchers in computer simulation techniques for music instrument tones, and also from auditory theorists and perceptual psychologists. Our ongoing development of this technique should provide many new insights into the simulation and perception of complex auditory signals.

In the simulation of natural tones in natural environments using loudspeakers, it is of utmost importance that the realism of the auditory images supercedes the physical presence of the loudspeakers themselves. In order to free sounds from the loudspeakers it is necessary to simulate the reverberation of a space as well as the localization cues of the sound within the space. To this end, artificial reverberation techniques based on Schroeder (1962), together with the results of several years of research into producing localization cues by means of loudspeakers, have been implemented in a general control program for the arbitrary localization of a source. Using interactive graphic display techniques, a user can specify the location an movement (trajectory) of a sound in a two-dimensional reverberant space. A program computes the control functions for azimuth, distance, and velocity which are used to modulate the signal to be applied to the loudspeakers. This program has been useful in the investigation of simulation algorithms for localization cues and for indicating the most potentially productive research areas for the future. Additional work which extends that of Schroeder on colorless reverberation has lead to the uncovering of a novel, higher-order reverberator that may be useful in extending the range of possible simulated environments. Our recent research has been very concerned with the maximal preservation of the characteristics of the input signal, hence our interest in and further development of colorless reverberation schemes.

All of the research to date has been done at the Stanford Artificial Intelligence Laboratory. It should be noted that we have received *no direct support for research in the form of salaries or purchase of hardware,* although we have been allowed to use the computer facilities through the generosity of its directors. All of the audio recording, amplification, and noise reduction equipment was purchased with income resulting from a license of the "Simulation of Moving Sound Source" invention described in Chowning (1971). In recent years, as the use of the A.I. facilities has increased by a significant amount, and consequently, the presence of any research group has been a much more apparent load on the system, the presence of a non-funded project has become an increasing burden on the resources of the lab. It has become clear, therefore, that we must seek external support in order to continue our association with the lab, by 1) significantly reducing our computing load and 2) paying, in part, for our use of peripheral equipment. Recent advances in computer technology, resulting in the availability of specialized hardware for real-time signal processing, make possible our development of a small satellite system which would not only significantly reduce our load on the system, but would immeasurably increase the rate of our research progress and would allow for many diverse and unforeseen real-time applications, some of which are listed below. The research and development of such a system is a major impetus for our seeking external support.

Stanford University has recognized the contributions of the research group by creating the Center for Computer Research in Music and Acoustics. This proposal is directed toward the Center's initial support.

PROPOSED RESEARCH

The results from the current and past research indicate clearly both the general direction and some of the specific steps for the future. The overall goal in the simulation of instrument tones is the development of synthesis algorithms which produce tones having the perceptual complexity and naturalness of those in the real world, but which also have the simplest possible physical representation in the computer. In order to achieve this goal, the analysis and two synthesis techniques will be applied to a larger set of tones with a view to capturing or confirming the significant perceptual features through the rigorous application of reduction techniques. As the empirical data accumulates, the reduction techniques will be 'formalized' as algorithms which are able to detect and preserve the perceptual features of a tone in the most concise representation. Similarly, algorithms for mapping the perceptual attributes of a tone into parameters and functions for FM synthesis will be developed. A convergence of the two synthesis techniques is anticipated in that the FM synthesis of many complex instrument tones requires a particular expansion of the technique which places it in part in the category of additive synthesis. Finally, the perceptual representations of tones will be used to formulate higher order algorithms which reflect a general model for the perception of a wide range of natural tones. Methods from experimental psychology will be used to help establish the dimensionality of this perceptual model and the relationships between the subjective dimensions and the physical properties of the tones.

In the case of the simulation of reverberant spaces and the localization of sources within the space, the overall goal is to be able to provide for a user the maximum control over localization of sources and over size, shape, and reverberant qualities of the apparent space. The major research to be done in achieving natural representation of real rooms is in the artificial reverberation algorithms. We plan to develop the techniques of Schroeder in conjunction with other techniques developed here, through the use of graphic computer analysis programs. The difficulty in constructing compound reverberation circuits is that there is no current method for formal prediction of their output (which in addition is very often counter-intuitive). The combination of such programs, together with subjective evaluation, appears to be the most effective manner of research. With the application of resonators to uncolored synthesized reverberation, we plan to simulate a number of real acoustical environments.

The research in localization will focus on the optimum number and arrangement of independently controlled loudspeakers which maximize the effective area of listening positions. Although the algorithms which we have developed for localization appear to be effective for four channels, we plan to further "tune" and evaluate the cues through subjective measurements for as many as eight channels.

The effectiveness of the research effort will be dependent upon the proposed development of a special purpose, interactive, acoustical signal processing system. The system will be able to synthesize in real-time a number of complex signals, localized in complex reverberant environments. Programs will be developed which will give the user a high level control over the total acoustical environment. The programs will include all of the digital simulation techniques which have been developed and which are proposed. The integrated circuit technologies, especially the rapidly developing field of large scale integration, suggest provocative applications of this research in digital simulation.

B. APPLICATIONS

In realizing the research proposed here, a tool of great complexity and power will be assembled. This tool will be specially designed for the production of sound. This tool is of sufficient generality that its applications extend to many fields. The hardware portion of this tool will be composed almost entirely of standard off-the-shelf equipment, making replication straightforward. One must remember that what is meant by "tool" here includes the programming as well as the computing equipment. This is why we call it a "system" and not just a machine. The computer is of little use without its software, and well-designed sophisticated programs are necessary to produce a complete system.

The research and development of powerful simulation algorithms that we propose will result in many contributions to related scientific areas, both of a theoretical and practical nature. The most closely related scientific area which will benefit from our research is that of psychoacoustics, the study of the auditory system. Psychoacoustics is now beginning to study the perception of auditory signals which resemble, in complexity and meaningfulness, sounds from our daily environments. The largest body of research in this domain has included the perception of speech signals, an interest which has many obvious payoffs. Little work has been done on the perception of the non-verbal signals, which constitute a sizable remainder of our complex, natural auditory environment. The few efforts which have been made have attempted to look at the perception of music instrument tones, a logical starting point for such investigation. We feel that the research which we will undertake will make many direct contributions to this growing area of knowledge, an area which will provide an ultimate test for any general model of auditory perception. Indeed, implications should result from our work for models of speech perception, in that we are dealing with a comparably complex auditory domain.

Another contribution of our proposed research to the study of hearing is in our development of a real-time digital system for the synthesis of sound. It would be conservative to note that over half of the time spent in research on auditory perception is consumed by the construction and debugging of special-purpose analog circuits. The problems with the stability of, and precision of control over, analog hardware has imposed implicit limitations on the complexity of the auditory stimuli which can serve as tools in research. It is for this reason that the study of the perception of complex, natural signals is out of the reach of most researchers; with digital synthesis of sound it becomes possible, but with the development of a real-time digital system for synthesis it becomes practical. We feel that both the research and development of a real-time system, which indeed supports interactive psychoacoustical experimentation, and the model which our research will provide to this branch of science, will represent a significant step in research possibilities to the scientific community. For example, in auditory research, special-purpose, low budget systems could evolve from our work which certainly would represent as much of a jump from the current limitations of equipment as was the jump from the use of tuning forks and Helmholtz resonators to the use of electronic oscillators and filters.

The algorithms which we are developing for the simulation of music instrument tones localized in reverberant spaces are clearly applicable to the study of higher-order auditory information processing. An extension of the range of auditory signals which fall in the domain of our simulation techniques would include many diverse naturalistic sounds that are not traditionally categorized with music instruments, but which occur in our everyday environments, such as various types of noise, mechanical sounds, and any of the many other sorts of non-verbal sounds which daily surround us. Techniques already implemented for the simulation of localized sources of sound in reverberant spaces would of course apply to the extended set of auditory signals. At this point, very powerful tools would exist for any social science research which desires to examine the behavior of man in a naturalistic environment, but further demands the control over environmental factors. The mounting interest in this level of work is demonstrated by the increasing use of the criterion of the relevance of research findings to real-life situations. The social psychologist could study the influences of various auditory conditions on human behavior, such as levels and types of noise. The study of the internal representations of naturalistic auditory signals and the cognitive operations which may be performed on these representations could make many uses of the tools which we are developing. Both short and long term memory for non-verbal, but familiar, auditory stimuli could also be investigated with these simulation techniques. They would also make feasible the controlled investigation of the effects of training and experience on the processing of, or the effects of the contexts which surround, natural signals. These are but a few examples of possible applications for the simulation algorithms and real-time system which we propose to develop.

It is interesting that the equipment and programs for the storage and poduction of sound have already found practical usage here at the Artificial Intelligence Laboratory. Dr. Kenneth Colby and his co-workers under an NIMH grant have been persuing aids to nonspeaking autistic children. In one of their programs, the computer is used as a teaching device. The child sits or stands before a graphic display device in the music room, which is equipped with a four-channel audio system, driven by the computer's digital-to-analog converter. The audio system was set up by the music department for the Center personel. As the child presses a key, a cartoon appears on the screen and sounds familiar to the child are reproduced through the audio system. The sounds have been digitized and stored on the computer's bulk storage and can be accessed randomly and played by the computer. This use of the audio system is a simple but important example of the application possibilities of the equipment. Complete details of the usage may be found in Colby [1968,1971,1973] and Smith [1972].

There is, perhaps, a more profound relationship of the research we propose to contemporary composition. Although psychoacoustics is not the principle thrust of our research, it is a major component of every aspect of our work in the development of simulation algorithms. The following quote seems particularly appropriate.

> *A few years ago, a colleague and I were asked to write a chapter on acoustics for a book on contemporary music. We assembled what seemed to us pertinent current and unexploited information. The editor rejected the chapter on the grounds that we had not related the material to examples of contemporary music. The only response we could make was that no relation was possible. Contemporary music and psychoacoustics had become completely disjoint fields.* - *J.R. Pierce*

II. RESEARCH PROPOSAL

note to the reader

In the following section we present the nature of our current and proposed research. Because of the scope of our research, this presentation is necessarily lengthy. To aid the reader in obtaining an overview of this material, we have adopted the following convention in this section. Introductory and summary materials, including references to pertinent Figures are presented at the head of all divisions of this section. More *detailed presentations* follow immediately as subsections which have *headings in lower-case italicized type (as here printed).* The reader who first desires to get an overview of our work, before becoming involved with the *details*, can use this format as a guide.

A. SIMULATION OF MUSIC INSTRUMENT TONES

In this part of the proposal we will discuss our approaches to the computer simulation of music instrument tones. The main goal of our research is the development of a powerful, general purpose technique for the simulation of auditory signals that will have the perceptual complexity and naturalness of the musical sounds which occur in the real world. The fundamental concern here is with the synthesis of natural timbres of the extremely varied and highly complex tones which occur in music. The definition of timbre most accepted in the literature of auditory theory is that stated by the American Standards Association (1960): "Timbre is that attribute of auditory sensation in terms of which a listener can judge that two sounds similarly presented and having the same loudness and pitch are dissimilar." It is added that: "Timbre depends primarily upon the spectrum of the stimulus, but it also depends upon the waveform, the sound pressure, the frequency location of the spectrum, and the temporal characteristics of the stimulus."

The immediate problem we face in designing a successful algorithm for the computer simulation of music instrument tones involves the exact nature of the psychophysical relationships in timbre, that is, the relationships between the physical properties of sounds and the subjective, psychological qualities by which they are perceptually differentiated when presented at the same loudness and pitch. We have found that there has been relatively little work done in the last century of auditory research investigating the psychophysical relationships in timbre perception. Even of that which has been done, most of the researchers have held such restrictive definitions of timbre that their findings are entirely useless for application in our present attempt to design simulation algorithms. This should be evident if we examine the real lack of clarity shown in the definition given in the last paragraph, where timbre is most precisely pinpointed by a negative formulation: whatever is not pitch and not loudness (and, we might add, not duration and not location) is 'timbre'. Of the extremely complex acoustical left-overs, researchers have mainly focussed on the (so-called) 'steady-state' spectrum of stimuli as the dominant, if not exclusive, factor in timbre.

It is becoming clear that this is only one of many factors in timbre - if indeed there even is an actual 'steady-state' in real tones, that is, a duration of stability in which the amplitudes of the harmonic components of a tone remain constant. From the results of recent analyses of real tones, as well as attempts to synthesize natural-sounding tones, it appears that real tones have complex and ever-changing physical properties, and that the nature of these temporal changes is most probably an extremely important factor in timbre (Luce, 1963; Risset, 1966; Strong & Clark, 1967a, 1967b; Freedman, 1967, 1968). The vagueness of the above-cited definition seems to accurately represent the actual state of knowledge to date concerning timbre perception, and to ask for a more precise and useful statement demands fundamental research which is yet to be done. We are now in the position to accomplish such research, given the possibilities presented by the digital computer.

Supporting our goal to develop a technique for the computer simulation of natural tones, therefore, is a concurrent research effort to formulate a model which is able to describe the perception of musical sounds. The theoretical problem is to establish a set of acoustical dimensions, those which are actually salient in the perception of musical timbre, and to design a computational algorithm that enables the user to exert the most direct control with respect to these dimensions for the purposes of simulation. The empirical problem which follows is the determination of those aspects of the signal which are actually important in the perception and identification of a sound. Necessarily included is a study of the distinctive features of signals, the investigation of physical conditions which contribute to the naturalness of a signal. Related research should examine the general characteristics of timbre perception, looking into the effects of such phenomena as the categorical identification of musical sounds.

The discussion which follows is concerned with a description of our systematic approach to a general model for simulation, which is based on perceptual verification at every step. The two strategies which we have used for simulation are additive synthesis which is based on the analysis of real tones, and frequency-modulation synthesis. The first method discussed below, additive synthesis based on analysis, presents the goal for research of data reduction. We start with the most complete, complex information about a real signal given through its analysis. We then systematically step in the direction of the most simple representation of the signal which can be used successfully to reproduce the original tone by additive synthesis. To this end we examine the perceptually important aspects of physical signals.

The second approach towards a model for simulation discussed below begins with a simpler, more easily-controlled process, frequency modulation synthesis of sound. This technique allows the user to directly manipulate aspects of the signal that we subsequently found to be very meaningful in terms of certain perceptual cues for music instrument tones. The success of this method first came as a surprise because the physical waveform that it generates is strikingly different from that of any natural signal. However, upon inspection, the reasons for this success have been determined, and we thereby began to learn what physical dimensions are perceptually important for timbre. The direction of this approach is to increase the complexity of the synthesis process, until there is control over a very wide set of features which occur in natural tones.

Following the detailed description of these two methods is a third section devoted to a discussion of the ultimate aim of our research: the development of powerful, general-purpose algorithms for the computer simulation of natural tones. This more general algorithm will be an outgrowth of the interdependency and convergence of our two approaches to synthesis. The two approaches do not proceed independently of one another, but interact at several levels. Findings in one technique can immediately be applied to the other, and a system for cross-verification is thereby established. In this way, a convergence of these two methods is approached. The simplicity and perceptual meaningfulness of specifications to the frequency modulation technique points out an important goal for the additive synthesis method. On the other hand, the complexities of tone which are revealed by analysis, and which are confirmed to be perceptually salient in the additive synthesis, point out necessary levels of complexity which must be accomodated by the frequency modulation technique. As the latter technique is then made more complex, it in fact enters the category of additive synthesis. The ultimate model for simulation will draw from the research findings using both methods.

A common aspect of research with both methods is the concern for perceptual verification of any particular results at hand. Experimental methods from perceptual psychology are employed for the rigorous verification of the success of simulation, in terms of the discriminability of a synthesized from real tones and in terms of the naturalness of simulation. In addition, to assist in the development of a general algorithm, we will have to formulate a general model for the perception of timbre. This general model will provide information for the construction of perceptually-based higher-order simulation algorithms. We employ a spatial model for the subjective structure of the perceptual relationships between signals. Research is directed at uncovering the dimensionality of the subjective space, the psychophysical relationships which are structurally correlated to this space, and the properties of the space. The existence of such constraints as categorical boundaries will be investigated in an attempt to assess the continuity of the subjective space for timbre. In the same regard, we will also examine the effects of musical training or context on the structure of the space. The model will be evaluated by our ability to predict the mappings of real and novel tones.

I. ADDITIVE SYNTHESIS BASED ON THE ANALYSIS OF REAL TONES

INTRODUCTION TO SYNTHESIS AND ANALYSIS TECHNIQUES

This section will introduce our approach to the simulation of music instrument tones using additive synthesis based on the analysis of tones from actual instruments. Additive synthesis considers a complex sound to be the sum of a set of sinusoidal components, or harmonics. A basic presentation of synthesis and analysis techniques will follow. These are based on computer processes that analyze a real tone, which has been recorded and digitized,' into time-varying frequency and amplitude functions for each of its harmonics. A concrete example of this is given in Figure 1 for the first four harmonics of a violin tone. Given the results of analysis, we can then reproduce the tone by additive synthesis, where the set of sinusoidal components are controlled in amplitude and frequency by the analyzed functions, and their outputs are added together to constitute the complex music instrument tone. Various other methods for displaying sets of amplitude and frequency functions are given in Figures 2 through 4.

synthesis

In additive synthesis we physically model a complex sound waveform as a sum of sinusoids with slowly time-varying amplitudes and phases. The process of synthesis involves specifying the amplitude and phase (equivalently, amplitude and frequency) for each component sinusoid as it varies with time throughout the duration of the tone. We will generally refer to this specification as being a time-varying function, amplitude or frequency, for a component sinusoid. These sinusoids are added together to produce the complex waveform. Equation (1) summarizes this formulation.

$$(1) \quad F_{\alpha} = \sum_{n=1}^{M} A_n \sin(\omega_n \alpha h + \theta_n)$$

Notation:
- α is the sample number
- h is the time between consecutive samples
- F_{α} is the sampled, digitized waveform at time αh
- A_n is the amplitude of the nth partial tone and is assumed to be slowly varying with time
- θ_n is the phase of the nth partial tone and is assumed to be slowly varying with time
- ω_n is the radian frequency of the nth partial tone

One can see that from this model, if we can determine the functions A_n and θ_n of a tone from a musical instrument, we can then synthesize an approximation to the waveform F_{α} from those functions by use of equation (1). The degree to which this form of synthesis has been successful will be discussed below. To determine the functions A_n and θ_n of a music instrument tone, we must assume the frequencies of the partial tones, ω_n, are nearly harmonically related. By harmonically related, we mean that the tone has a fundamental frequency, ω, and that the frequencies of all the partials of the tone are integer multiples of the fundamental frequency. That is, the frequency of the n^{th} partial, ω_n, is approximately $n\omega$.

It should be pointed out that equation (1) could have been formulated with time-varying frequencies and constant phases. This formulation is equivalent and for all practical cases, the one can be derived from the other. We will speak interchangeably of the 'phases' of the harmonics and the 'frequencies' of the harmonics as a function of time. In the context of the analysis of tones, it is most natural to produce the phases of the harmonics as functions of time, as is shown in Appendix A. For intuitive purposes, however, it is more instructive to view displays of the frequencies of the harmonics as functions of time, and we will therefore usually refer to the frequency (rather than phase) functions of harmonics. In Figure 1 we present an example of a set of time-varying amplitude and frequency functions for four component sinusoids of a tone. We would use these functions in additive synthesis to control the amplitudes and frequencies of four components of a complex tone. In fact, they would constitute the first four harmonics of the tone, their average frequencies being approximately 308 Hz, 616 Hz, 924 Hz, and 1232 Hz, respectively. We should note that these functions were actually the result of the computer-analysis of a real tone, which was tape recorded and then digitized.

analysis for additive synthesis and graphic techniques

The method we have found most useful for analysis we call the 'heterodyne filter.' This is described in detail in Moorer (1973) and is derived briefly in Appendix A. We take the digitized waveform of a single music instrument tone and for each harmonic under analysis, we perform the following operations: We form the products of the digitized waveform with a sine and cosine at the frequency of that harmonic and compute the average of each product over one period of the fundamental frequency of the tone. The square root of the sum of the squares of these two averages is an approximation to the amplitude of the harmonic in question at that point in time. The inverse tangent of the ratio of these two averages is an approximation to the phase of the harmonic in question at that point in time. We repeat this process throughout the duration of the note for all harmonics.

As aids to the researcher, we have designed several different methods for displaying the results of analysis. The output of the heterodyne filter can, of course, be displayed as a number of isolated amplitude and frequency functions, covering the individual components, as is shown in Figure 1 for the first four harmonics of a violin tone. The total duration of the tone is about 400 milliseconds and its fundamental frequency is about 308 Hz. Sixteen harmonics were actually analyzed for this tone, but we present the isolated plots for only the first four of these in Figure 1. Three more pages of such plots would cover the remaining harmonics, however, the first page is sufficient to get a feeling for the sort of information which can be obtained from this form of display.

To obtain a more easily-grasped picture of the relationships between all harmonics of a tone, it has been found most informative to view the entire set of harmonics together. One method designed for this is the three-dimensional perspective plot. Figure 2 shows such a plot of the amplitudes of all sixteen partials of the same violin tone. The fundamental appears as the backmost function in the picture, while the highest harmonic is represented as the frontmost function. This form of display allows us to more readily discover relationships among the harmonics. The perspective plot can be spatially rotated on-line by the computer, so that the observer is able to see the three-dimensional representation from any angle. This has been very helpful in getting a more comprehensive understanding of the behavior of the partials of a tone as a function of time.

Another form of display revealing the evolution of the partials of a tone as a function of time is the sequential line-spectrum plot. Here, we make use of animation techniques to present successive moments in the tone, presenting a plot of the amplitudes of all the harmonics at each moment in time. One plot is shown in Figure 3, taken from the middle of the violin tone. This strictly on-line display presents such two-dimensional frequency by amplitude plots of the partials for successive instants in time, and the viewer can follow the amplitude changes for the partials from the beginning to the end of the tone.

A fourth way of examining the output of the heterodyne filter, inspired by the conventional speech spectrograph, is given in Figure 4. The particular advantage of this form of display is that it presents both frequency and amplitude information at once in a concise plot, allowing us to view relationships between the two as functions of time. Here, the thickness of each bar is proportional to the log of the amplitude of that harmonic. The vertical position represents its instantaneous frequency, as determined from the phase drift of the harmonic. The utility of this display is its representation of the phase information with respect to amplitudes.

CURRENT RESEARCH

We begin the discussion of our current research with the results of the perceptual evaluation of the analysis-synthesis strategy, a necessary test of the usefulness of this approach in simulating natural tones. Experienced listeners verified that the strategy of additive synthesis based on the results of heterodyne analysis is indeed capable of producing tones that are perceptually indistinguishable from their respective original recorded and digitized tones. The main

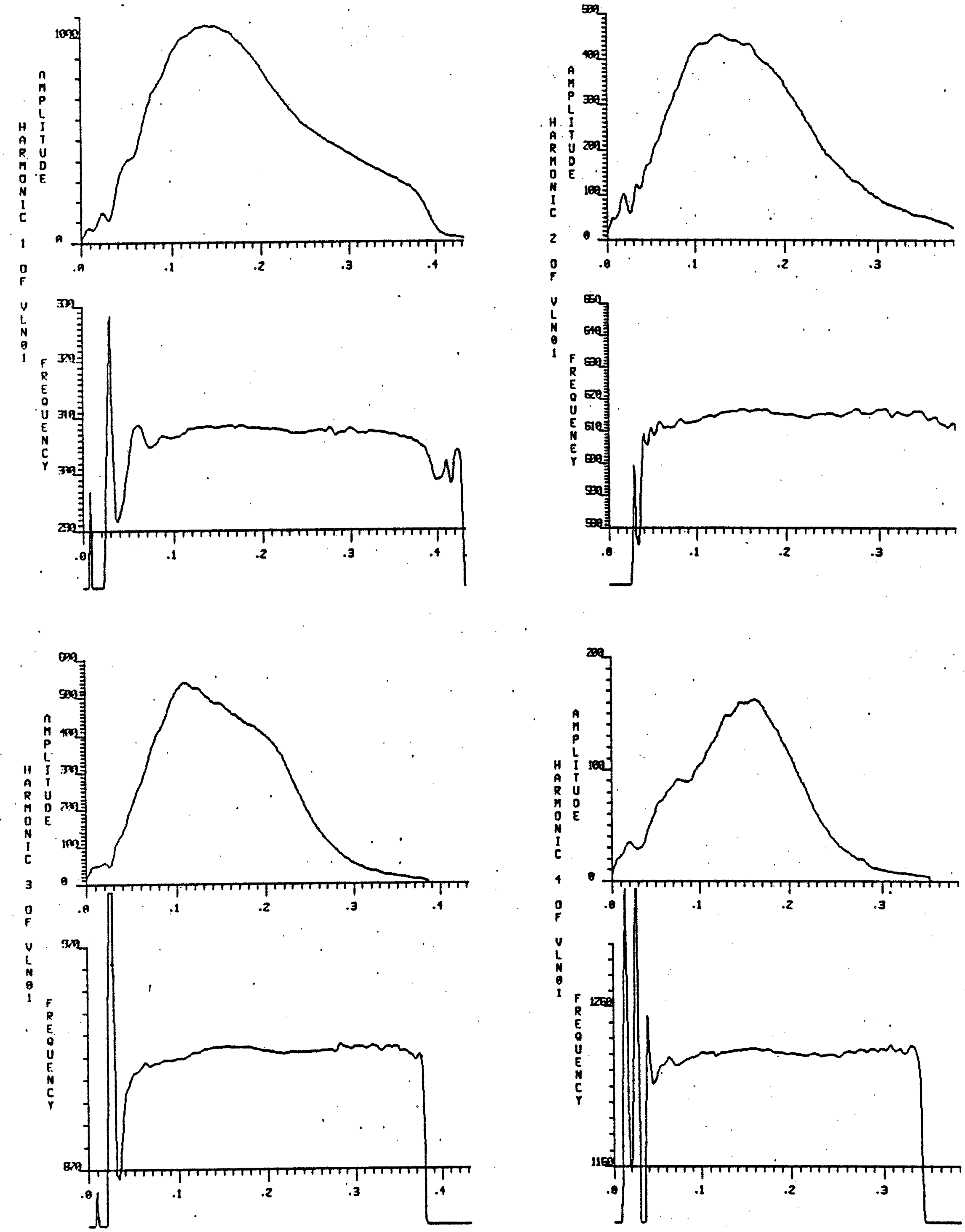


Figure 1. Time-varying amplitude and frequency functions for the first four harmonics of a violin tone. These were obtained from the analysis of a digitized tone with the heterodyne filter. (abscissa=time, ordinate=frequency or amplitude)

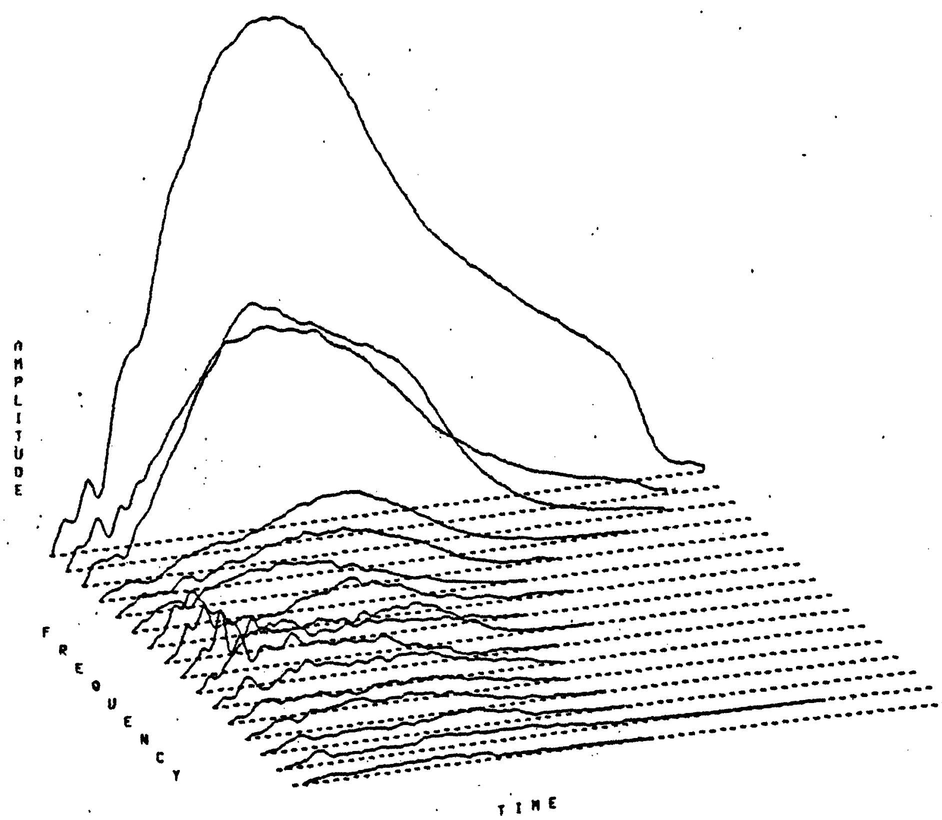


Figure 2. Three-dimensional perspective plot of the amplitude functions for the 16 harmonics of the preceding violin tone. (x=time, y=amplitude, z=frequency: fundamental is backmost) This plot can be rotated (on-line with the computer) by the researcher.

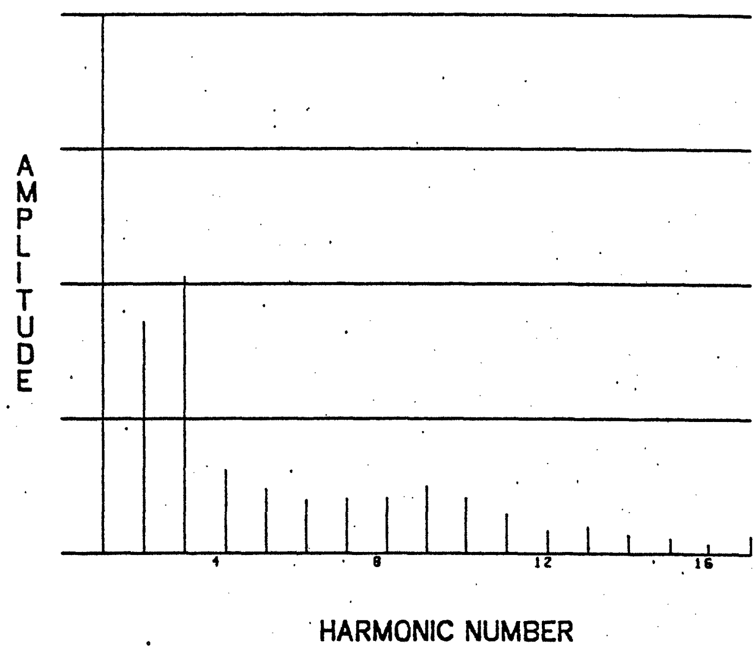


Figure 3. Average line-spectrum of the preceding violin tone, for a time segment of about two milliseconds, which occurs about half way through the note. (abscissa=harmonic number, ordinate=relative amplitude) This is but one of a number of sequential plots which, with (on-line) computer animation, display the changing amplitudes of the harmonics of a tone through time.

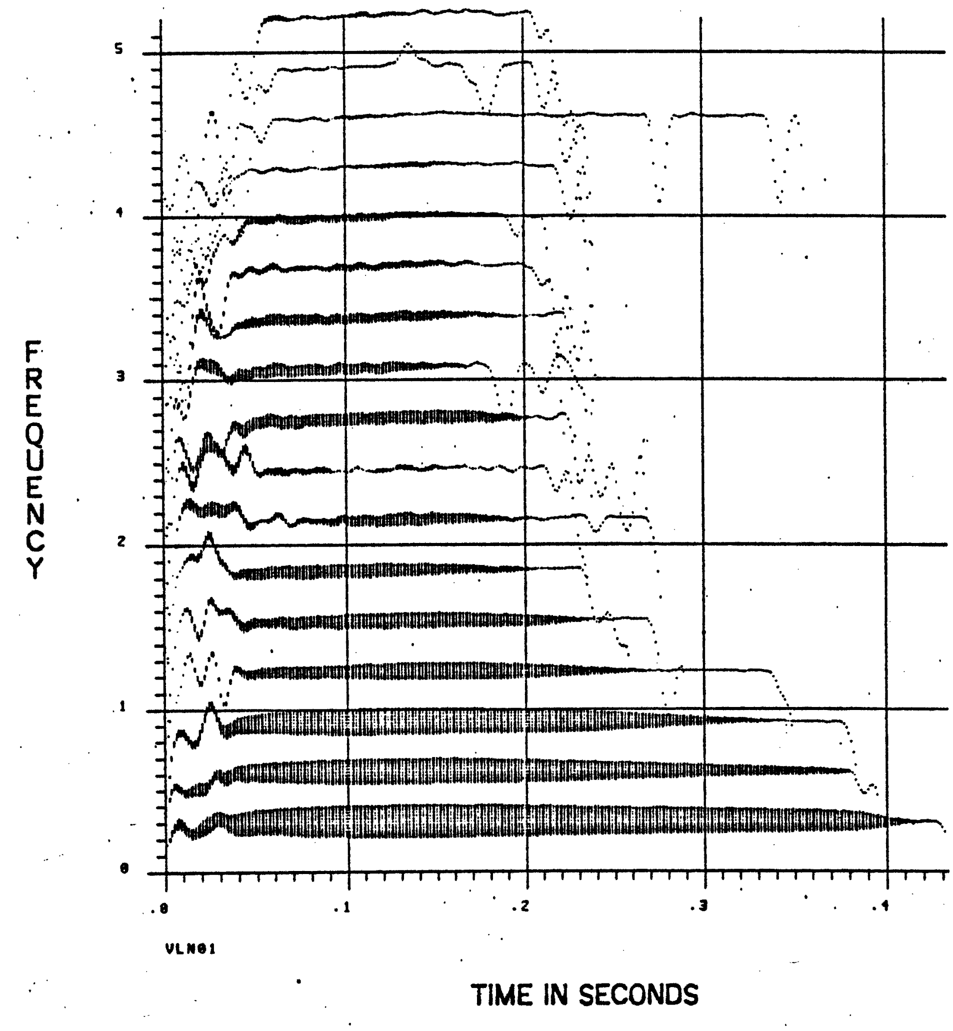


Figure 4. Spectograph of the preceding violin tone (abscissa=time, ordinate=frequency in kHz: bar-height=relative intensity where -40dB from the maximum intensity is a single dot).

function of the analysis-synthesis technique is to provide a starting point for simulating signals which retain the perceptually salient features of complex natural tones. We begin with the most complete set of data on the signal which can be provided by analysis. We are assured that additive synthesis based on this highly detailed information will produce a signal which is indistinguishable from the original (compare Examples 1 & 2, the digitized and synthesized versions of the violin tone given in Figures 1 through 4).

Our goal is to determine which aspects of the simulated signal, and therefore the original signal, are perceptually important. Some attempts have been made in the past to reduce the highly complex information derived from the analysis of real tones to just those perceptually important aspects of the signal which are necessary for its simulation (Risset, 1966; Strong & Clark, 1967a, 1967b). Our current research involves an extension of this type of work. We will first discuss the systematic modifications of tones which we have performed in our work, namely the filtering of signals in order to localize their most perceptually important components. The results of synthesizing tones from selected components rather than the full set of analyzed harmonics, a highly accurate form of 'filtering', was found to provide a means to study the important cues for the identification of instruments.

We will then describe a main direction in our current research: the simplification of the very complex data structure which is obtained from the analysis of the physical properties of a real tone. This process, generally referred to as data reduction, serves many of the goals which we have set for an ideal simulation algorithm. We will describe the remarkable success which we have had in using small numbers of line segments to represent the time-varying amplitude and frequency functions for the components of various tones. We will mainly cite the results of studies on the violin as one concrete example, and briefly mention related findings for other types of instruments. A strikingly successful reduction of the complex data obtained from the analysis of a violin tone (shown in Figures 1 & 2) was a representation of each amplitude and frequency function by only three line segments (shown in Figures 5 & 6).

perceptual evaluation of analysis-synthesis strategy

It is necessary to confirm the utility of our analysis-synthesis strategy for the simulation of musical tones on the basis of the perceptual success of the synthesized signal. That is to say, we must establish that the signal which is produced by our analysis-synthesis technique is indiscriminable from the digitized recording of the original sound. The critical test, then, is a comparison of the sound which has been produced by additive synthesis with the original musical tone that was analyzed with the heterodyne filter. Informal experimentation, in which experienced listeners compared the original recorded tones, played directly back after digitization, with the tones that were synthesized on the basis of analysis, has shown that the analysis-synthesis method produces an extremely convincing replication of the original signal. The strategy thereby has been perceptually verified as being capable of reproducing natural tones. This is in agreement with the findings of other investigators who have attempted to verify similar analysis techniques by comparing tones synthesized from analysis with the original signals (Risset, 1966; Freedman, 1967, 1968).

Among the types of natural musical signals which have been successfully simulated by the analysis-synthesis procedure are tones from the string, woodwind, and brass families of the orchestra. Specifically, we have been able to reproduce tones of various pitches and durations from the following instruments:

> ***violin, viola, cello, double bass, trumpet, trombone, French horn, baritone horn, oboe, English horn, bassoon, Bb clarinet, alto clarinet, bass clarinet, flute, alto flute, alto sax, soprano sax.***

The comparisons between the original digitized tones and their respective simulations by experienced listeners, musicians and acousticians, have demonstrated the potential power of our analysis-synthesis technique.

filtering of signals to localize perceptual cues

One sort of modification which we have employed to reveal perceptual cues for the identification of tones is that of filtering. Directly given by our analysis-synthesis method is the power to precisely select the harmonics which will be synthesized. We have applied this modification to a number of signals, to make a preliminary evaluation of its usefulness for localizing, in the frequency plane, perceptual cues for the identification of a number of music instrument tones. As we had expected, there is a fairly broad variation in the minimal number of lower harmonics which are necessary to transmit the identity of an instrument. This variation occurred even though many of the tones started with the same number of harmonics. A close study of the variation of identification with the low-pass cut-off frequencies for the individual tones gave a rough estimate of the location of various perceptual cues for these instruments. We concluded from these preliminary tests that the selective filtering of signals could indeed provide much significant information about the relationships between the physical properties of sounds and their perceived qualities. We propose further work below, but give here an example of the results of this testing on the violin.

The violin tone examined is the one displayed in Figures 1 through 4. A number of filtering operations were performed on this signal, and we had experienced listeners, including several musicians, attempt to identify which instrument had produced the tones. Identification of this particular source was increasingly difficult for most listeners in the low-pass filtering condition as the cut-off was reduced below the tenth harmonic. At that point the source of the signal was not definitely identified as a violin, but any of a set of string-instruments. This contrasted to the results obtained with various other instrument sources - some of these had an equal number of analyzable components to start with - which could be correctly identified with much lower cut-offs. For example, the clarinet, which also started with 16 harmonics, could be identified from only harmonics 1, 3, and 5. With the violin, however, the prominent activity in seventh through eleventh harmonics, especially during the attack, may be implicated to be of great perceptual importance. In a high-pass filtering condition, it was found the identification of the violin source was difficult when only the first three harmonics were absent. This was found with most of the other signals tested (however, prominent cues for certain brass instruments were associated with patterns of modulation during their attack segments, and these sources could often be identified from a single component which displayed the modulation pattern). More complex selective filtering strategies again confirmed the importance of the activity of the seventh through eleventh harmonics for the identification of the violin tone.

data reduction

A general strategy for data reduction is presently being persued. The complete data is initially represented by 400 to 500 line-segments per amplitude and frequency function per harmonic. (This, of course, is a reduction of data from the 25,000 points per second directly produced by the analysis, but does not significantly distort the complexity of microstructure within the functions.) We are in the process of determining the minimum number of line-segments per function which will allow for a successful simulation of the original signal. Data reduction will be found to depend on an empirical verification of the perceptual fitness of measures taken: the success or failure of a current data reduction strategy contributes to the understanding of the salient features in the perception of musical tones; this understanding, in turn, directs the next stage of data reduction.

Reduction efforts have met with a surprising degree of success. Simulations based on as few as three line-segments per amplitude and frequency function have been indistinguishable from the original signals by experienced listeners. The shapes of the reduced functions are now empirically derived from the originally complex curves. An example of a successful three line-segment data reduction of the amplitude and frequency functions for the violin tone shown earlier in Figures 1 & 2 is shown in Figures 5 & 6. This represents an enormous step in data reduction, where, for example, the violin tone referred to could be represented by a total of less than 200 numbers rather than over 16,000 (assuming approximately 500 segments per function). The resources of the computer are optimally used with respect to the storage of parameters for synthesis, the number of input-output operations required, and the size and complexity needed in the program used for synthesis.

The success of these reductions also present a major discovery about the perception of timbre, since the subtle micro-fluctuations which occur in the physical parameters of these signals seem to have very little perceptual importance, even in laboratory listening conditions where tones are presented in temporal isolation for comparison. At present, we are attempting to represent functions by two line-segments, and have had encouraging results with the violin tone, the only one tested so far. This instrument, we should note, has been one of the most challenging sources for data reduction attempts, and presents a good test of any particular technique. Previously, an effort was made to employ constants, instead of time-varying functions, for the frequencies of the components. This was found to produce a noticeable change in the quality of the violin tone, although several of the other signals tested suffered much less discriminable alterations. The change was described by listeners as a decrease in the strength of the attack of the signal. The tone was still considered to retain the quality of naturalness, as established from informal reports, including the response of an experienced violin player. Our general conclusion from this preliminary study was that time-varying frequency functions are necessary to exactly replicate certain second-order features of tones. This does not preclude the substitution of some other physical manipulation for these features.

three dimensional function editor

In order to assist the researcher in evaluating the importance of observed properties in an analyzed tone which has been vastly simplified, graphical techniques have been implemented to allow three dimensional function editing. The editor is most useful in performing manipulations on a set of functions which are composed of small numbers of line segments. It allows the user to view the set of functions in three dimensions and rotate them into any position. The functions can be altered by attaching a cursor to any breakpoints of the line segments. The cursor can be moved either by light pen or teletype commands, and the function is altered appropriately. The resulting sound will be calculated upon exiting from the editor, and a playing program can be called to play the sound and compare it with the original sound and any other versions desired.

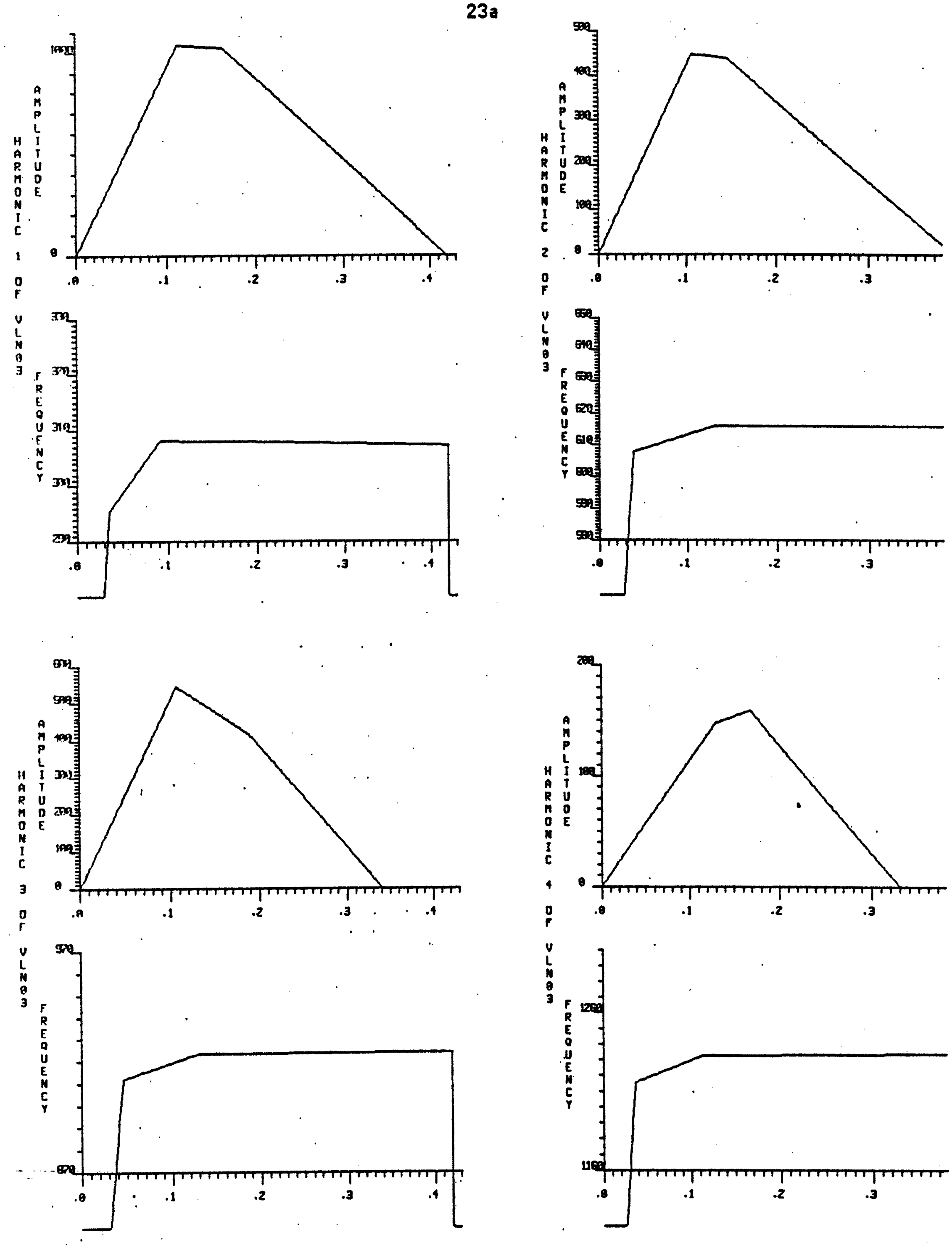


Figure 5.
Time-varying amplitude and frequency functions for the first four harmonics of a violin tone. These resulted from the 3-line segment data reduction of the functions in Figure 1. (abscissa=time, ordinate=frequency or amplitude)

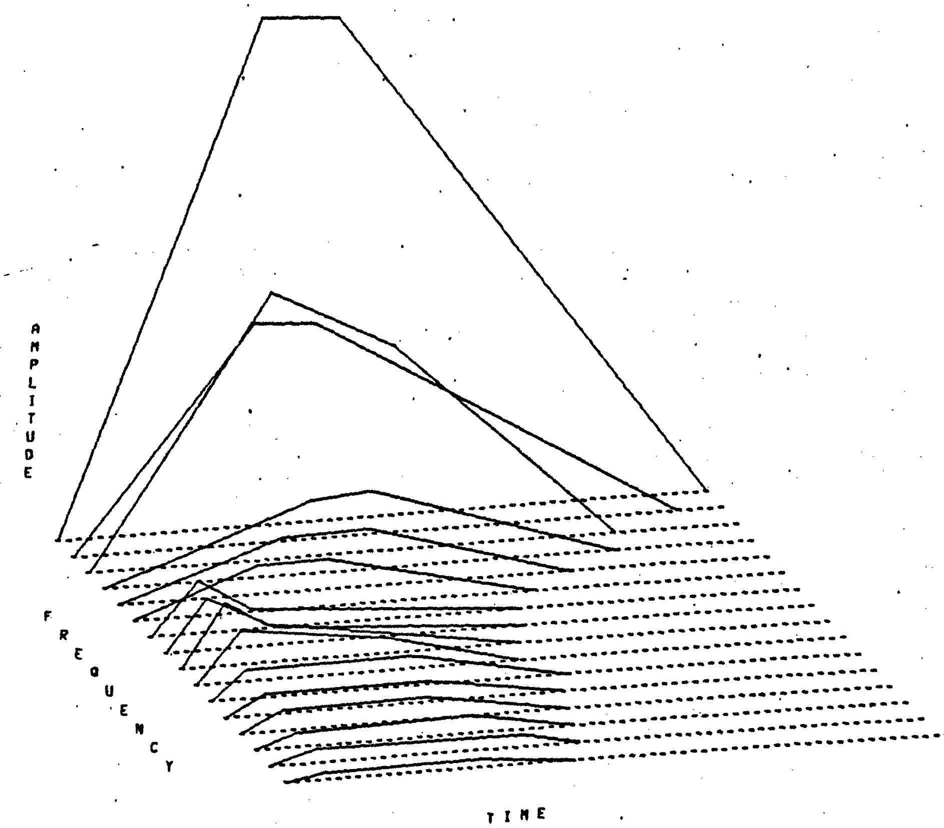


Figure 6. Three-dimensional perspective plot of 3-line segment approximations to the amplitude functions for the 16 harmonics of the violin tone shown in Figure 2. F1(x=time, y=amplitude, z=frequency: fundamental is backmost) This represents significant data reduction.

PROPOSED RESEARCH

We will now turn to our proposed research, and begin by briefly discussing the necessary extension of the range of timbres covered by the additive synthesis technique. The eventual development of truly general techniques are contingent upon this extension of cases examined. We next describe our plans for a systematic exploration of data reduction techniques, which include the rigorous testing of particular methods by perceptual scaling experiments. We then discuss a practical result from this exploration: the development of automatic data reduction algorithms. Reduced data structures for the physical attributes of music instrument tones provides the researcher with a better tool to investigate the more general aspects of timbre perception for whole sets of natural sources. This will be amplified in a latter section, devoted to the applications of multidimensional scaling techniques to timbre perception. We will here discuss the higher-order algorithms which should result for additive synthesis from the above research, algorithms which give the user perceptually meaningful controls and which make optimal use of computer resources for the simulation of tones.

extension of timbral range

A necessary step for the eventual development of truly general simulation techniques is the application of our methods to an extended set of sources. For this purpose, we are planning to gather a large collection of tones from string, woodwind and brass families of musical instruments. Notes at several durations, played in different manners, will be recorded throughout the ranges of all instruments in the above families. As our research progresses in time, we will cover a broader base of signals. We will thereby investigate the perception of a very diverse set of cases and be guided to a more general system for simulation. In that the goal of our endeavors is to develop a technique by which we can realistically simulate any instrumental sound, having any specific characteristics in any context that could occur in reality, our data base necessarily will be extensive. The widening of this data base is an important part of our future research, and an integral feature of all phases of investigation that are presented below.

systematic exploration of data reduction techniques

A series of rigorous discrimination studies are planned, in which a wide range of signals and reduction specifications will be investigated. The basic approach consists of observing the perceptual effects of systematic modifications and simplifications of the data which directs synthesis. Listeners will attempt to discriminate the original digitized tones, tones synthesized from their complete analyses, and tones which have been significantly simplified in their parametric data. The results of this testing will give us the strongest evidence for those aspects of the signal which are important to perception and those which are insignificant and need not be present in a simulation, with respect to a representative population of listeners with varying degrees of musical training.

The discriminability of signals is a standard perceptual measurement. The experimental procedure employed for the measurement of discriminability will involve the judgment of 'same' or 'different' for a pair of tones by the listener. The experiment is completely controlled by computer: pairs of tones are randomly selected from the stimulus set and played to the listener; his response is tabulated and the data is analyzed. It should be noted that the computation to synthesize the tones is done beforehand, and the digital waveform representing the tone is stored on the bulk-storage disk. On completion of the computation, the tones are played by the computer through the digital-to-analog converter. The only analog equipment used is the standard audio system.

In that we are concerned with the simulation of signals which are highly realistic to the listener, we will be interested in measuring the 'naturalness' of simulations for listeners, who have varying degrees of musical training. We will test tones both in temporal isolation and in complex sequences, to determine the relative effects on the evaluation of naturalness for tones induced by the context in which they are presented. We realize that this measurement could be subject to much variability, and we feel that it is important to carefully examine factors which might be correlated with this variability, such as the background of the listener and the context of the signals. It will be important to evaluate the adequacy of simulations with respect to these factors, if the techniques that we develop are to obtain generality. Experiments will have listeners apply an N-point rating scale of relative 'naturalness' to a particular set of tones, which will include the digitized real tones and discriminable simulations, some which are the results of drastically simplified methods, such as fixed-waveform synthesis where spectral dynamics are absent.

Our preliminary findings suggest a general success with as few as three line-segments per control function. Even if this vastly simplified representation of natural signals turns out to be the limiting case, rapid progress can be made in understanding the psychophysical relationships in their perception and identification. The investigation of these relationships, between the subjective, perceived qualities of tones and their physical properties, will be vastly facilitated by the simplified representation of their physical properties. The importance of relative slopes of attack and onset times of components, the ranges of variation permissible for spectral levels, and the necessity to exactly preserve various other overall characteristics of the analyzed functions for each harmonic will be closely studied.

automatic data reduction algorithms

As we examine a broader range of signals, we will be able to design algorithms for the automatic reduction of the data from analysis which will replace our initial empirical method of reduction. The first step will be the automation of the process of line-segment fitting of complex time-variant amplitude and frequency functions. The optimal type of fitting procedure, and the range of permissible variability in the reduction, will have been established in the research outlined above for a variety of signals. More sophisticated routines for data reduction will draw from related research on the perception of timbre which is described below.

higher-order algorithms

The systematic exploration of the relationships between the known physical properties of tones and their perceptual correlates will reveal the salient cues for their identification, hence the necessary features for their simulation. Perceptual scaling experiments described below are designed to uncover the dimensions and properties of the subjective space for timbre. With this information we will be able to begin to approach a general model for the auditory processing of complex natural signals. We will also benefit by developing a successful set of strategies for data reduction for the computer simulation of these signals. At this point we will be able to investigate the more central aspects of auditory information processing, the internal representations of complex natural stimuli and the perception of these stimuli in complex temporal contexts. This information will lead to very powerful computer simulation techniques able to produce realistic sounds in highly complex realistic contexts.

Higher-order simulation techniques will be a direct product of the research with additive synthesis in conjunction with findings from the frequency modulation approach described next. A higher-order simulation algorithm will simultaneously provide the user with a powerful level of control over salient aspects of tone while reducing and making more experientially relevant the type of input specifications to the simulation procedure. As we come to understand the perceptually important features of tone, the simulation algorithm will reflect this understanding. Features like the relative attack slopes and onset times of components, expressed in simplified graphical-relational form, such as the overall evolution of the bandwidth of energy distribution of the signal through time, will come to be directly dealt with by the user. Other possible important features of tone, e.g. the modulation of functions in amplitude or frequency or the existence of bandwidths of noise, will be controlled via meaningfully simple specifications by the user.

2. FREQUENCY MODULATION SYNTHESIS

INTRODUCTION TO SYNTHESIS AND ANALYSIS TECHNIQUES

The second method of synthesis is by means of frequency modulation (FM). Our research using this technique has yielded some altogether surprising results, in that a simple algorithm is capable of producing a large number of highly differentiated tones which can have a strong perceptual resemblance to those of natural musical instruments. The technique is based on the sinusoidal modulation of a carrier wave, where both the modulating and carrier frequencies are in the audio band. When the carrier and modulating frequencies are small integer multiples of some frequency, the components of the modulated carrier are in a harmonic sequence and when they are related by irrational multiples, they are inharmonic. Thus, by assigning the ratio of the carrier to modulating frequencies, a large number of harmonic and inharmonic spectra can be produced. The number of significant components in the spectrum is determined by the frequency deviation or amplitude of the modulating wave. Thus, as the deviation increases from zero, more and more components become significant, or we say the bandwidth of the signal increases. In fact, the bandwidth relates indirectly to the deviation and directly to the modulation index which is the ratio of the deviation to the modulating frequency. The power of this technique for synthesis is that the relationship of frequencies and the bandwidth of the signal can be controlled by two parameters. When the modulation index changes as a function of time, the bandwidth of the signal changes more or less proportionally, as shown in Figure 7.

synthesis

Here we produce the tone by varying the amplitudes and phases of a small number of sinusoids. The feature which distinguishes modulation synthesis from additive synthesis is that one or more of the parameters is no longer restricted to be a slowly time varying function. A more complete description of FM synthesis is given in Chowning (1973). We shall describe briefly the essence of the method. The basic equation is shown in equation (2).

(2) $F_\alpha = A \sin(\omega_c \alpha h + I \sin(\omega_m \alpha h))$

Notation:
F_α is the sampled, digitized waveform at time αh
A is the amplitude of the modulated carrier
ω_c is 2π times the carrier frequency
ω_m is 2π times the modulating frequency
I is the peak frequency deviation/modulating frequency
h is the time between consecutive samples

For the basic application, we restrict the modulating waveform to a pure sinusoid. 'I' is called the 'modulation index,' which is the ratio of the frequency deviation to the modulating frequency. For most of the useful applications we allow the amplitude A and the modulation index 'I' to vary slowly with time, while ω_c and ω_m are constant with time.

We may expand equation (2) as follows:

(3)
$$
\begin{aligned}
F_\alpha = A\{&J_0(I)\ \sin(\omega_c \alpha h) \\
&+ J_1(I)\,[\sin((\omega_c+\omega_m)\alpha h) - \sin((\omega_c-\omega_m)\alpha h)] \\
&+ J_2(I)\,[\sin((\omega_c+2\omega_m)\alpha h) + \sin((\omega_c-2\omega_m)\alpha h)] \\
&+ J_3(I)\,[\sin((\omega_c+3\omega_m)\alpha h) - \sin((\omega_c-3\omega_m)\alpha h)] \\
&+ \ldots\}
\end{aligned}
$$

where $J_i(I)$ is the ith Bessel function of the first kind as a function of the modulating index, I. Thus we see that the tone represented by the waveform of equation (2) consists of a series of sinusoidal components whose frequencies are determined by the sum of the carrier frequency, ω_c, and integral (positive and negative) multiples of the modulating frequency, ω_m. As the modulation index increases, the amplitudes of the components change in complex ways, but the general trend is that energy is shifted away from the carrier frequency. When the modulation index is zero, equations (2) and (3) degenerate to a sinusoid at the carrier frequency. As the modulation index increases, more energy is transferred to components representing larger and larger integral multiples of the

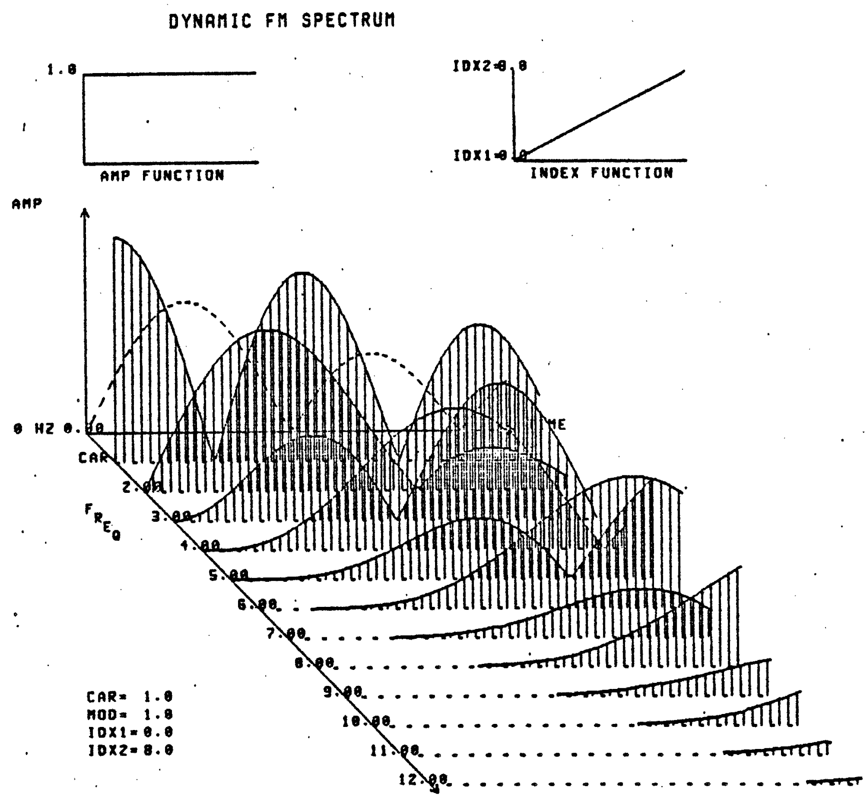


Figure 7. Increasing bandwidth with increasing modulation index.

modulation frequency, as shown in Figure 1 in Chowning (1973).

If the carrier frequency and the modulating frequency are small integer multiples of some other frequency, ω, then the components form a harmonic sequence. If the carrier and modulating frequencies are related by irrational multiples, then the components are inharmonic.

Both of these cases have use in synthesis and both are dependent upon reflected lower side-frequencies where all or some of the components produced by the terms $\omega_c - k\omega_m$ are less than zero. Since this is equivalent to a positive frequency with a phase shift, there are contributions which might not be obvious at first glance. For example, in the case $\omega_c = \omega_m$, the components in the negative frequency domain reflect around 0 Hz and add to the components in the positive domain thus forming the harmonic series where the partial at ω_c will have amplitude $(J_0(I)-J_2(I))$, the partial at $2\omega_c$ will have amplitude $(J_1(I)+J_3(I))$, Figure 4 in Chowning (1973). For the case where $\omega_m = 2\omega_c$, the only nonzero components will be at odd multiples of ω_c, with amplitudes again being sums of pairs of bessel functions. ω_c does not always represent the fundamental frequency of the tone. For instance, if $\omega_c = 2\omega_m$, then a harmonic series based on ω_m is produced. If, on the other hand, $\omega_m = 2^{1/2}\omega_c$, the reflected components will inter-leave with the positive components producing an inharmonic spectrum, Figure 6 in Chowning (1973).

It should be pointed out that while the power of this technique of producing periodic signals is very great, *it does require a degree of precision that can only be attained by digital synthesis techniques,* viz. if, in the case of a ratio of $\omega_c/\omega_m = 1/1$, there were frequency drift of 1Hz in either oscillator, the reflected lower side frequencies and the upper side frequencies would have a 2Hz difference producing a clearly audible beat. Although it is an effect that can be useful, it must certainly be under control.

FM predictive analysis and graphic techniques

At the present stage of research, we rely upon the intuition of the researcher to develop suitable parameter values and amplitude and index functions. There are tools we have given the researcher, however, to aid in exploring the consequences of the various choices. The most powerful of these tools is an interactive graphical program which allows the user to design amplitude and index functions for evaluating equation (3). The program makes a perspective plot of the amplitudes of the partials as a function of time. Figure 7 shows such a plot where the amplitude is constant, the modulation index increases in time from 0 to 8, and the ratio of $\omega_c/\omega_m = 1$.

After the display is produced, the user can then go back and alter his function shapes and parameters to try to converge on the desired spectral characteristics.

CURRENT RESEARCH

Frequency modulation synthesis, unlike additive synthesis, is not at the outset an obvious model for tones which have any likeness to those of natural instruments. The process by which these applications were discovered was initially founded in the ability of a trained 'musical ear' to penetrate the spectral complexity of the FM technique and to predict the effect of a change to the parameters of the equation. The development of analysis and synthesis programs, as described in the previous section, yields precise data describing a complex signal which significantly increases our ability to apply the FM technique to the simulation of music-instrument tones.

We establish in this section the correlation of the parameters of the FM equation to perceptual cues. We select ratios of the carrier to modulating frequencies that produce harmonic spectra appropriate to the simulation. For example, 1/1 for brass tones, which includes all of the harmonics and 1/2 for clarinet tones, which includes only the odd numbered harmonics.

The critical correlation for the production of natural sounding tones is the change of the modulation index as a function of time to the evolution of the bandwidth of the resulting spectrum. Based on the analysis of brass tones by Risset (1966), we were able to synthesize natural sounding brass tones by simply relating the evolution of the bandwidth to the amplitude envelope, as shown in Figure 8. The initial success of this simulation was a surprise in that no one could have predicted on the basis of psychoacoustical knowledge the possibility of such a simple representation of a brass tone.

The FM technique can be extended by adding another carrier wave to the system. If the ratio of the first carrier to the modulating frequency is 1/1 and the second carrier is nine times the first, but modulated by the same modulating wave, then additional energy is added in the region of the ninth harmonic which is similar to a resonance. Figures 9 and 10 show the spectral distribution resulting from the two modulated carrier waves. The bandwidth and amplitude of the resonance can be controlled independently from the the components produced by the first carrier, giving great additional flexibility to the technique.

periodicity - ratio of carrier to modulating frequencies

Since music instrument waveforms are for the most part quasi-periodic, where successive periods of the wave have only small deviations, we select ratios for the carrier and modulating frequencies which produce side frequencies that fall in the harmonic series. As we pointed out in the section on synthesis technique, small integer multiples of some frequency will produce periodic waveforms. Within this generalization there are a number of further options: $\omega_c/\omega_m = 1$, spectrum including all of the partials; $\omega_c/\omega_m = 1/2$, spectrum composed of odd numbered partials; $\omega_c/\omega_m = 1/3$, spectrum where every third partial is missing, and so on.

bandwidth as a function of time - modulation index

An initial period of experimentation resulted in convincing simulations of several music-instrument tones of which the most surprising were those of brass-instrument. These tones were analyzed in order to determine the perceptual correlations to the parameters of the equation (2). The equation is restated where A and I are functions of time.

$$(4) \qquad F_\alpha = A_\alpha \sin(\omega_c \alpha h + I_\alpha \sin(\omega_m \alpha h))$$

The amplitude A and the modulating index 'I' were selected to be slowly varying functions of time. The function A_α controls the time-varying amplitude of the frequency modulated carrier to produce the appropriate attack and decay characteristics of the tone. The function I_α is used to control the evolution of the bandwidth of the spectrum. Since in brass tones the energy of the higher partials increases roughly in proportion to the increase in intensity during the attack (Risset, 1966), the shape of the amplitude function was used for the index function in order to produce this correspondence. The value of the index was set to range from a minimum of 0 to a maximum of 5 according to the function shape, which produces nine significant partials at the peak amplitude of the wave. Figure 8 is a perspective plot of the spectrum of an FM synthesized brass tone with the functions controlling A and I at the top. The increase of bandwidth with amplitude can easily be seen.

The extent to which the application of this technique is able to simulate the natural qualities of a brass-tone is remarkable and serves to emphasize an important physical correlate to perception: *the temporal evolution of the bandwidth of a signal.*

Our preliminary investigations suggest that the evolution of the bandwidth is of varying degrees of perceptual importance, often primary, in all instrument tones and a large number have been synthesized using equation (3), where the only time-domain functions were amplitude and modulation index (Chowning, 1973).

resonances - multiple carrier waves

There are many signals which have spectra for which there is no approximation using the FM equation in its simple form; in particular, the spectrum may have two regions of significant energy which are caused by the natural resonances of the instrument. There may be no value of the modulation index applied to equation (3), which will produce a spectral curve that preserves the effect of these resonances. In this case, we extend the FM synthesis algorithm to include additional carrier waves whose frequencies are set at the harmonics around which the regions of spectral energy are centered. The expanded equations (5,6) include a second carrier wave. For the purpose of simplicity in the notation, A and I are assumed to be slowly varying functions of time as explicitly notated in equation (3).

(5) $M = I \sin(\omega_m \alpha h)$

(6) $F_\alpha = A \sin(\omega_{c_1} \alpha h + M) + K_1 A \sin(\omega_{c_2} \alpha h + K_2 M)$

Notation: ω_{c_1} is the first carrier frequency
ω_{c_2} is the second carrier frequency
K_1 is a constant to scale the amplitude function for the the second carrier
K_2 is a constant to scale the index function for the second carrier

An application of this technique is appropriate in simulating the spectral distribution of a string tone which has a resonance near the ninth harmonic. We break the desired spectrum into two parts, each of which outlines a spectral shape which seems similar to an FM produced spectral shape. In this case the ratio of the first carrier to the modulating frequency is set to 1, and the index ranges from a maximum of 4, during the attack, to 1.5, where the amplitude is maximum. This produces a spectrum shown in Figure 9. The second carrier has a ratio to the modulating frequency of 9/1, the index is scaled by the constant .5, to produce the index range 2 to .5, and the amplitude is scaled down by the constant 0.3. The spectrum of this part is shown in Figure 10. The sum of these two spectra has attributes of a string tone.

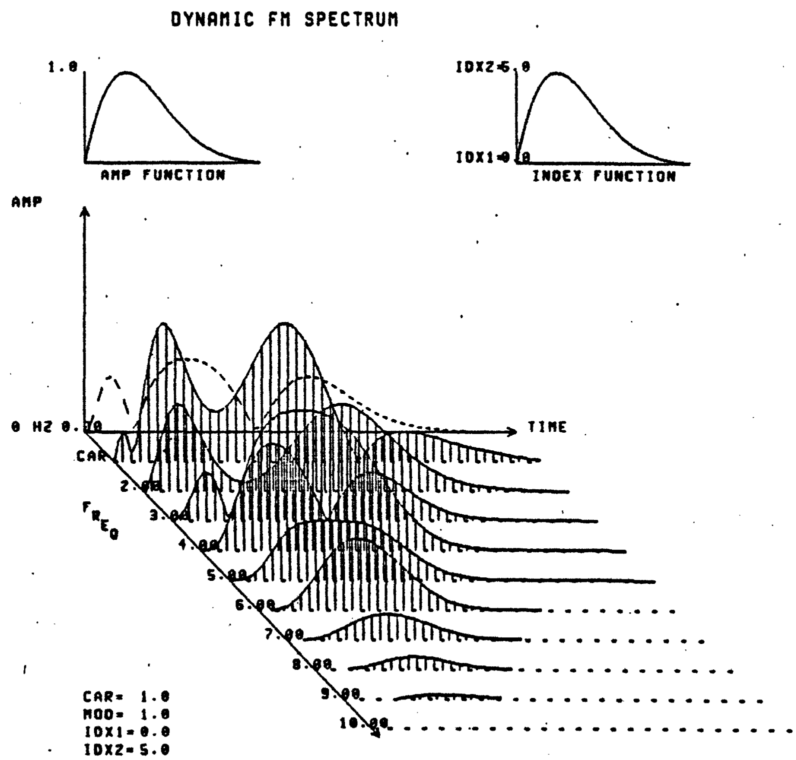


Figure 8. FM synthesized brass tone.

PROPOSED RESEARCH.

The research that we have described above has been heavily dependent upon the interaction in application of the two synthesis techniques to identical or similar signals. Each technique has provided insights into the critical cues which have been applied to the other as a means of confirmation. Although the number of signals we have considered is a small fraction of the total we need in order to generalize the synthesis procedures and ultimately theorize a timbre space, the number does include highly diversified spectra which have *given in* to our procedures of reduction and FM synthesis. We are therefore confident that within the general constraint of quasi-periodicity we can successfully apply these techniques to a large number of signals.

In order to pursue further the FM synthesis of some music-instrument tones, we must first extend the technique of the synthesis of resonances. Here, we rely upon the results of the analysis, data reduction, and additive synthesis to quantify the fixed resonances of instrument tones such as the violin. On the basis of this information we will construct a resonance table for the instrument. The table will contain three values for each of the tempered pitches in the normal range of the instrument describing, 1) the harmonic around which the harmonic occurs, 2) the relative amplitude of the peak of the resonance, and 3) the bandwidth of the resonance. Figure is an example of such a resonance where the values are harmonic = 7, amplitude = .3, and bandwidth = ca 1 at the peak amplitude of the tone (as noted before, naturalness is dependent on significant change in the bandwidth, especially during the attack).

A very powerful means of simulating secondary features of tones such as inharmonicity during the attack period, is the use of a *complex* modulating wave. In the case of string tones the noise or scratch is of very great importance. Of particular importance in relation to the additive synthesis technique where the attack noise resulted from asynchronous frequency perturbation of each of the harmonics, is the apparent latitude in producing this cue. We will investigate, therefore, the application of complex modulating waves to a variety of other instrument tones which have significant noise components.

To extend the timbral range of simulation, we will also apply the technique to the synthesis of non-periodic signals in order to measure the correspondence of critical cues such as bandwidth in the two classes of signals. The FM technique has very great potential in this class of timbres.

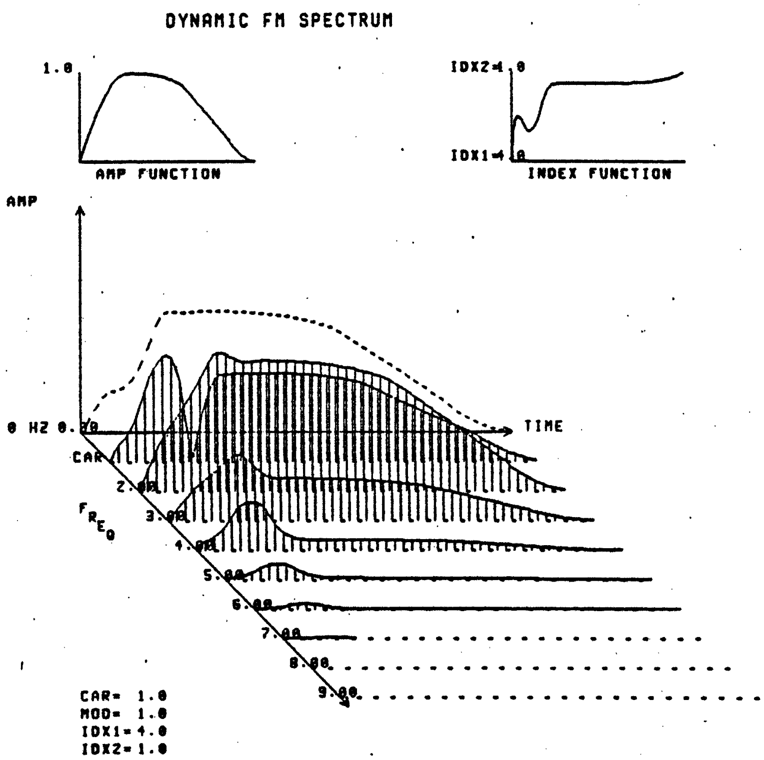


Figure 9. Lower region of spectrum for FM synthesized string tone.

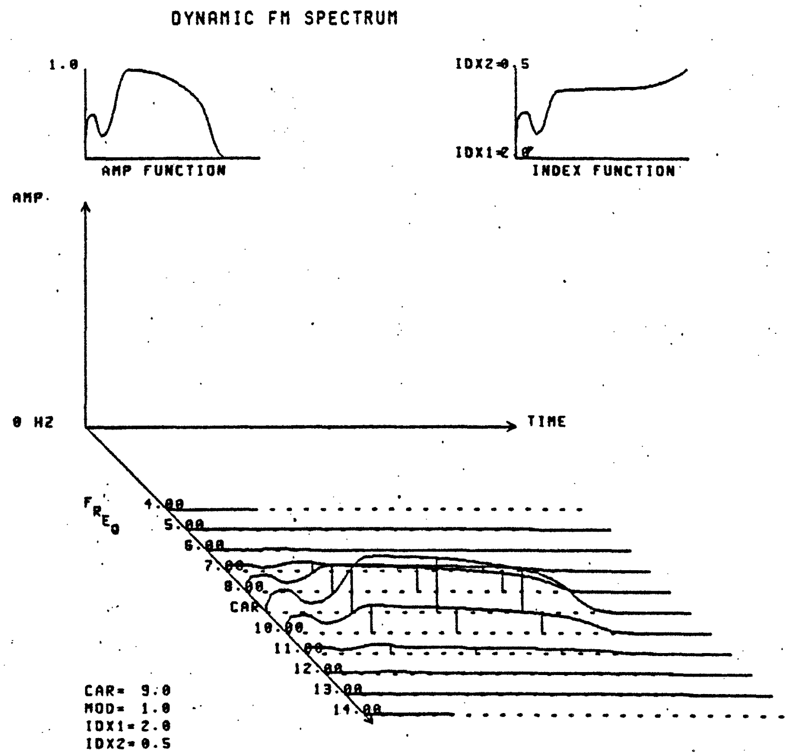


Figure 10. Upper region of spectrum for FM synthesized string tone.

simulation of fixed resonances

Our research in the simulation of music-instrument tones, using the additive and FM synthesis techniques, has focused on single representative tones. In order to extend and test the data reduction procedures we must apply them to the analyzed data from a music-instrument within a range of frequencies and articulations. It is well-known that many instruments have fixed resonances and that these resonances cause significant changes to the shape of the spectrum as the frequency of the fundamental changes. The analysis-synthesis-reduction techniques should highlight these effects and allow us to produce spectral envelopes in a concise form for the entire frequency range of an instrument.

In order to account for the effect of the fixed resonance we must provide for a higher level of control. The obvious analog to the resonances of a musical instrument are band-pass filters or resonators which are tuned to the observed resonances of the instrument in question. As the period of the signal changes, the harmonics shift their position in relation to the resonators thereby effecting the desirable spectral shaping.

Mathews (1973) has shown that a large number of resonances are exhibited by stringed instruments. If we were to digitally simulate the physical waveform of the violin, based on results of research in the physics of stringed instrument tone production, the computations required would be extremely extensive. One realization of resonances that is suggested by the FM technique is the form shown in equations (5) and (6) which has two carrier waves. Two carrier waves can synthesize one (or to some extent, two) fixed resonances.

The additional carrier wave provides for arbitrary placement, amplitude, and bandwidth of a resonance for any particular fundamental pitch period. The second carrier frequency would be set at the partial number which is closest to the center of the resonance for a given fundamental frequency, and the bandwidth and peak amplitude of the resonance would be controlled by the appropriate scale factors. It should be pointed out that in typical applications, the second carrier frequency may be 5 to 10 times the modulating frequency and the index relatively small. Since in such a case there are no significant *reflected* side frequencies, the components are symmetrical about the carrier and have spectral envelopes which are similar to those of natural resonances. The three controlling values would represent the best approximation of the spectral envelope as determined by the analysis-synthesis-reduction technique and as constrained by the form of equations (5) and (6). This procedure will be repeated for the n pitch periods of interest and a 3 by n table generated to store the values for harmonic center, amplitude scale, and index scale. The FM synthesis algorithm would simply extract the values according to the scale step and synthesize the tone accordingly. Figure 11 is a spectral representation of a set of values applied to the second carrier wave of equations (5) and (6). In this manner, we may synthesize with little additional computation the effect of a single fixed spectral resonance peak with a frequency-modulation instrument having two carriers.

inharmonicity - multiple modulating waves

One of the secondary characteristics of many tones, and in particular string tones, is the noise or 'scratch' which is apparent during the attack. The heterodyne analysis and additive synthesis techniques indicate and confirm respectively, that the noise is a result of frequency disturbance of the harmonics during the attack, shown in Figure 4. The reduction techniques have demonstrated that nearly any inharmonicity during the attack will preserve the characteristic 'scratch.' In order to simulate this using the FM technique, we introduce an additional modulating wave which has an irrational ratio to the fundamental and which has non-zero index only during the attack portion of the tone. Equation (7) shows two sinusoidal modulating waves and a single carrier wave.

$$(7) \quad F_\alpha = A \sin(\omega_c \alpha h + I_1 \sin(\omega_{m_1} \alpha h) + I_2 \sin(\omega_{m_2} \alpha h))$$

In this case where the modulating function is not a single sinusoid, but a sum of sinusoids, the resulting expansion is similar to equation (3); however, the components will be at the set of frequencies consisting of the *sum* of all integral multiples of the modulating frequencies added to the carrier frequency. The amplitudes of the components will be products of Bessel functions of the modulation indices. We can produce combinations of harmonic and inharmonic spectra simply by the choice of the modulating frequencies.

The simulation of a violin tone would thus far incorporate, then, both two carrier and two modulating waves for the closest approximation.

non-periodic tones

We have already demonstrated in the laboratory the utility of the FM technique for simulating non-periodic music-instrument tones such as bells, drums, gongs, etc. This class of tones has the common characteristics of primary frequency components which do not fall in the harmonic series, decay times which are long compared to the attack and steady state (if any), and decrease in bandwidth which follows, more or less, the decay. In FM synthesis the frequency distribution of such a tone is controlled by setting the ratio of the carrier and modulating frequencies to be irrational, and the bandwidth is controlled by allowing the modulation index to follow the decay curve. In most cases the spectrum degenerates to a sinusoidal oscillation as the amplitude of the tone goes to zero, in which case the index also goes to zero. The spectral evolution of a bell tone is shown in Figure 12.

The spectral evolution in time of non-periodic tones has features which are common also to periodic tones. The dimensions and structure of the timbre space must account for both classes of tones. As an example, we have determined through the FM technique, that the only difference between the tone of a plucked string and a bell is one of periodicity, or a wave composed of harmonically related frequencies, and non-periodicity, the amplitude envelope and evolution of the bandwidth being common. We can imagine, then, a spatial representation of timbre where the variance for this case is along a single dimension.

3. TOWARDS A GENERAL MODEL FOR SIMULATION

CURRENT RESEARCH

We will here discuss an important aspect of our current research which is directed toward our ultimate aim: the development of a general model for the computer simulation of natural tones. The two approaches to simulation described above, one using additive synthesis based upon analysis and the other using frequency modulation synthesis, have not proceeded independently of one another, but interactions have occurred at several levels. An example is provided by the initial use of frequency modulation for the simulation of a brass tone which was strongly influenced by the research of Risset (1966) on brass tones, using additive synthesis based on analysis.

The eventual development of a general model for simulation will be an outgrowth of the interdependency and convergence of these two approaches. Examples are given where findings using one technique are applied and tested with the other, providing a cross-verification of research discoveries. Moreover, the particular advantages of each technique influences the direction of research using the other technique, and, in this way, both approaches are in the process of converging on a single, more powerful and general model. This resultant model for simulation will have the advantages of both methods: the simplicity and perceptual

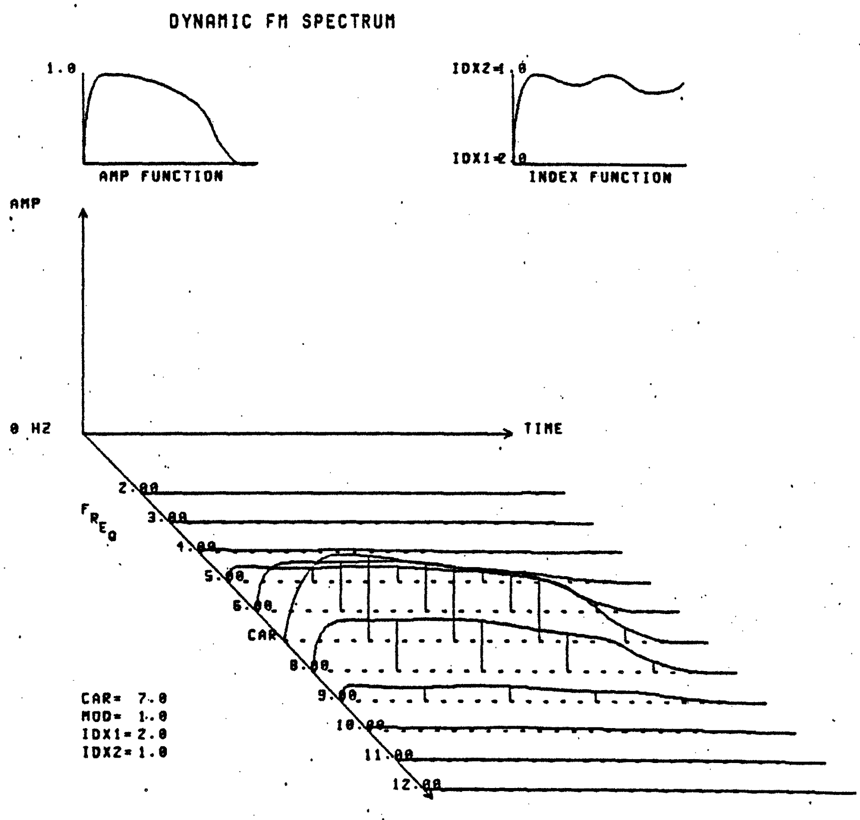


Figure 11. FM synthesized example of a single formant region.

meaningfulness of user-control over the FM technique, and the wide range of complex cases of natural tones handled by the additive technique.

interactions between additive and FM syntheses

One example of the interactions which have occurred between the two approaches discussed above, one using additive synthesis based on the analysis of real tones and the other using frequency modulation synthesis, is the initial discovery of the potency of the FM technique to synthesize periodic music instrument tones. The translation of the distinctive cues for brass instruments found by Risset (1966) - which he derived from analysis-based and data-reduced additive synthesis techniques largely analogous to those which we are using - into the parameters for FM synthesis resulted in amazingly successful simulations of brass tones. This provided a confirmation of the nature of the perceptual features seen for the brass family of instruments. It also indicated the power of the FM technique to simulate important perceptual attributes of tone.

Direct interactions between techniques have ensued in our research. Salient features for several instruments discovered in our approach using additive synthesis have been translated into the FM technique, providing a confirmation of the importance of the suggested cues. Modulations which occur in certain brass instruments, especially the French horn, have been found to be critical using both methods of synthesis. The successful simulation of the violin tone using additive synthesis was found to necessitate the preservation of the inharmonicity among the partials in the attack. The application of this finding to FM synthesis was noted above, and puts a new level of timbral complexity within the reach of the latter method. The success found in using any sort of inharmonicity for the simulation of the quality of the violin attack gives us insight into the critical feature of that attack and the range of alternative techniques which can be used to generate it - all of which do not duplicate the actual acoustical waveform of the real tone! Many other applications of findings from the additive approach to FM are in progress, which also include the reed family of instruments.

We find special significance in the convergence of the two approaches, as it relates to the development of a more general model for simulation. The simplicity and perceptual meaningfulness of specifications to the frequency modulation technique points out an important goal for the additive synthesis method. On the other hand, the complexities of tone which are revealed by analysis, and which are

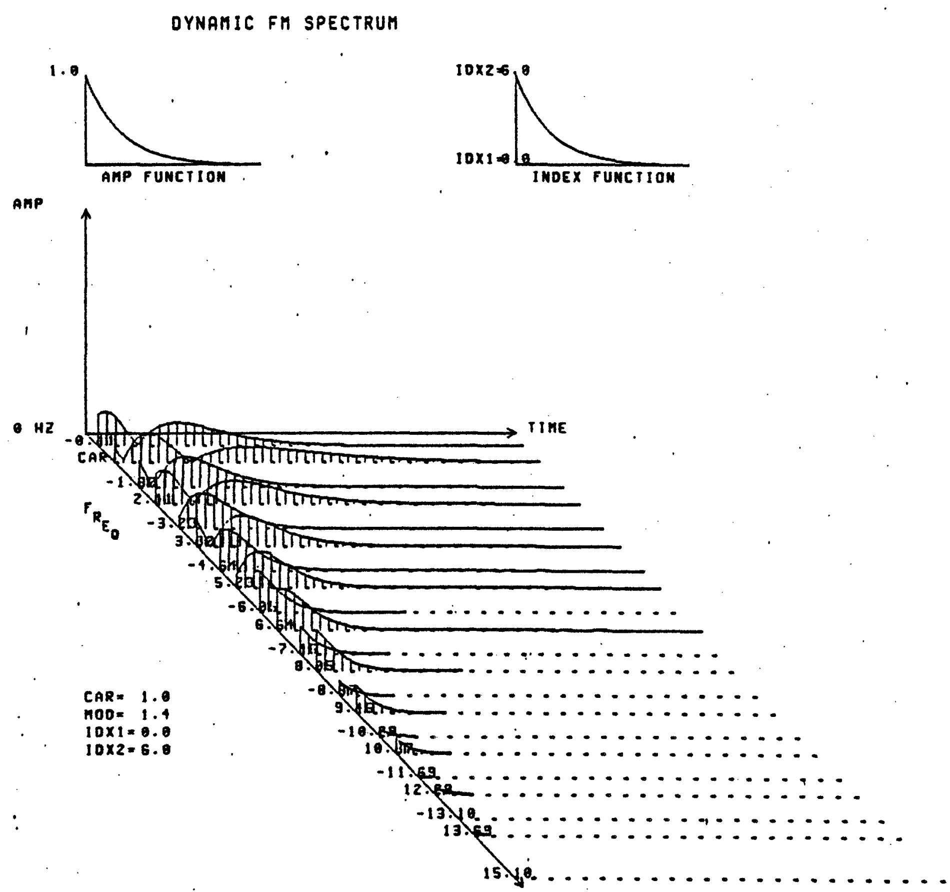


Figure 12. FM synthesized Bell tone, Recorded Example 2.

confirmed to be perceptually salient in the additive synthesis, point out necessary levels of complexity which must be accomodated by the frequency modulation technique. As the latter technique is then made more complex, it in fact enters the category of additive synthesis. The ultimate model for simulation will draw from the research findings using both methods.

PROPOSED RESEARCH

We will here discuss the proposed research which is centrally concerned with approaches to our ultimate aim: the development of a powerful, general, and easily-controlled algorithm for simulation which is based on a comprehensive perceptual model for natural tones. We first will mention our intention to explore the possible use of subtractive synthesis in the simulation of natural music instrument tones. We anticipate that this method will be useful for the simulation of percussion instrument tones such as drum and cymbal. In these cases, one advantage of the subtractive method is that inharmonic partials or even wide-band noise may be easily introduced into the sound. When simulating instruments with certain fixed resonances, one or more filters could be positioned at these resonances regardless of the fundamental pitch period of the exciting waveform. Other applications will be explored, and a significant third approach may result, which would have further advantages besides those of the above two, additive and frequency modulation syntheses, and which would be integrated into our approach towards a general model for simulation.

We will next describe experimental procedures used to investigate the perceptual processing of instrument tones, since, to assist in the development of a general simulation algorithm, we must formulate a general model for the perception of timbre. This will provide important information for the construction of perceptually-based higher-order simulation algorithms. We employ a spatial model for the subjective structure of the perceptual relationships between signals. In particular, multidimensional scaling techniques will be discussed. Research is directed at uncovering the dimensionality of the subjective space, the psychophysical relationships which are structurally correlated to this space, and the properties of the space. The existence of such constraints as categorical boundaries will be investigated in an attempt to assess the continuity of the subjective space for timbre. Of interest is the possible existence of a categorical mode of perception for musical sounds, as has been claimed for speech (Liberman, et. al., 1967). In the same regard, we will also examine the effects of musical training or context on the structure of the space. The model will be evaluated by our ability to predict the mappings of real and novel tones. For the purpose of investigating the properties of a subjective space for timbre, and for testing the existence of a categorical mode of perception, we are designing algorithms which produce new tones whose physical properties lie between two known music instrument tones. An example of one algorithm, designed for additive synthesis based upon the data-reduced analysis of real tones, is shown in Figure 13.

Based on these findings, we plan the design of an algorithm which maps the dynamic spectra of real tones into the FM parameters and time functions. The ability to write such an algorithm would indicate our success at having identified the perceptual dimensions of timbre. This is an important step in the convergence of synthesis techniques toward our ultimate goal stated above.

exploration of subtractive synthesis techniques

A form of synthesis which we have not yet discussed, which actually constitutes the only class of sound synthesis uncovered by our research, is that of subtractive synthesis. The procedure here is to take a simple signal with a wide bandwidth, such as a pulse train or a band-limited sawtooth wave, and apply spectral shaping filters to produce the desired partial tone amplitudes. We have not as yet used this form of synthesis, but intend to do so in the near future. This is the type of synthesis most commonly used in vocoders. We may thus assimilate the techniques of analysis and synthesis of human speech and apply them in a more general context. Two of the most useful methods seem to be the linear predictor (Atal & Schroeder, 1970; Atal & Hanauer, 1971; Markel, 1972) and the homomorphic vocoder (Oppenheim & Schafer, 1968; Oppenheim, 1969; Miller, 1973).

Generally, the technique would be as follows. A musical instrument tone would be analysed at discrete intervals, for instance, every 5 milliseconds. At each analysis point, we compute a filter whose frequency response approximates the spectral shape of the input waveform in the interval around the analysis point. To resynthesize the signal, we filter a pulse train, updating the filter parameters at each analysis point. In the case of the linear predictor, the filter is an all-pole filter. For the homomorphic vocoder, the filter is an all-zero filter. Since these methods are well documented in the literature, we shall not explain them here.

We anticipate that the linear predictor will be useful for analyzing percussion instrument tones such as drum and cymbal. In these cases, the excitation might be modeled as band-limited noise, rather than an impulse train, the spectral shaping being applied by the filter produced by the linear prediction algorithm. Although the same thing could be done with the homomorphic vocoder, a difficult convolution is then required.

The hope is that by using time-varying filters with both poles and zeros, a lower-order filter may be used. At present, the only way of automatically producing the parameters for such a filter directly from digitized sound is by time-consuming optimization techniques. We propose to determine the parameters initially in much the same way as one determines the modulation indices for FM instruments. This involves studying the results of heterodyne filter analysis, the manual preparation of parameter trajectories, and the testing of the results by listening to and analyzing the waveform synthesized by the prepared parameters. We hope to determine whether economical subtractive synthesis can be realized for a wide variety of instruments, and whether an efficient automatic method can be determined for the calculation of time-varying filter parameters.

One advantage of the subtractive method is that inharmonic partials or even wide-band noise may be introduced into the sound by adding such excitation to the driving pulse train. The total excitation will then be passed through the filter and will experience the same spectral shaping that a pure pulse train would experience.

When simulating instruments with certain fixed resonances, one or more filters could be positioned at these resonances regardless of the fundamental pitch period of the exciting waveform. This means that the size of the multiple tables of parameters as functions of the fundamental pitch period might be greatly reduced.

In the event that analysis-based subtractive synthesis proves to be a tool for the extension of the set of tones which we can simulate, it of course will be included in the research program. The future research which follows would in that case incorporate the use of subtractive synthesis in addition to the additive and FM synthesis techniques that we have found useful to date.

applications of multidimensional scaling to timbre perception

A spatial model will be employed to represent the judged relationships between music instrument tones for the purpose of uncovering the perceptual dimensions of timbre. If the reader is not familiar with the computer-based multidimensional scaling techniques discussed below, it would be most instructive at this time to read Appendix B, which introduces the basic concepts of multidimensional scaling and discusses the specific algorithms we will use.

Multidimensional scaling is initially useful for exploring the psychophysical relationships involved in the perception of timbre, that is, the relationships between the subjective, psychological qualities of tones and their physical properties. Interpretation of the perceptual configuration in terms of physical attributes or an actual correlation of the subjective dimensions to physical dimensions in the signals is most desirable. Various attempts have been made in the recent past. Plomp (1970) and Pols (1970) have employed the MDSCAL algorithm to investigate the perception of steady-state auditory signals generated from single periods of musical tones and vowels. Although several problems exist with the specific assumptions of their approach, the most unsatisfactory aspect is the restrictive definition of timbre which excludes the temporal qualities of sound. Wessel (1973b) has used MDSCAL and INDSCAL in a study of perceived and imagined relationships for a set of 9 music instrument tones. Two-dimensional spatial representations were interpretable with respect to the spectral distribution and temporal relations in the onsets of components of the tones as analyzed by speech spectography. The results of Wessel's preliminary study were consistent with a pilot experiment which we conducted with 14 music instrument tones last year. Both indicate great potential for multidimensional scaling and the fruitfulness of its application to a larger range of timbres.

We feel that there is special potential in the perceptual scaling of the computer-simulated music instrument tones described above. The ability to independently control the pitches, loudnesses, and durations of tones that are synthesized makes it possible to experimentally account for these dimensions in the stimuli. This is a non-trivial problem for experiments on timbre perception!

In the case of additive synthesis, tones which are indiscriminable from the original recordings and which are synthesized from very reduced data structures give the investigator a powerful advantage in the interpretation of psychophysical relationships. Not only are the physical parameters of the signals completely known, but they have been simplified to the extent that they are more easily dealt with. Nonessential physical characteristics have been removed independently, reducing the number of possible physical attributes which the investigator must consider in an interpretation of the perceptual space. The preliminary success of filtering experiments for the localization of perceptual cues for source identification, mentioned earlier for additive synthesis, suggests the value of an extended study using a wide set of signals. We plan to study both confusions in identification and perceived similarity for sets of instrument tones presented in various conditions, including filtering.

Likewise, tones generated by frequency modulation synthesis provide a means of controlling significant physical cues with a small set of parametric specifications. A rich timbre space results from the manipulation of a few basic controls which affect the dynamic evolution, bandwidth, and frequency ratios of the partials of a synthesized complex tone which has many of the characteristics of natural tones. Much insight may be gained from the perceptual scaling of such a simply-specified but multidimensionally rich space. The extension of FM into wider ranges of the timbre space, including the non-periodic music instrument tones, provides obvious advantages for scaling.

Various approaches will be taken to examine specific properties of the subjective space for timbre. Predictions for mappings of new tones will be made in order to confirm our interpretation of the physical correlations to the perceptual space. We are particularly interested in the perception of tones which the listener might not be already familiar with, such as real ethnic instruments or syntheses of tones which have no analog in the real world. The influence of familiarity with instruments and musical training on the perception of timbre will be explored.

Another interest is the continuity of the space, which centers on the nature of the regions which lie in between a mapped set of tones. Especially important is the consideration of the effects of the categorical identification of instruments on the perception of simulated tones which might physically lie in between the known tones. Equal steps along acoustical dimensions might not map as equal steps along respective perceptual dimensions, because of the influence on perception of the tendency to categorize input signals with respect to their sources of origin. We describe below independent perceptual testing for the existence of a categorical mode of perception for timbre. Multidimensional scaling will also be used to explore the space in between the known points.

graphic techniques for multidimensional configurations

Often researchers who have employed multidimensional scaling techniques have been somewhat limited by their capacity to view more than two dimensions simultaneously. In that the spatial solutions for timbre perception will most likely be interpretable in at least three dimensions, we have implemented a graphic technique based on stereoptics for the display of three dimensional configurations. It employs a split screen image and a system of mirrors to aid in the fusion of the two configurations so that the resultant image appears three dimensional. The visual cues of relative rotation of the two images, appropriate for the respective eyes to which they are reflected, and the relative brightnesses of the individual points in the configurations, appropriate to their respective distances from the viewer, are used to create a convincing three dimensional stationary image. In addition, the viewer can rotate the image about any of the three axis of the space, reflect the configuration about any axis, contract or expand the space, and display a second configuration of points in the opposite case-type of label lettering. We are planning to have some capacity for joining the points in any one configuration with vectors so to increase the cohesiveness of that one configuration and to further aid in the three dimensional coherence of the image. Further, we plan to employ various congruence optimization algorithms in aiding the viewer to associate the two different configurations which are simultaneously displayed.

We have no a priori reason to believe that timbre is only interpretable in three dimensions. In fact, researchers at Bell Labs (Carroll and Wish, 1973) have been able to interpret the scaling solutions for speech sounds in as many as six dimensions. We therefore would like to be able to explore graphical techniques for the display of higher dimensional spaces. One possibility which may prove useful in this regard involves the concept of nested three dimensional spaces.

investigation of categorical perception

As an independent test for the possible existence of cognitive constraints on the perception of simulated music instrument tones, we will employ procedures to examine the existence of a categorical mode of perception. Researchers at Haskins Labs report categorical effects for certain speech sounds. Equal steps along an acoustical dimension, interpolated between the modelled physical characteristics of two or more known speech sounds, are employed as stimuli. Listeners identify the stimuli as falling into definite categories having narrow overlap. In addition, the discriminability for pairs of stimuli, all of which are equally

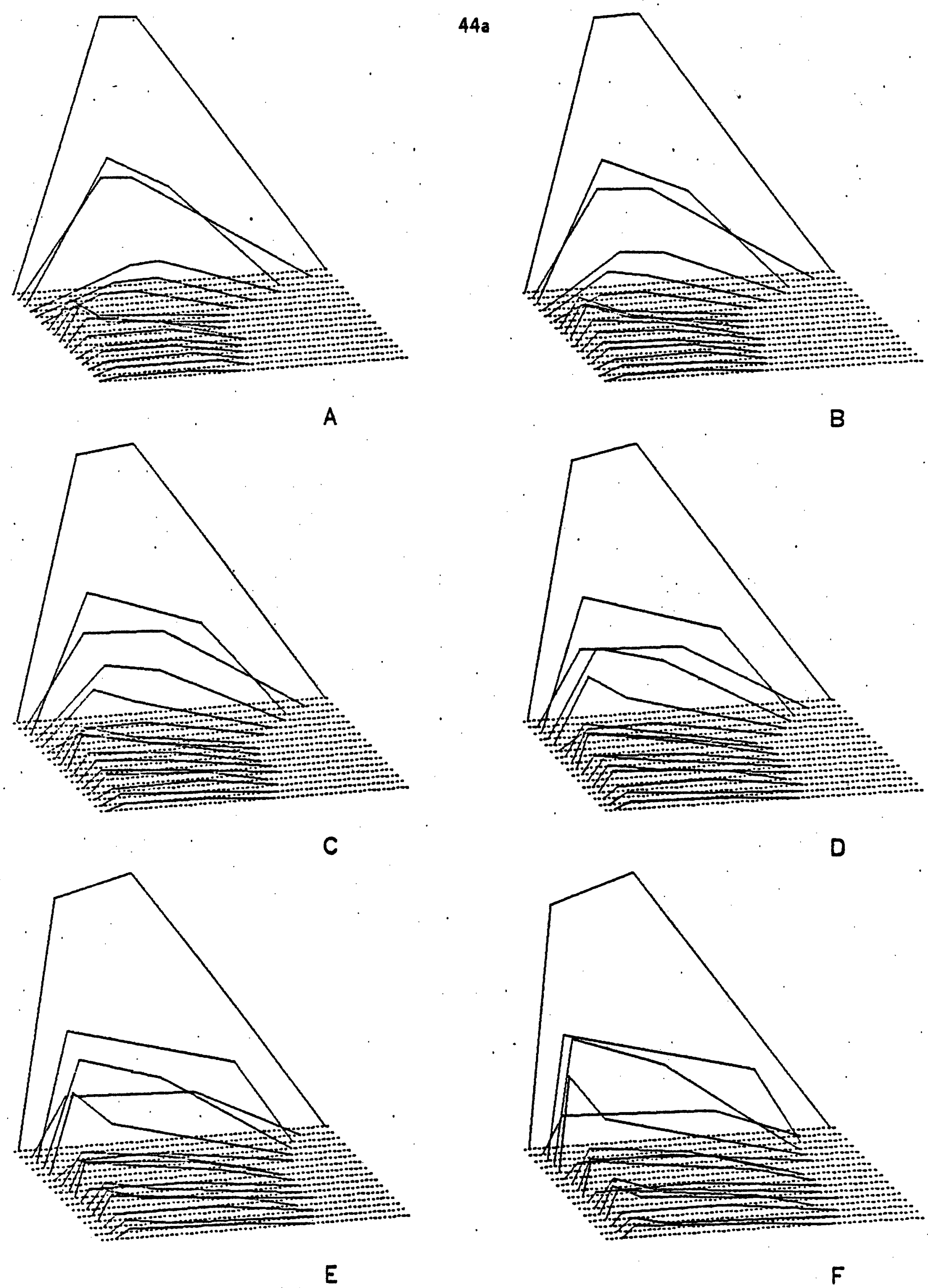


Figure 13. Interpolations (B - E) between violin (A) and saxophone (F) tones, based upon 3-line

separated along the physical dimension, is affected by their position with respect to these categorical boundaries. If both members of a pair fall within a single category, they are discriminated more poorly than if they fall within different categories. The Haskins group has used this as evidence for a very special mode of perception for speech, and the failure of other researchers to find a similar interaction between categorical identification and discrimination of stimuli is presented in support of their theory (see Liberman, et. al, 1967; Liberman, 1972).

We are interested in testing for the existence of categorical effects in the perception of another set of stimuli which are in the same sensory domain and obtain comparable complexity as the speech sounds. Primarily we are concerned with perceptual constraints on a simulation algorithm, but obvious fallout will occur for general theories of perception. Our specific procedure will be similar to the Haskins approach. Identification functions will be compared to discrimination functions for a set of stimuli consisting of tones physically interpolated between two known end-points, through a multidimensional physical space. Discrimination functions will be derived by means of the 'same-different' task discussed above.

We have been able to produce interpolations with both methods of synthesis. Additive synthesis with data-reduced physical representations of tones has given us a potential method for interpolating between known sounds, and the results have been strikingly successful. The algorithm has been based on the physical properties of tones. It consists of interpolations between the two-dimensional locations (time vs. amplitude or frequency) of joints in the three line-segment representations of the parallel functions for two known tones. An example of a set of interpolations between the violin tone, discussed above, and an alto saxophone tone at the same pitch and duration is shown in Figure 13. A particularly significant finding thus far is that the set of tones which are produced in this manner are identified as being either one or the other of the two known instruments, and that the categorical boundaries have been rather sharp. Frequency modulation synthesis provides another approach for interpolating between sounds. It gives us a set of parameters which can be systematically altered from one value to another. Two shapes of control functions, used for amplitude or index, can be interpolated between in various ways, as in the manner used for interpolating between functions explained above for additive synthesis. Successful interpolations have also been produced with this approach, including interpolations between periodic and nonperiodic sounds.

We must emphasize that we are only in the early stages of exploring the interpolation between sounds. We are presently employing algorithms based on the physical structures of tones, and will map these into a perceptual space, both using multidimensional scaling and the identification- discrimination procedure to test for categorical perception. It is clear to us from our investigation to date that we have uncovered very useful tools for the examination of properties of the perceptual space for timbre with respect to a categorical mode of perception.

automatic FM mappings from analyzed tones and the convergence of approaches

We see the automatic generation of parametric and functional data for FM simulations of instrument tones to be a significant step towards the development of a general model for simulation. The potential increase in facility and knowledge from higher-order algorithms is great. The difference in the two synthesis techniques provides us with a powerful means for evaluating hypotheses concerning the synthesis of timbres because of the difference of spectral control between the two: the additive technique allows independent control of the amplitude of each of the components in time, where the FM technique allows control of only the bandwidth. Based on the models for timbre perception which we are able to formulate with the use of multidimensional scaling, we propose to construct algorithms which map into FM data, first, the reduced data of the additive synthesis technique and second, the unreduced data from the original analysis. The difference, if any, between the analytical parts of the two mapping algorithms for the additive and FM techniques will be forced to converge where possible.

Extending the notion of mapping algorithms, and in the light of interactions which continue to occur between the two approaches to synthesis, we see the formulation of a most powerful simulation algorithm which combines the advantages of both approaches. It would enable the user to realistically simulate any known sound via the most perceptually relevant parametric specification. The power of user-control over the FM technique, in that the small number of parameters and time-functions are of such strong perceptual importance, provides a model for the optimal level of parametric specification. The degrees of freedom for the additive synthesis technique, which allows the synthesis of any arbitrary configuration of functions for the amplitudes and phases of the components of tones, provides the most powerful means of simulating any point in the physical timbre space. The course of future research is towards the convergence of the most powerful attributes of both methods, and the eventual formulation of an algorithm for simulation which truly satisfies the criteria we outlined in the introduction of section II: 1) the optimal use of computer resources, i.e. storage, efficiency, 2) the perceptual validity of the results in terms of naturalness, 3) the general applicability to the widest range of cases found in the repertoire of instrumental timbres, 4) the level of user-control of the algorithm such that parametric specifications are perceptually meaningful, 5) the efficiency with which hypotheses may be verified.

B. SIMULATION OF REVERBERANT SPACES AND LOCALIZED SOUND SOURCES

In this part of the proposal we will discuss our approaches to the computer simulation of reverberant spaces and localized sound sources. The goal of our research is the development of computer algorithms which can simulate a wide variety of natural reverberant spaces, and which are able to project arbitrary sound sources into such spaces at any localized stationary position or upon any moving path. The discussion which follows will be divided into two main topics: the simulation of realistic sounding artificial reverberation which can be controlled via perceptually meaningful parameters by the user; and efforts to maximize the area in which listeners receive convincing illusions of localized sources that are at apparent positions specified by the user. An ongoing consideration of this research is that the simulations must be accomplished with the minimal number of speaker-channels (independently controlled loudspeakers) and an optimization of computer resources.

The construction of algorithms for the simulation of reverberant spaces is aided by the quantity of research which has been performed in the field of room and architectural acoustics. From this rich field of theory Schroeder (1961) produced a model for the purpose of simulation of realistic sounding room reverberation using loudspeakers. The significance of this contribution is that it overcomes the most objectionable perceptual qualities of all previous attempts at artificial reverberation, and its computer implementation is both simple and economical. We will first describe the fundamental algorithms for the generation of reverberation, and indicate the perceptual correlates to the parameters which control these algorithms. We will then discuss the approaches which have been taken to utilize these basic algorithms in compound and multi-channel reverberation systems. Coloration of timbre, the qualitative effect of reverberation on the source signal, will be discussed as a consideration in designing complex reverberation systems. Proposed research concerns are then presented. One matter of concern is the ability to acoustically 'tune' the simulated space, using spectral shaping techniques, to increase user-control over the qualities of the resultant reverberant environment. Perceptual scaling techniques will be enlisted to determine the perceptual distinctiveness and relative importances of various features of reverberation networks, both as an aid to the development of optimal reverberation systems and as a test of the adequacy of particular systems. We will conclude this section with a discussion of our plans for the development of higher-order algorithms for the simulation of reverberant spaces, based on the above research, which give the user perceptually meaningful parametric controls.

The next section will present a discussion of our approach to the simulation of localized sound sources, projected within a reverberant space at user-specified stationary positions or paths of motion. We begin with a description of the features of a computer algorithm which we have designed for this simulation using four loudspeakers. Several perceptual cues for localization are controlled in parallel, using empirically-based functions to specify quantitative parameters of sound to the speakers. Future research includes a rigorous investigation of these functions by perceptual scaling techniques. Especially of interest are questions of optimization: maximization of the area for viable listening positions and minimization of the number of independent speaker-channels needed. The implementation of cues for localization is next discussed, including azimuth or angular displacement, distance, and altitude.

I. SIMULATION OF REVERBERANT SPACES

INTRODUCTION TO ARTIFICIAL REVERBERATION TECHNIQUES

We use three forms of the delayed feedback loop, the basic tool of artificial reverberation. The first and simplest is the comb filter, where the frequency response is periodic or comblike. Figure 14 shows a block diagram of the comb filter with its impulse response and frequency response. The second form is a simplification of the all-pass network of Schroeder (1961), so-called because, unlike the comb filter, it passes all frequencies equally well. Figure 15 shows the all-pass network and its impulse response. The third form is an oscillatory version of the all-pass which, while still passing all frequencies equally well, has an impulse response of a damped sinusoid. Figure 16 shows the oscillatory all-pass network and its impulse response.

the comb filter

The form of comb filter we use for artificial reverberation is identical to that discussed by Schroeder (1961), a delayed and attenuated feedback loop, Figure 14. The impulse response of the comb filter is a pulse train with exponentially decaying amplitude. The frequency response resembles a comb, i.e., peaks at integer multiples of the reciprocal of the delay time. This and other forms of the comb filter are described in Appendix C.

all-pass unit reverberators

A particularly useful unit of artificial reverberation is the all-pass network. We use two different forms: (1) a first-order unit of which the impulse response is a pulse train with exponentially decaying amplitude, and (2) a second-order unit of which the impulse response is a pulse train of which the amplitude is a damped sinusoid. These unit reverberators are shown in Figures 15 and 16, and described in full in Appendix D. We will introduce them briefly here.

The relationship between the input waveform to the first order unit and the resulting output waveform is formulated in the following recurrence relation:

(8) $$Y_n = GY_{n-m}+X_{n-m}-GX_n$$

This unit reverberator, as with all linear digital filters, consists of a weighted sum of delayed input samples added to a weighted sum of delayed output samples. The recurrence relation shows how the next output sample is to be computed. X_n is the n^{th} sample of the input waveform and Y_n is the n^{th} sample in the output waveform. (It is helpful to recall that the digital representation of a waveform is a series of samples, or the instantaneous numerical values of that wave at sampled points in time. The n^{th} sample is the value of the waveform at time nh, where h is the time between successive samplings.) X_{n-m} is that sample in the input waveform which occurred m samples before X_n, i.e. the value of the waveform delayed m samples with respect to X_n with delay time mh. This same delay relationship holds for the output samples Y_{n-m} and Y_n. Finally, as the gain, G, approaches unity, the impulse response of the reverberator decays more and more slowly. Essentially, this is the all-pass unit reverberator used by Schroeder (1961), except that we have realized it in the canonical form, thus saving one multiplication over the form previously used. A block diagram of this unit and a plot of a typical impulse response are shown in Figure 15.

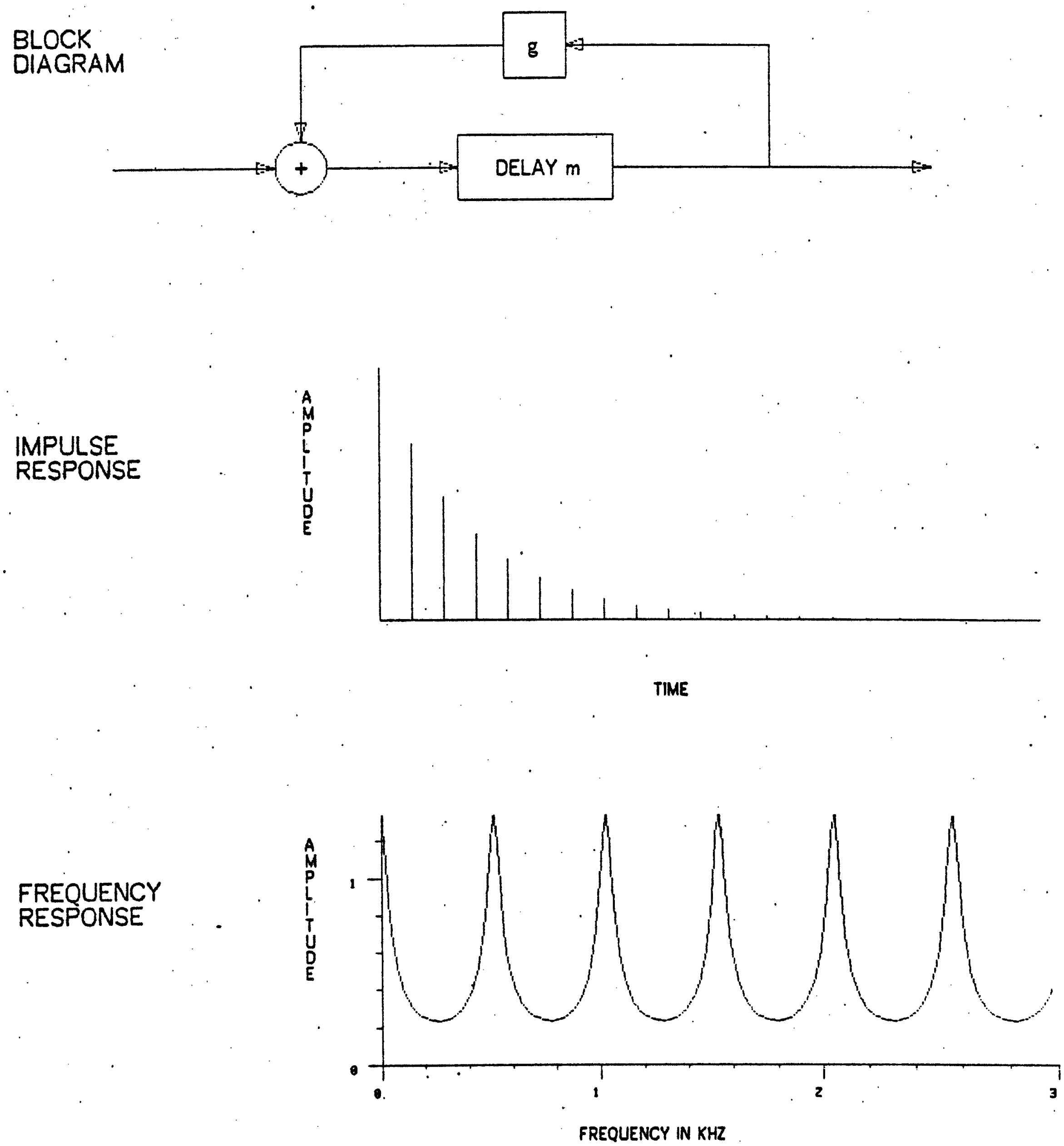


Figure 14. Comb filter with gain, G, and delay, m.

There is one important generalization of the all-pass network, the oscillatory all-pass. The recurrence relation for this second order unit is as follows:

(9) $$Y_n = G_1X_n+G_2X_{n-m}+X_{n-2m}-G_2Y_{n-m}-G_1Y_{n-2m}$$

where $G_1 = C_3/C_1$
$G_2 = C_2/C_1$
$C_1 = h^2(\omega_\theta{}^2+\sigma^2)+4\sigma h+4$
$C_2 = 2h^2(\omega_\theta{}^2+\sigma^2)-8$
$C_3 = h^2(\omega_\theta{}^2+\sigma^2)-4\sigma h+4$
ω_θ/m is the frequency of the oscillation
$7m/\sigma$ is the reverberation time, as in the first-order unit

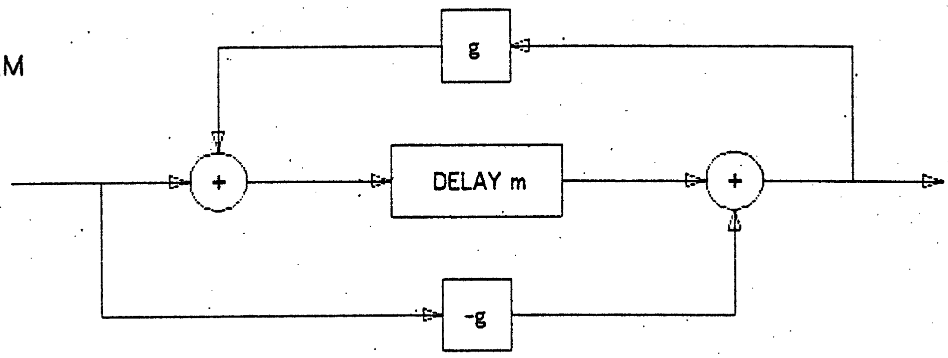
A block diagram of this unit and a plot of a typical impulse response are shown in Figure 16. At first glance, the sinusoidal property of the second-order unit might seem to be undesirable because one might perceive the frequency of the sinusoid as a spurious tone. We have found conditions, however, which suggest that this oscillatory characteristic may be used to advantage.

The parameters of both all-pass unit reverberators (equations (8) and (9)) and the comb filter are: delay time (pulse spacing), gain, and reverberation time (the time it takes for the reverberant signal to decay by 60dB, or approximately $7m/\sigma$). Only two of the three parameters are independent, i.e. the value of any one of the parameters may be expressed as a function of the other two. This construction is useful when linking unit reverberators to form a compound reverberator, discussed below. The second-order all-pass unit has as an additional parameter, the frequency of the sinusoid.

CURRENT RESEARCH

While the quality of reverberation of rooms is to some degree a matter of taste, there are some general attributes which all 'good' rooms seem to have: 1) an amplitude-frequency response which has no strong coloration, i.e. resonances or apparent periodicities, 2) an echo density which is sufficiently high that individual echos are not resolved by the ear, and 3) an echo response which is free from periodicities or flutter. Our goal, then, is to combine unit reverberators to form reverberation networks which preserve maximum similarity to the reverberation of real rooms. As with much that is described here, the strategies we use for combination of several comb filters and/or all-pass units in parallel and/or series are based on the work of Schroeder.

BLOCK DIAGRAM



IMPULSE RESPONSE

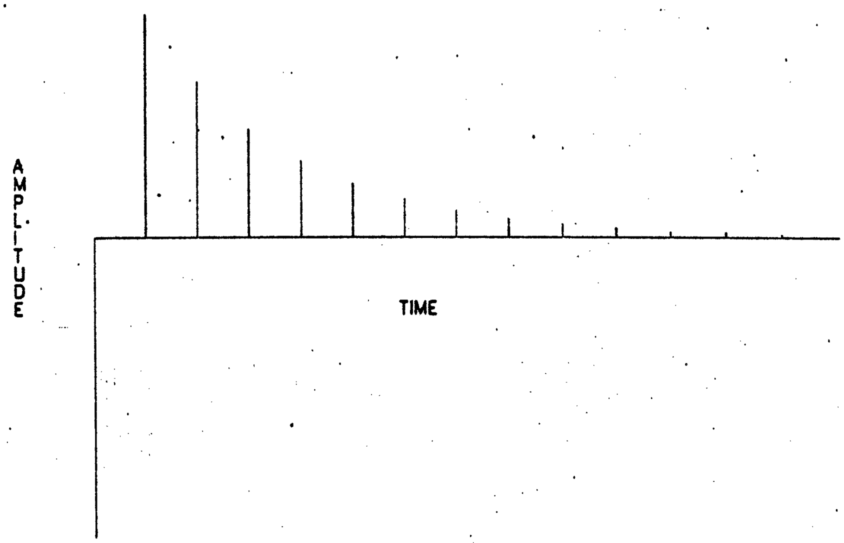


Figure 15. The first-order all-pass unit reverberator. By adding -G times the input into the output of the delay, we change a comb filter into an all-pass. Since the frequency response is unity, it is not shown here.

The most useful implementation is, perhaps, the combination of all-pass units in series. This network meets the above mentioned criteria of real rooms in the following ways: 1) the frequency response of the network is, by definition, flat, since each of of the units is flat, 2) the delay time of each of the units decreases exponentially, which increases the echo density to a point where individual echos can not be resolved by the ear, and 3) the delay times are set to be incommensurate with each other, which eliminates the possibility of periodicity or flutter in the echo response. Due to its flat frequency response, this network does not add its own 'color' to the reverberation; it is therefore called 'colorless' reverberation.

There is a fourth attribute of natural reverberation: its spatially diffuse character. In simulation techniques this is of utmost importance, since it determines the subjective impression of a real space. In order to simulate this quality it is necessary to produce uncorrelated reverberant signals from at least two speaker-channels (independently controlled loudspeakers). The obvious, and our first, implementation was to create totally independent but similar reverberation networks for each of the speaker-channels. This method has given exceptionally realistic reverberation, but it is also very expensive. A more economical method is to use a single reverberator which, at the final stage, branches to parallel (though not identical) unit reverberators. These final unit reverberators may then be connected each to one of the speaker-channels; or the output of the final unit reverberators may be mixed in differing proportion for each speaker-channel.

colorless reverberation

To form the all-pass, or 'colorless,' reverberator, several of the first-order all-pass unit reverberators are cascaded. We adjust the delay and gain of each unit reverberator until the impulse response of the compound reverberator is a smoothly decaying exponential with a perceptually relevant increase of density throughout the entire reverberation time. Thus far we have had the most success with delays and gains which decrease exponentially from unit to unit in the series. Figure 17 shows a typical impulse response as units are included in the series.

We have implemented this reverberation network in digital simulations, and the simulation is auditioned and tested via loudspeakers in real rooms. While we know that ultimately, due to the addition of signals from loudspeakers and the irregular response of the room, the signal cannot remain uncolored once it leaves the loudspeaker, it is important to be able to maintain the flat response at least as far as the loudspeaker. A coincidental resonance peak or depression between a colored

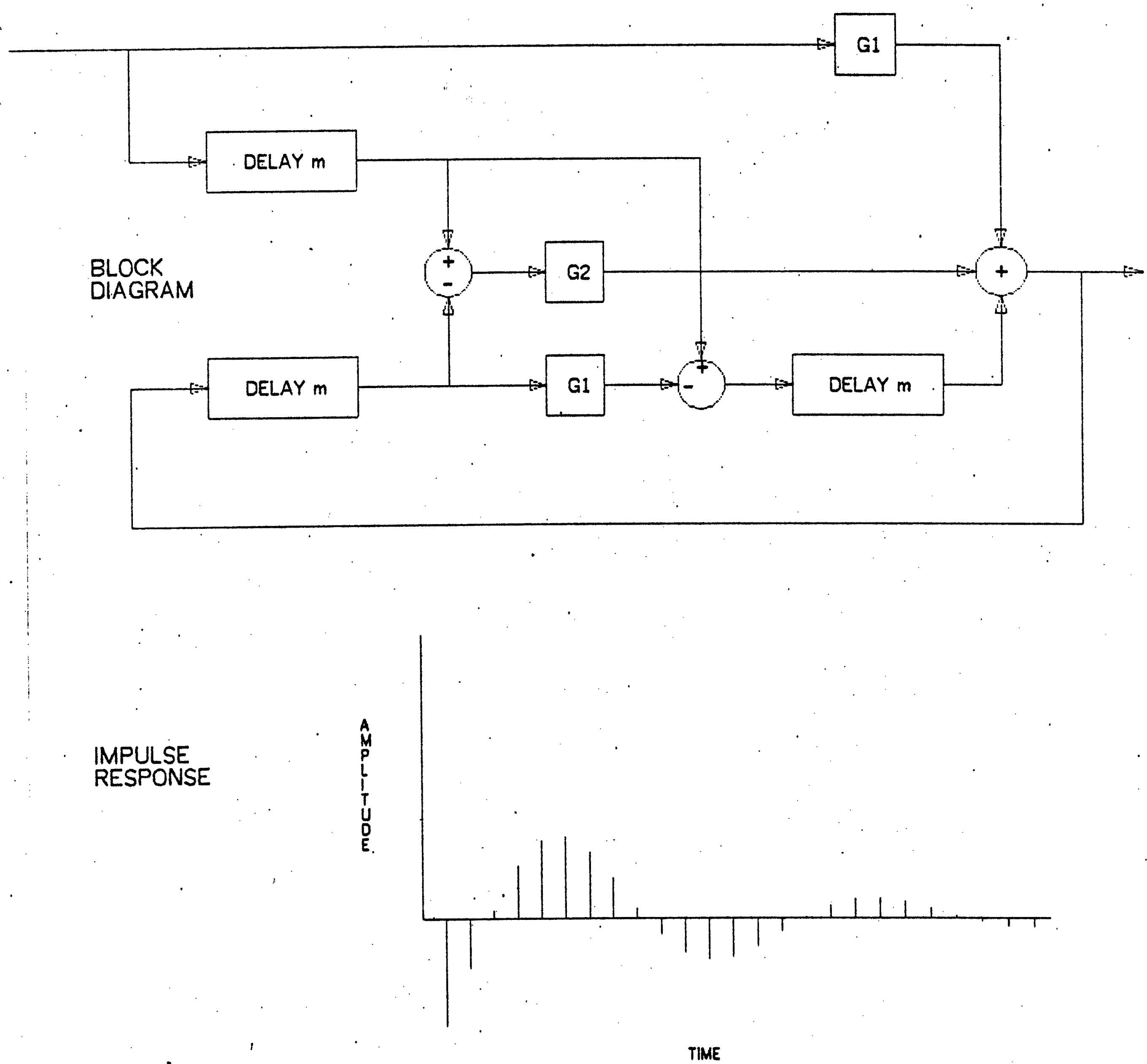


Figure 16. **The second-order (oscillatory) all-pass unit reverberator.**

reverberator and the room would tend to intensify the acoustical problems of both. There are cases, however, when one might actually want some coloration, or more likely, some specific coloration. For example, we may want to impose some special spectral shape on the output of the reverberator, and thus shape the room in which a simulated signal occurs. In this case, the resultant signal will conform exactly to the contour of the shaping filter only if the reverberator is colorless.

In an informal investigation we listened to a digitized violin phrase without reverberation, with colored reverberation, and with colorless reverberation. Although the digitized violin input is the same in all cases, the violin timbre seems more natural when comparing either of the reverberated signals to the non-reverberated signal and most natural with colorless reverberation. This suggests that the accuracy of testing timbral synthesis and data-reduction by perceptual comparisons would be improved by the use of artificial reverberation and by the condition that the reverberation is not adding its own spectral character to the sounds under consideration. Figure 18 shows the long-term discrete Fourier transform of a single digitized violin note, first without reverberation, then with colored reverberation (not extreme in this case), and finally with colorless reverberation.

Schroeder (1962) pointed out the difficulty of maintaining both uncolored reverberation and exponential decay of the pulse train when mixing the direct signal and the reverberator output. If signals are mixed by addition, then the result is no longer flat. However, flatness can be maintained by nesting the mixing process inside yet another all-pass network which, in turn, has the deficiency of producing non-exponential decay. Here, the imprecision of perception seems to be a help, for without exhaustive searching we have found delay-gain relationships for both types of mixing where deficiencies are not apparent. Although we cannot now suggest a theory for perceptually valid colorless reverberation, we are convinced that this is an area worth further investigation.

spatially diffuse reverberation

We have made extensive use of multi-channel reverberation projection to achieve a spacially diffuse quality, but using a combination of comb and all-pass unit reverberators first suggested by Schroeder. Our implementation of this reverberator, in 1967, was the first outside of Bell Telephone Laboratories.

The comb/all-pass reverberator takes the output of four parallel comb filters as the input of two or more cascaded first-order all-pass units. The delay and gain parameters follow closely the limits suggested by Schroeder. In order to avoid echo

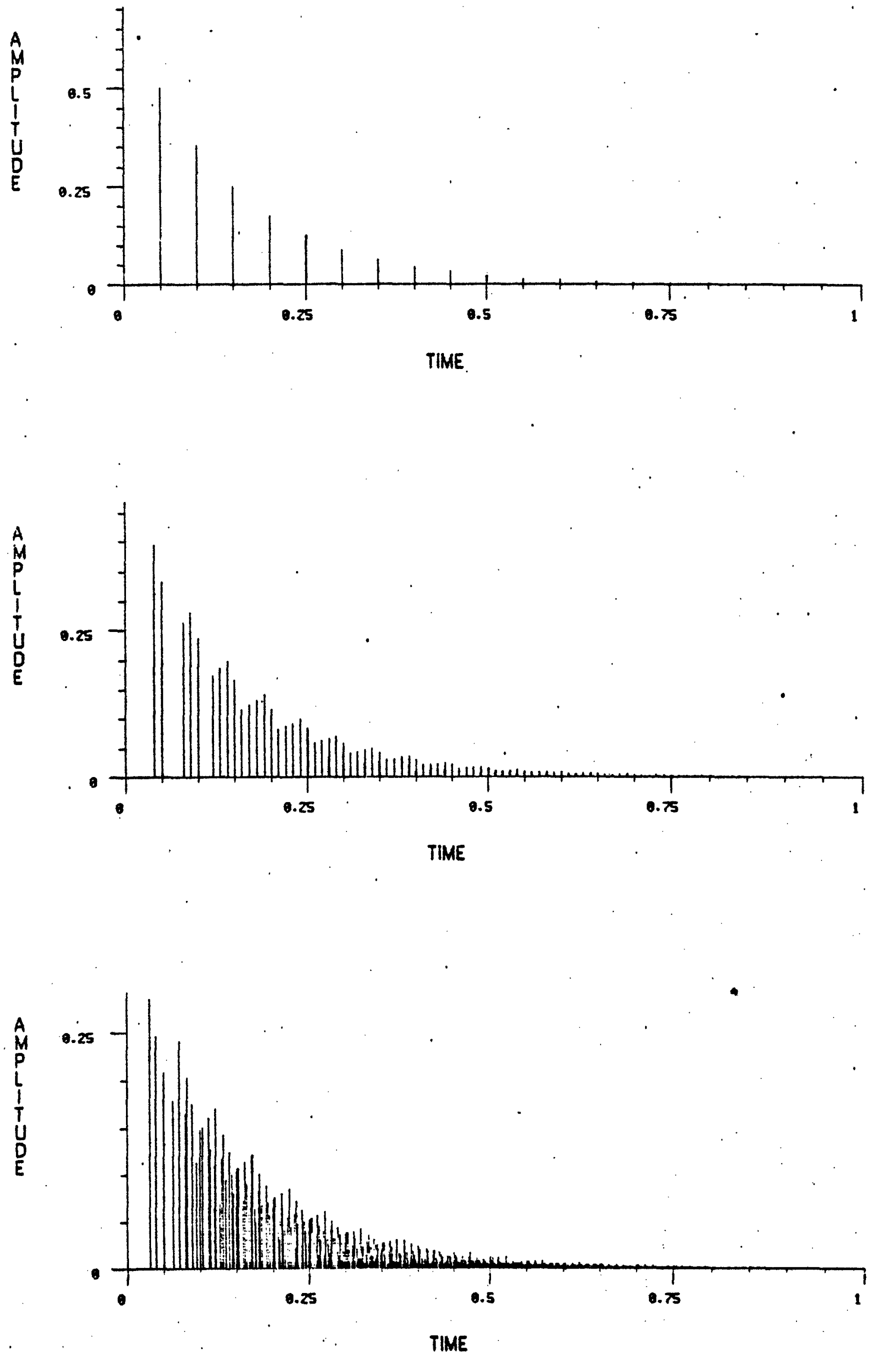


Figure 17. Absolute value of the impluse response of one, two, and then three first-order all-pass unit reverberators in series. Delays and gains are decreased exponentially from unit to unit in the series.

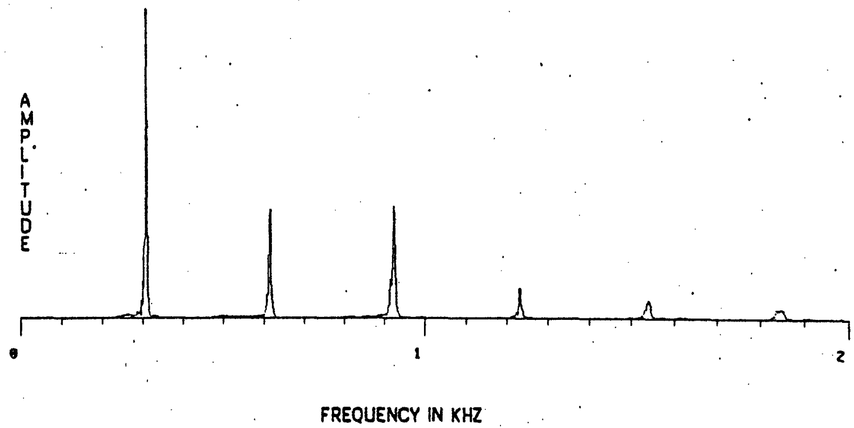
cancellation and superposition each delay is expressed as the prime number of samples which is the closest approximation of the delay time. The gains are expressed as a function of the delay time and the desired reverberation time. The time gap between the direct signal and the reverberation, the 'first delay,' is determined by the shortest comb filter delay. In creating a two-channel and then four-channel diffuse projection, the comb/all-pass reverberator was reproduced once for each channel. We used incommensurate delay ratios. The gains were expressed as a function of the delay time and the desired reverberation time. Experienced listeners have agreed that this method has given surprisingly realistic reverberation.

For better efficiency of computation, a variation of this comb/all-pass was implemented. We reverse the order of the comb and all-pass units of the comb/all-pass so that the output of three first-order all-pass units in series is taken as the input to each of four comb filters in parallel. The four comb filters are then added in various phase relationships (+ - + -, for example) with a different permutation for each loudspeaker. When the reverberation given by this smaller, more efficient, reverberator is compared to that of the large reverberator previously discussed, we have found that, for most applications, there is no loss of quality. However, the large reverberator is still used in the generation of some of the more unusual qualities of simulated rooms and localized sound sources. We have also considered an all-pass multi-channel reverberator, which will be discussed below.

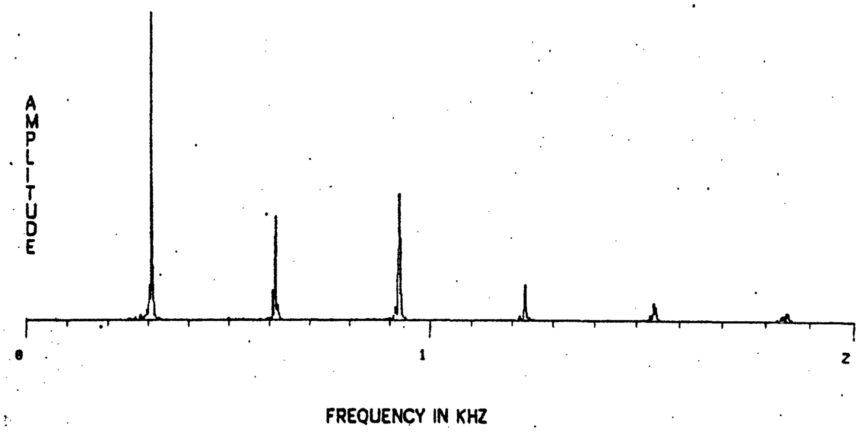
PROPOSED RESEARCH

We will here discuss the proposed research which is centrally concerned with approaches to our ultimate aim: the development of a general and easily-controlled algorithm for simulation which is based on a perceptual model for reverberant spaces. We plan to use the second-order (oscillatory) all-pass unit reverberator to implement the undulatory reverberation characteristic of good music rooms and determine whether this characteristic will intensify the realism of our simulation. In support of this enquiry, and also to create colorless and spatially diffuse reverberation, we will develop all-pass multi-channel reverberators which include second order-units. Based on analysis of multi-channel recordings of selected rooms we will explore the degree and nature of inter-channel reverberance differences, the overall spectral 'shape,' and localized resonances. We will then include the results of these studies in our simulation, which, in turn, will be verified by perceptual testing techniques similar to those already discussed. Based on these findings, we plan to develop an algorithm which can be controlled by the user via perceptually meaningful parameters such as reverberation time of the room, resonances and their location,

DIRECT SIGNAL



COMB/ALL-PASS COMBINATION



PURE ALL-PASS

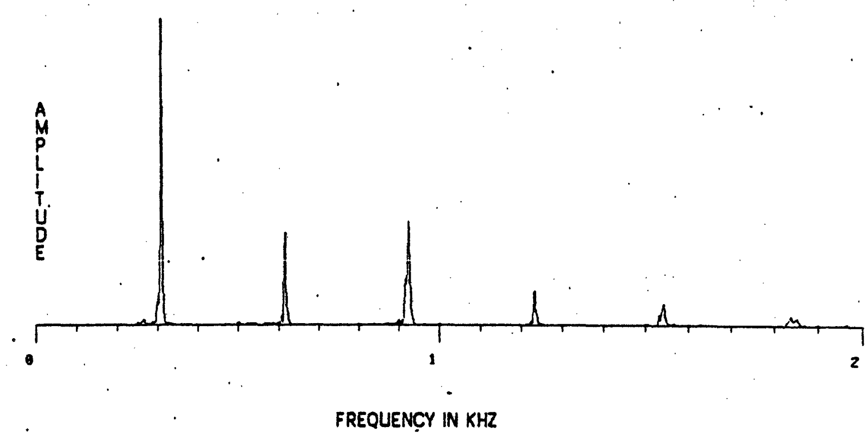


Figure 18. Long-term spectra of violin tone, reverberated with comb/all-pass reverberator, and with pure all-pass reverberator. The coloration of the comb/all-pass combination is often very slight.

reflective qualities, proximity of the walls to the listener, and the location of obstacles in the room.

applications of the second-order all-pass unit reverberator

Preliminary investigation indicates that the second order (oscillatory) unit reverberator will serve two aspects of our research. In discussing acoustical attributes of good music rooms, Knudsen (1963) characterizes the reverberation as having not only a relatively smooth decay but also a 'slightly undulating' characteristic. As mentioned above, we have produced smoothly decaying and colorless reverberation. But it is only with the development of this oscillatory unit that we can consider as an additional characteristic the undulatory quality of the reverberation. When this unit reverberator is used as a unit of a compound reverberator both the frequency and amplitude of its contribution can be controlled. This suggests that we will be able to determine whether the undulatory characteristic will intensify the realism of our simulation and also, perhaps, discover some sort of parametric range of control.

The second application of this unit is for spatially diffuse reverberation. The fact that it produces phase relationships which are of another order of complexity compared to the first-order unit suggests that it could be used as an important feature of an all-pass multi-channel reverberator. For example, the output of three or four first-order units in series can be taken as the input of a number of second-order units (one for each channel) which have incommensurate delays and incommensurate periods of oscillation. It appears that such a reverberator could produce reverberation which is as diffuse as that of the all-pass/comb variation, described above, but also colorless.

There is no intuitive way to predict the nature of the complexity and the effect of the damped oscillation when this unit reverberator is in series with other like units or with first-order units. We have found, however, that much can be learned from visual and aural observation of the impulse response of the output as the units are connected. We are therefore developing a graphic predictive analysis program which both displays and 'plays' the impulse response of any combination of unit reverberators. Such experimentation is totally dependent upon the combination of computer implementation and well trained 'ears.'

simulation of real rooms

The problem of simulating a real room is significantly more difficult than simply producing spatially diffuse reverberation, although the requirement for multi-channel output is the same. Multi-channel recordings made in real rooms preserve, to a large extent, the important localized resonances and produce the perceptual impression of a physical space. The artificial reverberation techniques have not so far produced an equally strong impression.

In order to develop our basic data, we plan to make 4-channel recordings of selected signals in a number of real rooms which fulfill the general requirements of good reverberation. Through the use of a high precision 4-channel analog to digital converter (described in section III, Research Facilities), we will digitize the signals, analyze the data for frequency and impulse response, and examine through graphic techniques the degree and nature of inter-channel differences.

We plan to begin our research in the simulation of real rooms through processing techniques applied to all-pass networks. The networks will be optimized such that four unit reverberators in series will branch to n units in parallel, where the outputs are passed to the n speaker-channels.

The value of artificial reverberation which has a flat response is that spectral shaping filters can be arbitrarily applied to the reverberator output. Two of the the filters we intend to use are the digital resonator and anti-resonator, described in Appendix C. The general synthesis algorithm is designed to allow operations to be performed on the simulated tone and the simulated reverberation independently. (In the following section on localization, the need for this control will be made clear.) Therefore, resonator and anti-resonator circuits, having different cues, can be applied to the reverberation of each of the speaker-channels according to the *localized* resonances of a real room. It was pointed out above, that spatially diffuse reverberation is dependent on uncorrelated signals at the two ears. Our preliminary investigations suggest that, in addition to an uncorrelated impulse response, localized resonances are of considerable perceptual importance to spatially diffuse reverberation.

Other filters which may be useful are the more traditional high-pass, low-pass, band-pass, and band-stop filters, as well as more special purpose filters.

perceptual scaling and testing

We project a special use for formal perceptual scaling and testing of the qualitative differences between various simulated reverberant spaces. Of central interest is a determination of the optimal algorithm for the production of naturalistic reverberation, having good spatial diffusion. The smallest total number of unit-reverberators which can be used to create such high-quality reverberation will be a major concern, since this would be an important saving in computer resources. In this regard, we are interested in the amount of apparent uncorrelation between the speaker-channels which can be produced by the smallest number of of independent unit-reverberators per channel. We will in addition differentiate the usefulness of the various types of unit-reverberators, in combinations, with respect to the above concerns.

The ability of spectral shaping filters to create convincing cues for natural environments will also be studied. We are especially interested in a test of the importance of the localized resonances which can be produced by independent shaping filters in different speaker-channels for the creation of natural sounding spaces. An evaluation of the relative efficacy of localized resonances via spectral shaping filters and uncorrelated speaker-channels via independent reverberator networks is a central concern.

higher level algorithms

Following the research in the development of algorithms for the simulation of real rooms and the evaluation of the results through the use of the scaling techniques described above, we plan to generalize this data in the form of interactive algorithms. Of principle importance will be the organization of the data required for the simulation techniques into perceptually appropriate 'visible' controls for the user.

We are currently using an interactive reverberator compiler as is an aid in exploring the technical aspects of digital artificial reverberation. The program utilizes the parameters of delay time, gain, reverberation time, and first delay. We are currently expanding this program to also display and play the impulse response of the reverberator as each unit reverberator is defined. As the next level of control, we plan to develop algorithms which will allow us to deal with problems of reverberation and the simulation of reverberant spaces in terms of their perceptual attributes. Or more specifically, we plan to move beyond considerations of the characteristics of a particular complex reverberator and describe the room we wish to simulate in terms such as reverberation time of the room, resonances and their location, reflective qualities, proximity of the walls to the listener, and the location of obstacles in the room.

2. SIMULATION OF LOCALIZED SOUND SOURCES

CURRENT RESEARCH

The simulation of reverberant spaces, as described above, has led to consequent research into the localization cues for a simulated source within the space. The control of the reverberant signals emanating from two or more loudspeakers allows the simulation of a space which is largely independent of the actual reflecting surfaces of the space in which the loudspeakers are located. The natural consequence of this illusion, then, is the arbitrary placement of a signal at some point in the illusory space which may, in fact, be beyond the walls of the physical space.

Our interest is in the further development of techniques which require the minimum number of loudspeaker-channels to provide the perceptual cues for azimuth, distance, and altitude in a natural environment. There have been a number of studies into the perception of location of a sound source (see Mills, 1972, for a comprehensive review of the psychoacoustic literature on auditory localization). Most findings are based on unnatural listening conditions where constraints are placed on the environment or the listener or both: a large number of loudspeakers, anechoic chamber, headphones, or fixed position of the listener. The conditions we assume in our research are few speaker channels, a room having a moderate reverberation time, a space which will accomodate a number of listeners, and no headphones.

We first describe the simulation of the cues for azimuth, distance, and velocity as currently implemented in our system. These cues are demonstrated on sound Example 5.

simulation of azimuth and distance cues

The ability of a listener to localize a sound source is dependent upon cues for three dimensions: azimuth or horizontal angular displacement, distance, and altitude or vertical angular displacement. Of the three the last is the least critical in our ordinary (horizontal) environment and is the least discriminated.

The cues for angular localization are: 1) the different arrival times of the signal to the two ears when the source is not centered in front of or behind the listener, 2) the pressure-level difference of short wavelength signals at the two ears resulting from the shadow effect of the head when the signal is not centered, 3) cues, currently not well-understood, provided by the asymmetricality of the pinna for very high frequency energy.

The cues for the distance of a source from a listener, when the distance is greater than a few feet, are: 1) the ratio of the direct signal energy to the reverberant signal energy where the direct signal decreases in intensity inversely according to the square of the distance, 2) the loss of high frequency components with increasing distance between the signal source and the listener, and 3) the loss of detail in the signal with increasing distance. Low intensity components are lost.

At the present, our simulation system consists of four loudspeakers which are independently driven by the digital to analog converter output of the computer. For the purpose of localization simulation, the speaker-channels are arranged in a square and the listener is assumed to be at the center, equi-distant from the four loudspeakers as shown in Figure 1, app. F. Adjacent speaker pairs form an angle of 90 degrees relative to the listener.

In order to simulate the azimuthal cue, we distribute the sound to one adjacent pair of speakers at a time. When simulating a source directly behind one speaker, all of the signal comes out of that speaker. As the source moves from one speaker to another, we decrease the amount of signal in the speaker the source is leaving and increase the amount of signal in the speaker the source is approaching. The particular functions we use to distribute the sound are Θ/Θ_{max} for one speaker and $(1-\Theta/\Theta_{max})$ for the other where Θ is the angle of displacement and Θ_{max} is equal to 90 degrees. Because the computations are based on the ideal position of the listener, there is a positional distortion of the phantom image for any other listener. This distortion is constant for any position and has not been found to be objectionable unless a listener is very close to one of the loudspeakers.

The simulation of the distance cue is dependent upon the availability of artificial reverberation. For the simplest case, the reverberation is set to be constant in intensity and the direct signal is scaled to be inversely proportional to the distance in question. For simplicity, he unit distance, 1, is assigned to be the point which divides the line joining two adjacent speakers and is, therefore, the nearest point which can be simulated. With increasing distance, the reverberant signal remains constant while the direct signal decreases in intensity, thereby meeting the principle requirement for the distance cue: a change in ratio of direct to reverberant energy.

The first extension of this technique involves a scaling for the reverberant signal with distance as well. The reverberant signal is attenuated according to a function which decreases less rapidly with distance, for example, in inverse proportion to the square root of the distance. As the source moves away, the total sound from the source, including reverberation, will decrease.

There is another important detail in our current technique for the simulation of the distance cue. At distances beyond the echo radius (that distance where the intensities of the direct and reverberant signals are equal) the direct signal becomes masked by the reverberant signal, thereby eliminating the azimuthal information. In order to overcome this deficiency we divide the reverberant signal into two parts: 1) Global, which is distributed equally to all channels, but which is now attenuated with distance of the direct signal according to 1/distance; 2) Local, which is distributed between speaker pairs with the direct signal and is increased with distance of the direct signal according to 1-(1/distance). Thus, when the source is close to the listener, the reverberant signal is equally distributed in all channels and as the source moves away, the reverberant signal becomes concentrated in the direction of the source.

Note that this localization of reverberation is in addition to the scaling for distance, which is inversely proportional to the square root of the distance. The reverberation, thus, has two attenuation factors. One factor which is equal in all speakers, and one factor which favors one pair of speakers.

The direct signal itself has also two attenuation factors. One which attenuates the signal with distance, and one which distributes the direct signal to two adjacent speakers.

moving sources and velocity cue

The localization technique has been extended to include the simulation of a moving sound source. This capability is a very powerful potential tool for tuning the simulation algorithms for localization cues.

A special program has been written which allows the user to specify an arbitrary path in a two-dimensional space by means of a light pen or a computed geometry. The program evaluates the trajectory and then derives the time functions which control distance and angle for the simulation. There is an additional component which is present in the case of a moving source and which is derived by the program: Doppler shift of the frequency as a result of the radial velocity of the source relative to the listener. This frequency shift has been found to be an essential cue in the simulation of moving sources. The program then applies these computed functions to the synthesized signal which can result in convincing illusions of spatial movement for the listener.

PROPOSED RESEARCH

The programs and techniques briefly described above and more fully in Chowning (1971), have proven to be sufficiently powerful to convince us that they include all of the significant cues for localization and that the method of control is valid. There is, nevertheless, a great amount of research to be done relative to the internal algorithms for the individual cues and the optimum number and relative positions of the speaker-channels. In their current state, they represent the obvious approximations to the natural physical processes. It may very well be, however, that special distortions of the space and/or amplitude relationships may provide a significant enhancement of the localization cues when projected through a few loudspeakers. In this section we propose further research by examining each of the cues independently.

azimuth

We have found the cue for azimuth to be the most problematical in simulations using four loudspeakers. The energy is distributed between speakers to provide a phantom source for the listener at the center of the space circumscribed by the loudspeakers. The centrally positioned listener can perceive a spatial distribution of 360 degrees. For positions of increasing distance from the center, the closest loudspeaker increasingly dominates the phantom source, until the *worst case*, a position next to a loudspeaker, where the listener perceives a spatial distribution of only 90 degrees. The perceived space, then, decreases monotonically as the listener position is further from the center, but the exact nature of the function is as yet unknown.

With five loudspeakers arranged in a circle, the *worst case* position allows a perceived space of 108 degrees, an increase of 20 per cent, six speakers 120 degrees, eight speakers 135 degrees, etc. There are clearly diminishing returns with additional speaker-channels.

It is in our planned research to thoroughly investigate the area of the listener space which is circumscribed by the maximum acceptable spatial distortion for off-center positions within that space and how that useable area increases with the addition of speaker-channels. Of particular importance is the cost-effectivness for n speaker-channels.

The problem of optimum number and placement of speaker-channels relates also to the simulated reverberant space, described in section IIB1, above. The perceptual criteria for realistic reverberation and azimuth cues are different, however. For reverberation, the perceptual impression of diffusion or unlocalized source is desired, whereas, for the the azimuth cue of a source, the utmost localization is desired. It is probable that the number of speaker-channels required for diffuse reverberation is less than the number required for azimuthal localization. We will consider, therefore, optimizations in our simulation algorithms to maintain the minimum number of uncorrelated reverberant speaker-channels independent of the number required for the most effective simulation of azimuth.

In the simulation algorithm we distribute the energy between speaker pairs in proportion to the angle of displacement. It may be that other functions can be used to better "fill the hole" between speakers, e.g. the tangent of the angle of displacement. We plan to apply, here, the multi-dimensional scaling techniques to subjective evaluations of various modifications to the algorithm.

distance

The validity of the distance cue in our simulations is the most convincing because it is independent of both the number of speaker-channels and the position of the listener within the space. Our algorithm, as stated previously, is based upon the attenuation of the intensity of the signal in inverse proportion to the square of the distance. This relation for the direct signal seems to be inviolate and effective.

The attenuation of the reverberant signal with increasing distance of the direct signal does not suggest an equivalent absolute relationship as does the direct signal. In a small space, the overall intensity of the reverberant signal changes little, whereas in a large space the change may be significant. In order to evaluate the perceptual significance of this changing reverberation, we propose to make four-channel recordings of signals at a variety of distances in a variety of spaces and analyze these signals by means of computer analysis techniques. Evaluation of a number of cases should provide us insight into the amount and perceptual effect of the change in the reverberant signal as a source moves away from the listener. It could well be that the change in the amount is not as great as the change in the directional emphasis, since, in moving away from the listener, the source is moving toward some reflecting surfaces and away from others.

Our experience has shown that the distance cue becomes obscure when the total amount of reverberant energy is great and the simulated source is not in immediate proximity to the listener. In addition, it is obvious that with no artificial reverberation, any attenuation for distance imposed on a source signal will be perceived only as a difference in loudness and not distance. Since the only reverberation is that which is natural to the room, it will change in intensity in constant proportion to the change in intensity of the source signal. We propose to determine the maximum and minimum ratios of reverberant to direct signals which will fully preserve the distance cue.

location of source as indication of room size

In addition to the room information described in section IIB1 above, the localization cues of a simulated source can also provide information. The cue for distance, for example, can be either in contradiction to, or confirmation of, room information carried in the reverberation itself. As an example of contradictory cues, very short first delays could be projected from the listener's left, indicating a near reflecting surface, while from the same direction, the direct signal cue for distance indicates far. This implicit power of digital synthesis to precisely and independently control cues, we see as being of enormous usefulness in establishing their potency and relative dominance.

altitude

It is apparent that the most powerful realization of simulated localization and reverberant spaces must also include cues for altitude. Localization studies on the vertical plane show that the processing of the pinnae of the ear of frequencies greater than 7000 Hz provides the critical cue. In addition, there are large spaces, most notably cathedrals, where the vertical reverberant component makes a clear contribution to the subjective impression of the space. We propose to investigate, therefore, arrangements of speaker-channels which most effectively and efficiently provide such an impression. An obvious arrangement of four speaker-channels would be in a tetrahedron. For an ideally located listener, four may suffice, however, in order to accomodate a larger listener space, a larger number is most probably required.

perceptual scaling and testing

We are planning to employ psychological techniques for the determination of the relative weightings of the various cues for localization and the perception of the motion of sound sources. We hope to obtain more detailed knowledge of the subjective scaling of several cues: the optimal amount of Doppler shift needed to create specific trajectory images; the perceptual scale which describes the relationship between apparent distance of the source and the ratio of direct to reverberant signal; the minimum amount of reverberation required to detect the distance cue; the effect of first delays as opposed to distance of a source as a subjective determinant of apparent room size. Of particular importance in this research, will be the rigorous evaluation on the basis of subjective measurement, of the optimum number and relative positions of speaker-channels for clear angular position. Multidimensional scaling techniques, described in Appendix B, may be of use in this regard - the spatial model for perceived relationships seems especially applicable.

III. APPENDICES

APPENDIX A: THE HETERODYNE FILTER

As was mentioned before, in additive synthesis we physically model a complex sound waveform as a sum of sinusoids with slowly time-varying amplitudes and phases. The process of synthesis involves specifying the amplitude and phase for each component sinusoid as it varies with time throughout the duration of the tone. These sinusoids are added together to produce the complex waveform. Equation (1) is duplicated here as equation (A1) and summarizes this formulation. The sinusoids in equation (A1) represent the components of the complex tone.

$$\text{(A1)}\quad F_\alpha = \sum_{n=1}^{M} A_n \sin(\omega_n \alpha h + \theta_n)$$

Notation:
- h is the time between consecutive samples
- α is the sample number
- F_α is the sampled, digitized waveform at time αh
- A_n is the amplitude of the nth partial tone and is assumed to be slowly varying with time
- θ_n is the phase of the nth partial tone and is assumed to be slowly varying with time
- ω_n is the radian frequency of the nth partial tone

To be more explicit, we should point out again that A_n and θ_n are functions of time. We will indicate this in the following text by appending the subscript α to each of these functions:

$$A_n = A_{n\alpha} \quad \text{and} \quad \theta_n = \theta_{n\alpha}$$

This brings out the time dependence more clearly.

It should be pointed out that this is not a Fourier series representation despite its outward similarity. The Fourier series approximates a periodic function with a sum of orthogonal sinusoids with fixed frequencies and phases and (possibly) exponential amplitudes. The model described in equation (A1) is not necessarily periodic, has time varying phase, and has time varying amplitudes that are not necessarily exponential. The individual components are certainly not orthogonal in the general case.

We have avoided use of continuous analysis and have defined all our functions in the discrete domain. This is done because the ultimate realization of these processes is on a digital computer. Inaccuracies can result from doing the mathematics in the continuous case and assuming it can be converted to the discrete domain merely by sampling. The conversion to the discrete domain must be done with some care. Since discrete mathematics and the science of digital signal processing have become so well developed, there is little reason to define our processes in the continuous domain, only to subsequently realize them in the discrete domain.

We have borrowed the notation of numerical analysis for functions sampled at equally spaced intervals by placing the sample number as a subscript. This is done partly for notational convenience and partly because in the computer, sampled functions are represented by matrices.

We seek a method of determining the functions $A_{n\alpha}$ and $\theta_{n\alpha}$ of a tone from a musical instrument, we can then synthesize an approximation to the waveform F_α from those functions by use of equation (A1).

We write equation (A1) using sinusoids at time-varying phase angles rather than as sine and cosine quadrature components because it is more efficient to synthesize the waveform in the form shown in equation (A1).

Let us turn to the problem of determining the functions $A_{n\alpha}$ and $\theta_{n\alpha}$ of a music instrument tone. To aid the analysis, we must assume the frequencies of the partial tones, ω_n, are nearly harmonically related. That is, there is some frequency, ω, such that ω_n is approximately $n\omega$. We shall call this frequency ω the fundamental frequency of the tone.

Basically, the method is as follows: First, compute the following two summations at each point in time α.

$$\text{(A2)} \quad a_{n\alpha} = \sum_{i=\alpha}^{\alpha+N-1} F_i \sin(n\omega_0 ih+\phi_0)$$

$$\text{(A3)} \quad b_{n\alpha} = \sum_{i=\alpha}^{\alpha+N-1} F_i \cos(n\omega_0 ih+\phi_0)$$

The initial phase angle, ϕ_0 is included for generality. It will be shown below that the method is independent of the initial phase angle. From these, we calculate two more sequences:

(A4) $A_{n\alpha} = (a_{n\alpha}^2 + b_{n\alpha}^2)^{1/2}$

(A5) $\theta_{n\alpha} = \mathrm{atan}(a_{n\alpha}/b_{n\alpha})$

The summations are taken to be over one period of a sinusoid of frequency ω_θ, that is, $Nh\omega_\theta = 2\pi$. This places somewhat of a restriction on the frequency of analysis, ω_θ, because in the discrete domain, the period, N, is restricted to integral values. This has not proved a problem in our experience.

If the partial tones are nearly harmonically related, if the amplitude and phase functions of the tone vary slowly with time, and if ω_θ is nearly equal to the fundamental frequency of the tone, then $A_{n\alpha}$ and $\theta_{n\alpha}$, as computed by equations (A2) through (A5), will indeed be approximations to the actual amplitudes and phases of the partials of the tone under analysis.

To review, we do the computations indicated in equations (A2), (A3), (A4), and (A5) for each of the partials of a tone, over the entire time interval spanned by the tone. The output $A_{n\alpha}$ and $\theta_{n\alpha}$ may then be used in equation (A1) to synthesize a tone based on the analysis.

Now let us compute the response of the heterodyne filter as defined by equations (A2), (A3), (A4), and (A5) to a sinusoid of constant amplitude and phase. We do this by substituting for F_i in equations (A2) and (A3) the function $\sin(\omega ih)$. We may compute the summations without error by use of the summation calculus (Hamming). Using the fact that $Nh\omega_\theta = 2\pi$, we may calculate $A_{n\alpha}$ and $\theta_{n\alpha}$ explicitly.

$$(A6)\quad A_{n\alpha} = \frac{1}{4N^2}\sin^2(\omega Nh/2)\left\{\frac{1}{\sin^2[(\omega+n\omega_\theta)h/2]} + \frac{1}{\sin^2[(\omega-n\omega_\theta)h/2]}\right.$$

$$\left. + \frac{2\cos[n\omega_\theta h - 2\phi_\theta]}{\sin[(\omega+n\omega_\theta)h/2]\ \sin[(\omega-n\omega_\theta)h/2]}\right\}$$

The expression for $\theta_{n\alpha}$ is a bit long and is thus not included here. As we consider the limit as ω approaches $n\omega_\theta$, we find great simplification of the results.

Let us define $\Delta\omega$ as $(\omega - n\omega_0)$.

$$\text{(A7)} \quad \lim_{\omega\to\omega_0} A_{n\alpha} = \frac{1}{4N^2}\{0 + N^2 + 0\} = 1/4$$

$$\text{(A8)} \quad \lim_{\omega\to\omega_0} a_{n\alpha}/b_{n\alpha} = \frac{\sin\{2\omega_0 h[(N-1)/2+\alpha]\} + N\sin\{\Delta\omega h[(N-1)/2+\alpha]\}}{\cos\{2\omega_0 h[(N-1)/2+\alpha]\} + N\cos\{\Delta\omega h[(N-1)/2+\alpha]\}}$$

If N>>1 then (A8) reduces greatly.

$$\text{(A9)} \quad \lim_{\omega\to\omega_0} a_{n\alpha}/b_{n\alpha} = \tan\{\Delta\omega h[(N-1)/2+\alpha]\}$$

Thus we see that in the limit, $A_{n\alpha}$ approaches one quarter of the amplitude of the input sinusoid and $\theta_{n\alpha}$ is a term related to the difference in frequency of the input sinusoid with the analysis frequency. With instruments with partials whose frequencies deviate from the ideal, this provides a dynamic estimation of those frequencies.

Notice that neither $A_{n\alpha}$ nor $\theta_{n\alpha}$ are functions of ϕ_0, the initial phase angle. Under the assumptions stated, this process is independent of initial phase.

Notice also that $A_{n\alpha}$ is no longer a function of α, the time parameter, but $\theta_{n\alpha}$ is. Fortunately, $\theta_{n\alpha}$ is a linear function of α whose slope is simply $\Delta\omega h$.

Figure A1 shows a plot of $A_{n\alpha}$ as indicated in equation (A1) for a range of values of ω. In this case, ω_0 is $2\pi(125 \text{ Hz})$ and n is 4. We see that there is a zero of transmission at all integral multiples of ω_0 except the n^{th} multiple.

This technique is useful as long as the amplitudes and phases of the partials of the input waveform change slowly with time. If the frequencies of the partials deviate from integral multiples of the fundamental by too great an amount, further error may be introduced.

One must remember that the heterodyne filter as defined here is a *nonlinear* filter.

Although we derived the response to a sinusoid, superposition does not hold in such a filter. The calculation of $a_{n\alpha}$ and $b_{n\alpha}$, however, is quite linear and superposition again applies. The general statements made above still hold, but only because of the special nature of the input signal.

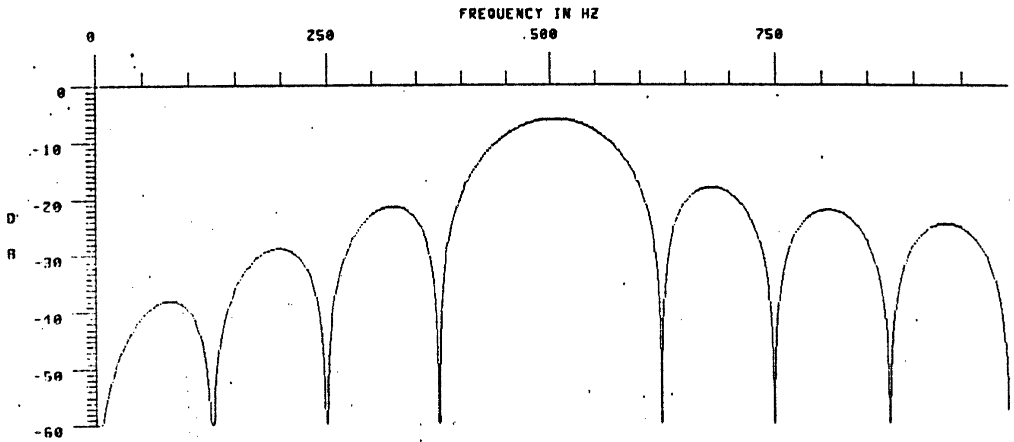


Figure A1. Log magnitude frequency response of the heterodyne filter for a base frequency of 125 Hz and j=4. There is a zero of transmission at every multiple of 125 Hz except the fourth one.

APPENDIX B: MULTIDIMENSIONAL SCALING TECHNIQUES

SPATIAL MODEL FOR PERCEPTUAL RELATIONSHIPS

It has been found most useful to employ a spatial model to represent the judged relationships between sets of stimuli, such as auditory signals. Computer-based multidimensional scaling algorithms have been developed for the reduction of the very complex data obtained from the subjective evaluations of perceived relationships between all members in a set of stimuli. The result is displayed in a form which is much more easily comprehended and interpreted by the investigator, that of a geometric configuration of points which represent the individual stimuli. The structure of the subjective evaluations of the set of stimuli is then mapped into an n-dimensional space, where the distances between the points are determined by some measure of the psychological distance between all pairs of stimuli.

The psychological measures which can be mapped into a spatial model in terms of distance include the confusability or the judged similarity of stimuli. One experimental procedure involves the identification of individual signals, perhaps presented under different conditions, and the result is a square confusion matrix of stimulus by response. A transform of this matrix yields the relative psychological distances, directly related to confusability, of all points in the data matrix. Another experimental procedure involves the rating of relative similarity for a pair of stimuli in the set. The similarity ratings can be placed in another square matrix, where each entry point is for one particular of all the possible ordered pairs of stimuli. The psychological distance in this case is inversely related to the similarity of stimuli, and is directly related to their dissimilarity.

The common characteristic of the scaling programs we find useful is their generation of an empirically-based representation of the relationships between the stimuli, rather than some theoretically imposed, a priori representation. We proceed from the perceptual data and will compare the representation of this data to the known physical attributes of the stimuli. The uncovering of psychophysical relationships, that is, making correlations between the subjective, psychological judgments of the stimuli and their physical properties, is essentially a matter of interpreting the representation of the perceptual data in terms of the known physical data for the stimuli. Note the significant difference of this from most experimental approaches, which begin with some a priori notions of the nature of the results, and the experiment is constructed and data analyzed around these notions. The multidimensional scaling approach is useful when the stimuli are inherently so complex that we have no a priori notions and are willing to rely totally on the empirical method of analysis. We are concerned with both the dimensionality and the general properties of the space. Correlations with physical parameters are sought. Various programs exist which are useful in assessing the correspondence of data structures, so that it would prove fruitful to formally represent the physical data.

MULTIDIMENSIONAL SCALING ALGORITHMS

We will briefly describe the two programs for multidimensional scaling found most useful in our research, MDSCAL and INDSCAL. These programs both attempt to represent input data matrices in the form of a configuration of points located in an n-dimensional geometric space, where n, the number of dimensions, is specifiable by the user. The points correspond to the stimuli, whose psychological distances are given in the input matrices. The coordinates of the points are obtained by an iterative computational algorithm which optimizes the correspondence of interpoint distances in the spatial representation to the measured psychological distances between the stimuli.

MDSCAL peforms a non-metric multidimensional scaling. The optimal spatial representation of a subjective response matrix is one in which the rank order of the values of psychological distance be the same as the rank order of the interpoint distances in the n-dimensional geometric configuration. A monotonic function maps psychological values into distances in the spatial representation (for a discussion of the theory and procedures of this algorithm see Shepard, 1962a, 1962b; Kruskal 1964a, 1964b). The use of MDSCAL in conjunction with other programs which deal with psychological distance matrices can be particularly informative. One such program, HICLUS (Johnson, 1967) produces a tree-structure which represents the hierarchical clustering relationships of stimuli in the matrix as inferred from their psychological distances (the use of HICLUS with MDSCAL for the analysis of confusion matrices for speech signals is demonstrated by Shepard, 1972).

INDSCAL is a metric multidimensional scaling program, which was developed to utilize the individual differences in sets of response matrices for analysis. It generates a single n-dimensional representation for the complete set of matrices. It analyzes the variations in the set of individual data matrices to uniquely determine a rotation for the axes in the space. Also produced by this analysis is a representation of weightings which account for individual response variations along the spatial dimensions. The weightings are mapped into an n-dimensional spatial configuration which can be used to assess the relationships between individual subjects or experimental conditions (see Carroll, 1970, for a complete discussion of the theory and operations of this technique; and Carroll and Wish, 1973, for an application of the method for the representation of confusion matrices for speech signals).

APPENDIX C: SPECTRAL SHAPING FILTERS

We use several basic filters. We shall describe only the ones that are different from the more common low-pass and high-pass networks, as these are documented extensively in the literature.

SIMPLE RESONATORS AND ANTI-RESONATORS

Let us begin with the digital resonator. This is most simply expressed as follows:

$$\text{(C1)} \qquad T(Z) = \frac{1-qZ^{-1}}{1-2r\cos(\omega_0 h)Z^{-1}+r^2Z^{-2}}$$

With recurrence relation

$$\text{(C2)} \qquad Y_n = X_n - qX_{n-1} + 2r\cos(\omega_0 h)Y_{n-1} - r^2Y_{n-2}$$

This has a resonant frequency of ω_0. r determines the Q of the resonator. q is a zero of transmission. This resonator is used extensively by Gold and Rader (1969), and thus will not be examined further here.

Another filter that is useful is the anti-resonance, or notch filter. In the continuous case, the transfer function is given by:

$$\text{(C3)} \qquad H(S) = \frac{S^2+2\delta S+\delta^2+\omega_0^2}{S^2+2\sigma S+\sigma^2+\omega_0^2}$$

To prevent warping of the frequency axis, the Boxer-Thaler (1956) transformation is used :

$$\text{(C4)} \qquad T(Z) = \frac{C_1+C_2Z^{-1}+C_3Z^{-2}}{C_4+C_5Z^{-1}+C_6Z^{-2}}$$

With its related recurrence formula:

$$\text{(C5)} \quad Y_n = (C_1X_n+C_2X_{n-1}+C_3X_{n-2}-C_5Y_{n-1}-C_6Y_{n-2})/C_4$$

where
$$\begin{aligned} C_1 &= h^2(\delta^2+\omega_0^2)+12h\delta+12 \\ C_2 &= 10h^2(\delta^2+\omega_0^2)-24 \\ C_3 &= h^2(\delta^2+\omega_0^2)-12h\delta+12 \\ C_4 &= h^2(\sigma^2+\omega_0^2)+12h\sigma+12 \\ C_5 &= 10h^2(\sigma^2+\omega_0^2)-24 \\ C_6 &= h^2(\sigma^2+\omega_0^2)-12h\sigma+12 \end{aligned}$$

This gives a magnitude-frequency response that is unity everywhere except near ω_0. The strength at ω_0 is determined by the values of σ and δ. If both σ and δ are small compared to ω_0, then the magnitude of H(S) at $S=j\omega_0$ will be approximately δ/σ. If δ is zero, the magnitude goes to identically zero at ω_0. If $\delta=\sigma$, then we have an all-pass network like the one used in appendix B as the second order reverberator.

The magnitude-frequency response of the filter in equation (C4) is plotted in figure C1 for five different values of δ/σ. The value of ω_0 was 2π(500 Hz), and σ was 2π(50 Hz). The ratios δ/σ were 4, 2, 1, .5, and .25.

THE COMB FILTERS

The comb filter comes in four forms, two of which are all-zero filters and two of which are all-pole filters. The two all-zero filters have the following recurrence relations:

$$\text{(C6)} \quad Y_n = X_n+X_{n-m}$$

$$\text{(C7)} \quad Y_n = X_n-X_{n-m}$$

These have magnitude frequency responses as follows:

$$\text{(C8)} \quad |T(e^{j\omega h})| = \{[1+\cos(m\omega h)]^2+\sin^2(m\omega h)\}^{1/2}$$

$$\text{(C9)} \quad |T(e^{j\omega h})| = \{[1-\cos(m\omega h)]^2+\sin^2(m\omega h)\}^{1/2}$$

It is clear that equation (C8) is zero at $m\omega h = (2n+1)\pi$ and equation (C9) is zero at $m\omega h = 2n\pi$. Thus, applying the filter specified by equation (C7) to a perfectly periodic waveform of period mh/n of arbitrary harmonic content will exactly annihilate the waveform. The filter of equation (C6) is somewhat more subtle. It will annihilate, for instance, the odd harmonics of a waveform of period 2mh.

By placing a constant in the recurrence relations, we may move the zeros off the $j\omega$ axis:

(C10) $\quad Y_n = X_n + gX_{n-m}$

(C11) $\quad Y_n = X_n - gX_{n-m}$

This causes the frequency response to approach zero, but never become identically zero unless g is exactly unity.

We may invert the spectrum of these filters by placing the delay in the feedback path, making recursive filters of these two:

(C12) $\quad Y_n = X_n - gY_{n-m}$

(C13) $\quad Y_n = X_n + gY_{n-m}$

The fact that the signs are reversed in the two equations is no accident. Equation (C12) does indeed correspond to equation (C10), and likewise equation (C13) to (C11). These two filters are like the previous ones except that they have resonances where the others had anti-resonances.

The magnitude-frequency responses for the filters in equations (C7) and C(11) were plotted in figures C2 and C3 respectively for $1/(mh) = 400$ Hz and for four different values of g, which were .25, .5, .75, and 1. It might be noted that the graphs are truncated at 20 db. Actually, for g=1, the response in figure C2 goes to zero, which would require our plot to extend to $-\infty$. Likewise, the response in figure C3 goes to $+\infty$. We have somewhat arbitrarily placed a maximum excursion on the plot at + or - 20 db.

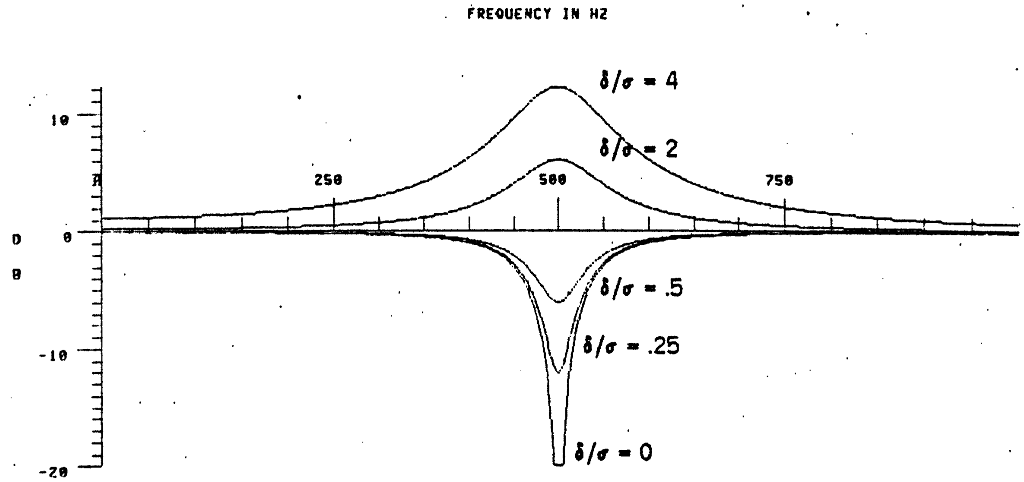


Figure C1. Log magnitude frequency response of digital resonator-antiresonator, plotted for δ/σ = 4, 2, 1, .5, .25, and 0.

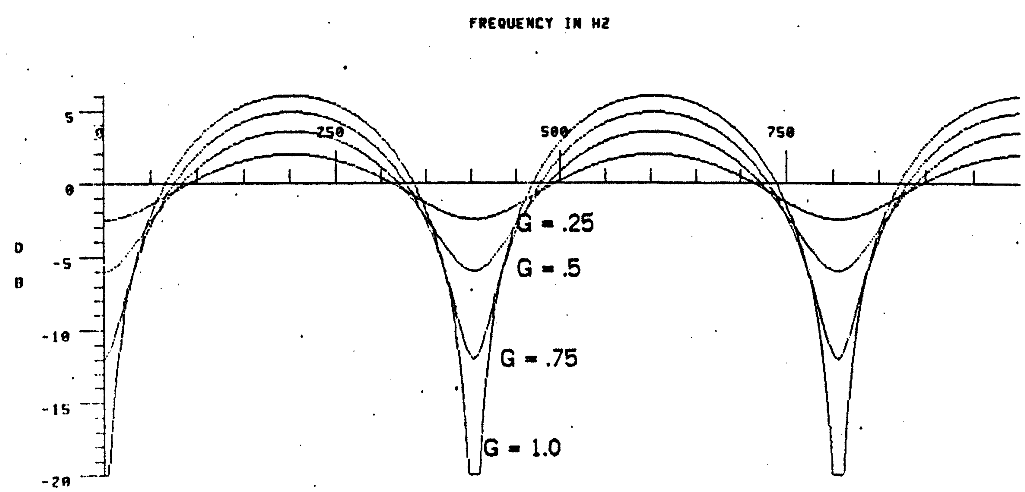


Figure C2. Magnitude frequency response of all-zero comb filter with values of gain G = .25, .5, .75, and 1.0. Zeros of transmission are at integral multiples of 400 Hz.

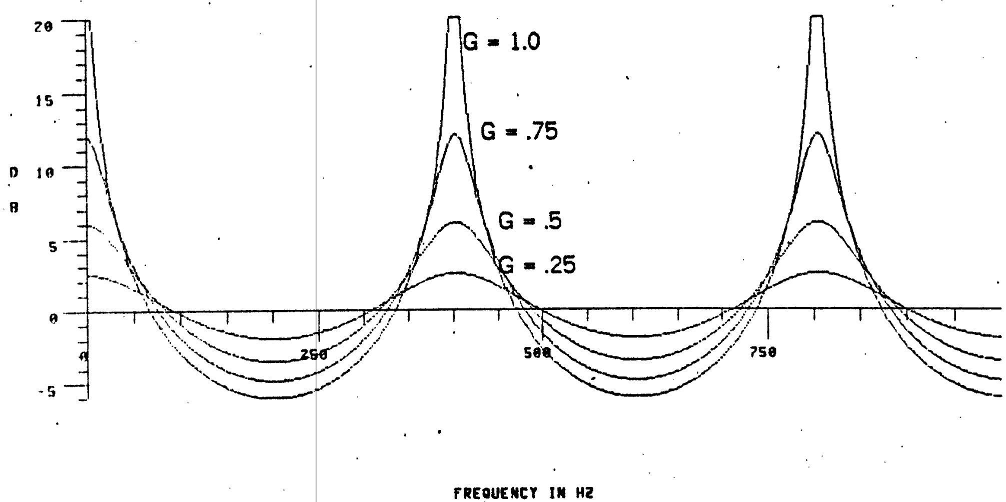


Figure C3. **Log magnitude frequency response of comb filter used as harmonic resonator with poles at integral multiples of 400 Hz. Four values of gain, G, are shown. They are G = .25, .5, .75, and 1.0**

APPENDIX D: ALL-PASS UNIT REVERBERATORS

We use two different forms of the all-pass network: (1) a first-order unit of which the impulse response is a pulse train with exponentially decaying amplitude, and (2) a second-order unit of which the impulse response is a pulse train of which the amplitude is a damped sinusoid.

A simple all-pass filter is the pole-zero pair, symmetrically located about the $j\omega$ axis. The complex frequency response in the continuous case is shown in equation (D1).

$$\text{(D1)} \qquad H(S) = \frac{S-\sigma}{S+\sigma}$$

where S is complex frequency and $1/\sigma$ is the decay time. This filter has a pole at $-\sigma$ and a zero at σ. We may convert this to a digital filter by use of the bilinear transform (Gold and Rader).

$$\text{(D2)} \qquad T(Z) = -\frac{(\sigma h+2)Z^{-1} + (\sigma h-2)}{(\sigma h-2)Z^{-1} + (\sigma h+2)}$$

where Z^{-1} is the unit delay operator and h is the time between consecutive samples. Since this is an all-pass, the magnitude of T(Z) anywhere on the unit circle is the same. This being the case, we may raise Z in equation (D2) to any power without altering the magnitude of T(Z) on the unit circle. If we raise Z to the power of $-m$, the frequency response will cycle m times as we go around the unit circle. The decay time becomes m/σ.

$$\text{(D3)} \qquad T(Z) = -\frac{(\sigma h+2)Z^{-m} + (\sigma h-2)}{(\sigma h-2)Z^{-m} + (\sigma h+2)}$$

Which implies the recurrence relation:

$$\text{(D4)} \qquad Y_n = \{(\sigma h-2)X_n + (\sigma h+2)X_{n-m} - (\sigma h-2)Y_{n-m}\} / (\sigma h+2)$$

This is essentially the unit reverberator used by Schroeder (1961), except we have realized it in the canonical form here, thus saving one multiplication over the form previously used. The frequency response is identically constant around the unit circle. The impulse response is a pulse train with exponentially decaying amplitude. The ratio $(\sigma h-2)/(\sigma h+2)$ is called the gain of the reverberator. As the gain approaches unity, the impulse response of the reverberator decays more and more slowly. We call this the 'first-order' unit reverberator, although technically it is of order m. A block diagram of this unit and a plot of a typical impulse response is shown in Figure 15.

There is one important generalization of the all-pass network. If we begin with an all-pass filter which has complex conjugate poles rather than a single real pole, we realize a reverberator that differs significantly in character from the one in equation (D3). We begin with the following filter, again shown in the continuous case:

$$\text{(D5)} \qquad H(S) = \frac{(S^2-2\sigma S+\sigma^2+\omega_0^2)}{(S^2+2\sigma S+\sigma^2+\omega_0^2)}$$

where ω_0 is the resonant frequency of the filter. This is again an all-pass. Let us transform it via the bilinear transform and substitute a unit delay of m as was done above.

$$\text{(D6)} \qquad T(Z) = \frac{C_1Z^{-2m}+C_2Z^{-m}+C_3}{C_3Z^{-2m}+C_2Z^{-m}+C_1}$$

Which, in turn, implies the recurrence relation:

$$\text{(D7)} \qquad Y_n = (C_3X_n+C_2X_{n-m}+C_1X_{n-2m}-C_2Y_{n-m}-C_3Y_{n-2m})/C_1$$

$$\begin{aligned} \text{where} \quad C_1 &= \omega_0^2h^2+\sigma^2h^2+4\sigma h+4 \\ C_2 &= 2\omega_0^2h^2+2\sigma^2h^2-8 \\ C_3 &= \omega_0^2h^2+\sigma^2h^2-4\sigma h+4 \end{aligned}$$

This, then, is another all-pass unit reverberator of a different character, where the impulse response is a pulse train of which the amplitude is a damped sinusoid. We call this the 'second-order' unit reverberator. It should be noted that with the substitution of Z^{-m} for Z^{-1}, the frequency of the sinusoid for this unit (equation (D6)) becomes $\omega_{0/m}$. The decay time likewise becomes m/σ.

A block diagram of this unit and a plot of a typical impulse response is shown in figures 16.

IV. BIBLIOGRAPHY

American Standard Acoustical Terminology. S1.1-1960. American Standards Association, Inc., New York (1960).

Atal, B. S. ,and Schroeder, M.R., Adaptive Predictive Coding of Speech Signals, Bell Systems Technical Journal, vol 49, 1973-1986 (1970)

Atal, B. S., Hanauer, S.L., Speech Analysis and Synthesis by Linear Prediction of the Speech Wave, J. Acoust. Soc. Am., 50, 637-655, February (1971).

Backhaus, H. Uber die Bedeutung der Ausgleichsvorgange in der Austrik. Zeit. Tech. Physik 13, 31-46 (1932).

Beauchamp, J. W. A computer system for time-variant harmonic analysis and synthesis of musical tones. in Music by Computers, H. von Foerster & J. W. Beauchamp, ed. Wiley, New York (1969).

Boll, Steven Frank, A Priori Digital Speech Analysis, PhD Thesis, University of Utah report UTEC-CSc-73-123, March (1973).

Boxer, R., Thaler, S., A Simplified Method of Solving Linear and Nonlinear Systems, Proc. IRE, vol 44, 89-101, Jan.(1956).

Boxer, R., A Note on Numerical Transform Calculus, Proc. IRE, vol 45, #10, 1401-1406, Oct. (1957).

Carroll, J. D., and Chang, J. J. Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition. Psychometrika 35, 283-319 (1970).

Carroll, J. D., and Wish, M. Models and methods for three-way multidimensional scaling. in Contemporary Developments in Mathematical Psychology. R. C. Atkinson, et. al., eds. W. H. Freeman, San Fransisco (1973).

Chowning, J.M., The Simulation of Moving Sound Sources, J. Audio Eng. Soc., 2-6, (1971).

Chowning, J. M. The synthesis of complex audio spectra by means of frequency modulation. J. Audio Eng. Soc. 21, 526-534 (1973).

Clark, M., Robertson, P., and Luce, D. A. A preliminary experiment on the perceptual basis for musical instrument families. J. Audio Eng. Soc. 12, 199-203 (1964).

Clark, M., and Milner, P. Dependencies of timbre on the tonal loudness produced by musical instruments. J. Audio Eng. Soc. 12, 28-31 (1964).

Colby, K. M. Computer-Aided Language Acquisition in Non-speaking Children, Archives of General Psychiatry, Vol 19, pp641-651, (1968).

Colby, K. M., and Smith, D. C., Computers in the Treatment of Nonspeaking Autistic Children, Current Psychiatric Therapies, (Masserman, J., Ed.), Vol. 11, pp1-17, (1971).

Colby, K. M., The Rationale for Computer Based Treatment of Language Difficulties in Nonspeaking Autistic Children, Journal of Autism and Childhood Schizophrenia, Vol 3, pp254-260, (1973).

Freedman, M. D. Analysis of musical instrument tones. JASA 41, 793-806 (1967).

Freedman, M. D., A Method For Analysing Musical Tones, J Audio Eng. Soc., Vol 16, #4, p419, Oct. (1968).

Freedman, M. D., A Technique for Analysis of Musical Instrument Tones, PhD dissertation, University of Illinois, Urbana. (1965).

Freedman, M. D. A method for analysing musical tones. J. Audio Eng. Soc. 16, 419-425 (Oct. 1968).

Freedman, S. J., and Fischer, H. G. The role of the pinna in auditory localization. in The Neuropsychology of spatially-oriented behavior. S. J. Freedman, ed. Dorsey Press, Homewood, Ill. 135-152 (1968).

Gold, B., and Rader, C. M. Digital Processing of Signals. McGraw-Hill Book Company, New York, (1969).

Johnson, S. C. Hierarchical clustering schemes. Psychometrika 32, 241-254 (1967).

Knudsen, V. O. Architectural acoustics. Scientific American 209:30, 78-92 (1963).

Kruskal, J. B. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. Psychometrica 29, 1-27 (1964a).

Kruskal, J. B. Nonmetric multidimensional scaling: a numerical method. Psychometrica 29, 115-129 (1964b).

Liberman, A. M., Cooper, F. S., Shankweiler, D. P.,and Studdert-Kennedy, M. Perception of the speech code. Psych. Reveiw 74, 431-461 (1967).

Luce, D. A. Physical correlates of nonpercussive musical instrument tones. PhD thesis. MIT (1963).

Luce, D. A., and Clark, M. Duration of attack transients of nonpercussive orchestral instruments. J. Audio Eng. Soc. 13, 194-199 (1965).

Markel, J.D., Digital Inverse Filtering - A New Tool for Formant Trajectory Estimation, IEEE Trans. on Audio and Electroacoustics, Vol AU-20, #2, pp129-137, June (1972).

Mathews, M. V. The Technology of Computer Music. MIT Press, Mass. (1969).

Mathews, M. V., and Kohout, J. Electronic simulation of violin resonances. JASA 53, 1620-1626 (1973).

Mills, A. W. On the minimum auditory angle. JASA 30, 237-246 (1958).

Mills, A. W. Auditory Localization. in Foundations of Modern Auditory Theory, Volume II. J. V. Tobias, ed. Academic Press, 301-348 (1972).

Miller, N. J., Filtering of Singing Voice Signal from Noise by Synthesis, PhD Thesis, University of Utah, Dept. Of Computer Science, technical report UTEC-CSc-74-013, May, (1973).

Moorer, J. A. The heterodyne method of analysis of transient waveforms. Artificial Intelligence Lab, Stanford University (1973).

Oppenheim, A.V., Schafer, R.W., Homomorphic Analysis of Speech, IEEE Trans. on Audio and Electroacoustics, vol AU-16, #2, pp221-226, (1968).

Oppenheim, A.V., Speech Analysis-Synthesis System Based on Homomorphic Filtering, J. Acoust. Soc. Am., 45, pp458-465, February (1969).

Plomp, R. Timbre as a multidimensional attribute of complex tones. in Frequency Analysis and Periodicity Detection in Hearing. R. Plomp and G. F. Smoorenburg, eds. A. W. Sijthoff, Leiden (1970).

Plomp, R. and Steeneken, H. J. M. Effects of phase on the timbre of complex tones. JASA 46, 409-421 (1969).

Plomp, R. and Steeneken, H. J. M. Pitch vs. timbre. Seventh Int. Congr. on Acoustics, Budapest (1971).

Pols, L. C. W. Perceptual space of vowel-like sounds and its correlation with frequency spectrum. in Frequency Analysis and Periodicity Detection in Hearing. R. Plomp and G. F. Smoorenburg, eds. A. W. Sijthoff, Leiden (1970).

Rader, Charles M., Gold, B., Digital Filter Design Techniques in the Frequency Domain, Proc. IEEE, vol 55, #2, 149-171, Feb. (1967).

Risset, J. C. Computer study of trumpet tones. Bell Telephone Labs, Murray Hill, New Jersey (1966).

Risset, J. C., and Mathews, M. V. Analysis of Musical-Instrument Tones. Physics Today, 22, no. 2, 23-30 (1969).

Ritchie, D. M., and Thompson, K., The UNIX Time-Sharing System, Bell Laboratories, Murray Hill, New Jersey 07974

Ritchie, D. M., C Reference Manual, Bell Telephone Laboratories, Murray Hill, New Jersey, 07974

Schroeder, M.R., Natural Sounding Artificial Reverberation, J. Audio Eng. Soc. 10, 219 (1962).

Schroeder, M.R., Logan, B.F., Colorless Artificial Reverberation, J. Audio Eng. Soc. 9 #3, 192, July (1961).

Schroeder, M.R., Improved Quasi-sterophony and colorless artificial reverberation, J. Acous. Soc. Am., 33, 1061 (1961).

Schroeder, M.R., Noll, A.M., Recent Studies in Speech Research at Bell Telephone Laboratories, Proc. 5th International Congr. Acoustics, Liege, (1965).

Shepard, R. N. Analysis of proximities: Multidimensional scaling with an unknown distance function. I. Psychometrika 27, 125-140 (1962a).

Shepard, R. N. Analysis of proximities: Multidimensional scaling with an unknown distance function. II. Psychometrika 27, 125-140 (1962b).

Shepard, R. N. Psychological representation of speech sounds. in Human Communication: A Unified View. E. E. David and P. B. Denes, eds. McGraw-Hill, New York. 67-113 (1972).

Smith, D. C., Newey, M. C., Colby, K. M., Automated Therapy for Non-Speaking Autistic Children, Spring Joint Computer Conference, (1972).

Strong, W., and Clark, M. Synthesis of wind-instrument tones. JASA 41, 39-52 (1967a).

Strong, W., and Clark, M. Perturbations of synthetic orchestral wind-instrument tones. JASA 41, 277-285 (1967b).

Thaler, S., Boxer, R., An Operational Calculus of Numerical Analysis, part II, Circuit Theory, IRE Convention Record, March (1956).

Wedin, L., and Goude, G. Dimension analysis of the perception of instrumental timbre. Scand. Journ. Psych. 13, 228-240 (1972).

Wessel, D. L. report to psychometric society, San Diego (1973a).

Wessel, D. L. Psychoacoustics and music: a report from Michigan State University. in press, Bull. Comp. Arts. Soc. (1973b).

Wish, M., and Carroll, J. D. Applications of "INDSCAL" to studies of human perception and judgment. in Handbook of Perception, Vol. 2. E. C. Carterette and M. P. Friedman, eds. Academic Press, New York (1973).

V. RESEARCH FACILITIES

A. EXISTING FACILITIES

1. HARDWARE FACILITIES AT THE STANFORD AI LAB

All of our research to date has been done at the Stanford Artificial Intelligence Laboratory. The facilities include the following:

Central Processors:
Digital Equipment Corporation PDP-10 and PDP-6

Primary Store:
65K words of 1.7 microsecond DEC Core
65K words of 1 microsecond Ampex Core
131K words of 1.6 microsecond Ampex Core

Swapping Store:
Librascope disk (5 million words, 22 million bits/second transfer rate)

File Store:
IBM 3330 disc file, 6 spindles (leased)

Peripherals:
4 DECtape drives, 2 mag tape drives (7 channel IBM compatible), line printer, Calcomp plotter, Xerox Graphics Printer

Communications Processor:
BBN IMP (Honeywell DDP-516) connected to the ARPA network

Terminals:
58 Television displays, 6 Information International vector displays, 3 IMLAC displays, 15 teletype terminals, 4 Texas Instrument terminals

Satellite Processors:

Digital Equipment Corporation PDP-11/45,
Signal Processing Systems 41.

Satellite Memory:

96K 16-bit words of Intel MOS memory

Audio Equipment:

14-bit 50Kc Analog to Digital converter with 4 channel multiplexing, 16-bit 200Kc Digital to Analog converter with 4 channel multiplexing, Scully 4-track tape recorder, 4 dolby system A noise reducers, Sony 4-track tape recorder, 2 Dynaco stereo 70 watt amplifiers, 4 Altec 804 monitor speakers

Special Equipment:

4 television computer-controlled cameras, 2 mechanical arms, remotely computer-controlled vehicle, laser depth ranging device

The bulk of this equipment has been purchased through contract research supported by the Advanced Research Projects Agency, with some additional funds provided by the National Science Foundation, and the National Institute of Mental Health.

The analog audio equipment (tape recorders, Dolby unit, etc) is owned by the Stanford Department of Music. Much of this was purchased from funds generated by the music group.

The facility is organized as an interactive time-sharing display-oriented system. Powerful graphical and interactive techniques are available as well as a host of computer languages, library functions, utility routines, and text manipulation programs. The system is well human-engineered so as to make research convenient and natural.

The laboratory itself serves Stanford professors, research associates, and graduate students, principally in the field of Artificial Intelligence research, but with other concerns, such as Mathematical Theory of Computation, Computer aids to Autistic children, and several others. It is a delightful place to work due to the very diverse interests and skills of the participants, as well as the computational power of the facility.

It should be mentioned here that we have recently designed and built a 16-bit D/A converter and a 14-bit A/D converter, which replaced the previous 12-bit units. The system is a conventional low-noise, high precision conversion system but has one novel feature. The digital to analog converter has a mode in which it accepts 9-bit bytes of compressed data from the PDP-10 and from this recovers the 16-bit sample. The compression method was discovered by simulating many different compression schemes. The method which provided the most reduction with the least distortion was the floating-point incremental method. The results using this method are indistinguishable from direct 16-bit conversion to even the most discerning of ears. The encoding consists, roughly, of taking the difference between consecutive samples, converting that to a floating-point number, and truncating the exponent to 4 bits and the mantissa to 5 bits. The exponent and mantissa together form a 9-bit byte which is sent to the D/A conversion unit, where the floating number is first fixed and then added into a 16-bit register which is then used as the input to a 16-bit digital-analog conversion module. The encoding method is actually slightly more complex than this, in that we do not subtract consecutive samples, but instead subtract the predicted state of the 16-bit register in the conversion unit from the current sample. This keeps errors from accumulating. Our experiments show that this could be reduced to an 8-bit byte (4 bits exponent, 4 bits mantissa) without loss of fidelity.

2. EXISTING RESEARCH SUPPORT SOFTWARE

In the course of our research, we have developed a number of original and useful programs. It is beyond the scope of this proposal to give a complete description, or even a complete list of all the programs we have written. Some of the principal ones are briefly listed below:

GENERAL-PURPOSE PROGRAMS

The heart of our experimental sound generation is a highly flexible acoustical compiler which is an ALGOL-based descendant of Bell Laboratory's MUSIC V system. This program enables us to produce sounds by any known synthesis technique as well as experiment with new techniques. We have been able to program versatile reverberation routines, spacial localizing routines, and many others.

To connect the synthesis programs with the audio system, there is a set of programs which communicate with that system. For playing synthetic tones, there are programs which can send a disk file to the digital-analog converter for any combination of available sampling rates and channel selections. Monaural, stereo, and quadraphonic files can be played and recorded on any of the tape recorders. Natural sounds can be digitized and stored on disk files. There are sound file editors which can display a segment of a digitized waveform, synthetic or natural, which can display the discrete Fourier transform of a segment of sound, select sub-segments from longer sound files, and many other splicing and editing functions.

A general purpose function generating program allows the user to specify the Fourier components of a complex wave, with control over amplitude and phase of each of the components. In addition, the program is used to generate time-domain functions by means of numerical specification or by use of a light-pen. The program stores all functions as a file on the disk memory.

There is a sampling rate conversion program which can convert between any two sampling rates that are rational multiples of each other. It selects the proper low-pass filter and accomplishes the conversion with minimum distortion.

PROGRAMS FOR REVERBERATION AND LOCALIZATION RESEARCH

We have written an interactive reverberator compiler which aids the researcher in the design of a compound reverberator. The program calculates reverberation parameters, generates and displays the resultant program language description of the reverberator as it is being constructed, displays and plays (on command) the impulse response of the reverberator as each unit reverberator is defined, and reverberates an existing sound file.

An interactive graphic program has been developed for the control of a moving sound source trajectory in the simulated two-dimensional space. The program accepts input from the teletype to specify a computed trajectory, or from the light-pen to specify a 'hand' determined trajectory. The program derives the control functions for azimuth, distance, and velocity, displays them, and stores them on the disk, on command.

PROGRAMS FOR MUSIC INSTRUMENT RESEARCH

A set of programs exist which implement the heterodyne analysis of music instrument tones. The first pass includes FFT's for overlapping windows, which are processed in the next pass to determine average frequencies of the harmonics of the tone. In the next pass, these frequencies set the heterodyne filter both in window size, the fundamental period, and center frequency, the particular harmonic being analyzed. The output is a set of time-varying frequency and amplitude functions for each harmonic listed. The final pass consists of a heuristic program which extracts, by an examination of phase variation and amplitude, the analyzed sound segment which most probably corresponds to the actual tone. It scans the functions for new estimates of the average frequencies of the harmonics, their peak amplitudes, and extracts various other information on the tone. At this point, or above, data compression by any power of two can be performed by an averaging technique.

Another set of programs are used to display, plot, and modify the analyzed functions obtained above. Three-dimensional rotation of perspective plots of the amplitude or frequency functions for the whole set of harmonics, temporal line-spectrum displays, spectrographic displays, and many other forms of graphics are obtainable (as described above in Section IIA1). Furthermore, a host of operations can be performed on the analyzed functions, including smoothing, light-pen modification, line-segment approximation, spectral envelope modification, amplitude and frequency modulation, and many more which serve as tools for research.

Another program is used for optimized additive synthesis, based on the directly analyzed or modified functions. Tones can be synthesized at arbitrary frequencies, amplitudes and durations, from any specified list of amplitude and frequency functions, paired as desired. This allows for the generation of a tone with a reduced set of harmonics, or a set of harmonics from different instrumental origins, or the pairing of arbitrary amplitude and frequency functions from different instruments or having differing operations performed beforehand. Furthermore, procedures can be written for this synthesis program which allow for further data manipulation, as in the interpolation between two tones (described in Section IIA3). The program is capable of reverberation, and accepts any arbitrary note/procedure list which specifies a temporal ordering of events, each having associated parameters for synthesis operations.

A dynamic display program is used for evaluating the results of decisions in FM synthesis. The amplitude of the frequency components are displayed as a function of time where the modulation index is interpolated between any two values. The program asks for a ratio of carrier to modulating frequencies, beginning and ending values of the index, and the number of steps in the interpolation. The displayed information includes the value of the index at every step, so that the user can interrupt the interpolation to examine the spectral shape in relation to the value of the index.

An equally useful representation of the same data exists in a three-dimensional interactive display program. In this program, the effect of complex functions for both amplitude and modulation index can be examined as they affect the spectrum shape.

PROGRAMS FOR PERCEPTUAL EXPERIMENTATION

A whole host of programs exist which conduct on-line psychoacoustical experiments, playing tones over the D-A converter; allowing listeners to hear stimuli again, often in any order. (It should be mentioned that the proposed real-time system would make it possible for listeners to manipulate various parameters of stored signals and get immediate feedback. This would give us a much greater tool to uncover the most important aspects of the perception of instrument tones.) Responses of listeners are taken, stored, and eventually analyzed by other programs, e.g. analysis of variance and multidimensional scaling.

A particularly useful program exists for the display of the results from multidimensional scaling experiments. Using a series of mirrors, it allows the viewer to see convincing three-dimensional configurations in realistic perspective. Rotation can be performed, and two configurations can be observed and compared simultaneously.

B. PROPOSED FACILITY

1. HARDWARE FACILITY

The process of analysis and synthesis as described in the preceeding sections demands great quantities of computer and human time. The computations can take as much as 100 times the length of the produced tone. This requires great patience from the researcher. Since many of the results are empirical, many experiments have been done and must continue to be done. For this reason, we propose a special-purpose computing system. The system would initially be a satellite to the Artificial Intelligence Laboratory, but would serve not only to reduce the computation load of the PDP-10, but by means of special-purpose hardware, actually synthesize tones in reverberant environments in real time. The system would be powerful enough to stand alone if necessary. Since it would be extremely expensive to purchase a system with the human engineering of the AI lab facility, we propose to use the AI lab facility as an interface to the special-purpose system, thus minimizing change-over inconveniences and initial setup price. As work continues, it would be possible to upgrade the system to stand alone and eventually provide a high degree of service without the aid of the AI facility. To this end, we propose using an existing time-sharing system as the resident monitor in the special-purpose system. This saves us the trouble of having to write device controllers, memory management programs, and other system-level functions. Since we propose to begin with a time-sharing system, upward compatibility is assured. Programs will continue to run unmodified as the system is upgraded. Although the detailed hardware budget is given in section V, we shall discuss the main items here. The following is a list of the principal components of the proposed facility:

Digital Equipment Corporation PDP-11/45 Computer
with floating-point unit
and memory management module

Bell Laboratories UNIX time-sharing system and
software support package

Digital Equipment Corporation RP03 disc drive.
Provides 20 million 16-bit words of storage.

Systems Concepts signal processor with digital
reverberation module

Digital Equipment Corporation GT-40 graphics terminal

Audio equipment, including 8-track tape recorder,
quiet booth, acoustically treated room,
8-channel amplifiers and matched speakers

The heart of the system is a PDP-11/45 with floating-point processor and memory management module. This is a powerful mini-computer capable of high-speed arithmetic and basic time-sharing operation. The purpose of such a computer is to provide a versatile test bed for research. The floating-point processor greatly aids numerical computations of the type which are so common in digital signal processing and multi-dimensional scaling. The memory management module provides for time-multiplexing of multiple tasks, so as to more effectively utilize the computer by running processes while others are idle. It also provides for cooperating parallel processes which aids decomposition of large tasks into smaller modules.

A computer system of the power we propose needs a sophisticated monitor to manage the multiplicity of tasks demanded of it. We propose the use of the UNIX time-sharing system, developed at Bell Telephone Laboratories, together with its support software. This system provides not only a multiple-user service, but it provides system support for cooperating parallel processes. Since music is an inherently parallel process, this seems like a natural choice. With the UNIX monitor itself comes a complete support package, including FORTRAN, ALGOL, the C compiler (a system programming language based on BCPL), editors, assemblers, numerical routines, and much more. A description of the UNIX system is available in a writeup by Dennis Ritchie [Ritchie]. A reference manual for the C compiler is also available to the public [Ritchie].

For bulk storage, an RP03 disk system is included. This is necessary for real-time operation. The bulk of the disk is to store digitized audio and control information for the signal processor. In analyzing music instrument tones, we digitize recorded natural tones to a precision of 14 bits (stored in 16-bit PDP-11 words) at rates up to 50,000 samples per second. One can easily see that to store individual notes of each of the orchestral instruments in all of their characteristic playing modes would be a staggering amount of storage. The RP03 disk can provide storage for up to 400 seconds of sound. This is a compromise between price and working requirements, representing the minimum amount of storage that can be effective and useful.

An alternative to getting an RP03 disk would be to use the AI facility's disk storage.

This is unwise because of the required rate of data transfer between the machines and would require a high-speed data channel for communication with the AI PDP-10. The most efficient solution is inclusion of some amount of bulk storage on the PDP-11 itself, and for real-time operation, this is the only solution.

The most important item of all is the Systems Concepts Signal Processor. This is a highly-parallel, special-purpose, programmable, digital processor designed especially for the generation and processing of audio signals. Together with the 48K reconfigurable bulk storage for programmable reverberation, it provides enough power and flexibility to synthesize all the tones we have produced to date in real time, even with spacial localization and 4-channel reverberation involved. It can even do the computations in the heterodyne filter analysis in real time. The processor is designed to be controlled by a small computer and a PDP-11 interface is a standard option.

When one attempts to generate sounds using a new technique, often there is no good way to make an *a priori* prediction on the range of the controlling parameters. A good example of this is deciding over what range to sweep the modulation index of an FM instrument. With our current turnaround time, one must "shoot in the dark" in attempting to find the correct parameter, often wasting tremendous amounts of computer time as well as personal time in the process. The PDP-11/45 by itself offers no speedup, but combined with the Systems Concepts Signal Processor, provides the solution to the problem. It would be possible to directly connect a knob, via some PDP-11 support software, to synthesis parameters, thus allowing the experimenter to directly control the parameter as the sound is generated. This increases the efficiency of the research process immensely. It also makes possible experiments which would be otherwise impractical or even impossible. One example of this would be a two-knob experiment where one knob controls the duration of a note and one controls the loudness. To do this experiment without real-time generation of the sound would require preparation of a large number of sample sounds beforehand. With three knobs, the size of such samples exceeds even the bounds of the entire AI lab bulk storage. It is much more efficient to store a sound as the program which can synthesize it rather than the digitized waveform itself. A whole new domain of experiments immediately suggests itself, based on interactive control of complex attributes of synthesized sound.

We wish to add to the signal processor the reverberation memory option. This device provides for a number of variable-length digital delays which are easily interfaced with the signal processor itself to provide reverberation in all the forms we have realized to date, and have enough generality to provide for any future forms of reverberation we may discover. Again, the parameters of the reverberation could be easily attached to knobs, giving the user direct real-time control over the character of the reverberation. This would greatly enhance the productivity of the researcher.

It should be mentioned again that the Signal Processor is of such generality and power that it is easily capable of synthesizing speech using the methods that are currently popular. Most speech synthesis is done be taking an excitation function of some kind, often a pulse train for voiced phonemes and white noise for frication, and applying spectral shaping filters to produce formants. This is the idea of the synthesis by linear prediction of the speech waveform [Atal]. Since the Signal Processor is capable of generating a large number of excitation functions and realizing a large number of digital filters all in real time with complete freedom of interconnection, it would seem to be capable of synthesizing several voices, possibly as many as four, in real time. Although our research is not directly concerned with speech synthesis, the techniques for programming and controling the Signal Processor would certainly be applicable to the synthesis of speech.

Again, such a device as the Systems Concepts Signal Processor could be interfaced directly to the AI PDP-10, but it requires real-time control of the type that is not generally available in time-sharing systems.

There are alternatives to purchase of the Signal Processor that should be discussed. One alternative would be to design and build the device in house. Since the parts costs for such a device run about one-fifth of the purchase price, this might seem like a bargain. It is not. One must also consider engineering and assembly costs. It is estimated that it would require an engineer about 6 months of full-time effort to design the device, and would require a technician about 6 more months to build it. The debugging effort could probably be completed in another 6 months. Although this comes out to less money than the purchase price of the device, this places the birth of the device at the end of the funding period, leaving no time to do any research. This also means that the research personnel would be engaged in hardware design rather than in research. This could be viewed as somewhat of a misdirection of talent. It would seem more appropriate to contract a hardware house (like Systems Concepts, Inc.) to supply the device than to require a group of researchers to attempt hardware design.

Another alternative might be to start from an existing design, like the digital filter unit by Alles of Bell Telephone Laboratories. This unit is, however, dedicated to digital filtering and has little versatility. With only a slight modification, one could get other functions such as ring modulation and random number generation. Alles, however, has not planned for any signal generating devices. This would be another major design effort of dubious utility. Again, we hope that the system could be easily replicated at any other site, so that the fruits of our work could be used elsewhere. This is greatly enhanced by the use of standard readily available off-the-shelf equipment.

The GT-40 display console is a general-purpose graphics terminal. It provides a direct graphical interface to the PDP-11 and thus to the Signal processor. As was noted many times in the previous sections, the use of computer graphics is an essential piece of human engineering that we take advantage of constantly. The AI facility has made graphics highly available and easily used. It thus has crept into many programs as a debugging aid as well as a research aid. The Simmographics tablet is a precise, versatile, and inexpensive input device that is invaluable for graphically editing or inputting complex functions.

The audio equipment is the final stage of the sound production. The waveform must be converted to sound by an audio system whose quality matches the extreme quality of digital synthesis. We propose to place 8 speakers in an acoustically treated room for listening tests. The number 8 is somewhat of a compromise with cost, as we have little *a priori* reason to believe that 8 channels will be enough. Our only real evidence is our success with 4 channels.

The importance of the listening room is great. The environment of a computer facility is very noisy. Computers require extensive cooling and air conditioning, making the computer room sound somewhat like a continuous hurricane. In addition, the particular building we are in uses forced-air cooling in each room, adding a gentle but disturbing hiss to all offices. To do perceptual testing, it is essential that the acoustical environment be controlled entirely by the investigator.

A small quiet booth is also proposed for recording and analyzing sounds from real music instruments. The 8-track tape recorder provides off-line storage of acoustical data. Audio tape is still the most economical way to store large quantities of audio, at the cost of some loss of fidelity and signal-to-noise ratio.

2. PROPOSED RESEARCH SUPPORT SOFTWARE

An amount of software would have to be written to integrate the proposed system into the current system. This could be done in layers, each preserving compatibility with the previous systems but each advancing the researcher's capabilities. The following discussion attempts to summarize the software which would have to be developed.

THE PDP-11 MONITOR

Although we intend to use the UNIX time-sharing system from Bell Laboratories as a base, it will need some amount of modification to deal with our specific set of peripherals: the GT-40, the RP03 disk, the PDP-10 interface, and the Signal Processor. It is not expected that this would require a great deal of effort to get a basic monitor up and running. Improvements and streamlining can always be added later as the need arises.

THE SIGNAL PROCESSOR

Writing software for the Signal Processor presents somewhat of a problem. One must provide methods not only for setting up the internal computation flow, but also for synchronization and data transfer with cooperating PDP-11 processes. One example might be interactive control of mixing natural and synthesized sounds. The PDP-11 would be reading analog to digital converter output, interactive knob settings, as well as controlling the signal processor and providing it with the parameters read from the knob and the waveform from the A/D converter all at once. The theory of cooperating parallel processes, however, is rather well developed such that straightforward methods can be applied here with little difficulty. It is a great advantage that the UNIX time-sharing system provides for parallel processes and has a highly developed inter-process communication system. Only slight modifications would have to be made to provide for real-time interaction. The efficiency of the inter-process communication system would have to be examined and possibly reprogrammed to assure that it does not incur prohibitive overhead.

One would write a program for the Signal Processor just as one writes a program for a computer. At first, the language would be a low-level one, corresponding to an assembler for a computer. As we learn more about controling such a device, the language could be upgraded to a higher level. One would specify communication with other processes through a system of ports. At run time, the user would be able to specify which ports should be connected to which channels. It is then that the user would specify which communication ports should be directed to which knobs, and which ports should be connected to which disk files, and so on. This maintains a level of generality which permits great flexibility in testing and repeating a run.

At first, we would merely seek to construct an interface with our existing PDP-10 synthesis programs such as to ease the change-over problems. This would allow for direct synthesis of all the sounds we have generated to date, and would allow research to proceed at a much greater rate, but would not allow for real-time interaction until the assembler for the Signal Processor was completed.

Since many of the analysis routines, as well as synthesis routines, could be executed on the Signal Processor, one would want ways to send digitized waveforms to the Signal Processor. It should be possible to come in directly from the analog-to-digital converter as well as retrieve stored waveforms from the disk. The output of the signal processor could be directed to the digital-to-analog converter as well as directed to a disk file. The system could also get at files on the AI Lab IBM 3330 disk, but only at a limited rate.

This ability to store the output of the signal processor is quite important. The most compelling reason is one of growth. As we progress, it is quite possible that we will demand computations so complex that even the Signal Processor can not complete them in real time. In this case, we must have the option of directing the output to the disk. In this manner, we could break up the computations into smaller runs and mix them in the PDP-11 for later playback.

There are also many utility routines that would have to be written for the PDP-11. We would need an implementation of the fast Fourier transform algorithm, an intermediate-level graphics package for convenience in displaying complex functions, routines for getting at the digital-analog converters, as well as ways to direct asynchronous processes in a uniform and convenient manner.

All of this is programming support that would be needed for interfacing our current research to the new system. Eventually, we would begin writing research programs directly on the PDP-11.