Center for Computer Research in Music and Acoustics

Department of Music
Report No. STAN-M-8

SIMULATION OF MUSIC INSTRUMENT TONES

Final Report

by

J. Chowuing, J. Grey, J. A. Moorer, L. Rush, L. Smith

May 1978

Research sponsored by the National Science Foundation at Stanford investigated the perception of timbre, making use of recent developments in the fields of digital signal processing and multidimensional perceptual scaling. Various interwoven phases of research were initiated; they are oriented towards uncovering the distinctive features and dimensions in timbre perception. The four phase of research discussed in this report are: Analysis by synthesis, Multidimensional scaling, Categorical perception, and Frequency modulation synthesis.

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

I

t

FINAL TECHNICAL REPORT

NSF contract DCR 75-00694

Simulation of Music Instrument Tones

An ongoing research project at Stanford has investigated the perception of timbre, making use of the most recent developments in the fields of digital signal processing and multidimensional perceptual scaling. Various interwoven phases of research have been initiated; they are oriented towards uncovering the distinctive features and dimensions in timbre perception. We will describe the four important phases of our research program below.

Analysis by synthesis

The first is the *analysis-by-synthesis* approach. In this phase of research, we are interested in the perceived differences between tones having physical representations that vary in complexity. A physical representation of a tone is defined as the parametric information used to control the process of synthesizing that tone. For example, one form of physical representation is a set of time-variant functions that control the amplitudes and frequencies of each harmonic of the tone. Given this particularform of representing a tone, the *complexity* of data used to synthesize a tone may vary considerably. In the control of the frequencies of the harmonics, one extreme of complexity would be to directly use the time-variant functions that were analyzed for a real tone that we are attempting to resynthesize; the opposite extreme of simplicity would be to use constant harmonic frequencies.

The most complex physical representation for a signal is directly derived from the digital analysis of the natural sound that we are attempting to resynthesize. In our past research we have used a heterodyne filter analysis technique (Moorer, 19751 to determine the time-variant amplitudes and frequencies of the harmonics of isolated tones. Tones resynthesized from the analyzed information were almost indistinguishable from the original tones - differences that existed related to the existence of tape noise on the originals which was not faithfully replicated in the resynthesis; there were no significant differences in the naturalness of the tones.

We compared different types of simplification of the highly complex, originally analyzed data for many tones. The results of this study indicated that the very complex control functions may be successfully replaced with functions composed of a small number of line segments; there is no important loss of'tone quality as long as the prominent changes in the amplitudes and frequencies are modeled. Further simplifications were not uniformly successful, however. The unsuccessful simplifications pointed toward perceptually important aspects of the original signals. Two simplifications that were revealing in their lack of success with several tones were: 1) the elimination of any existing low-amplitude precedent activity in the attack portion of tones, and 2) the substitution of constant harmonic frequencies for the time-variant frequency functions of the p~rtials. These failures in simplification point to the perceptual importance of certain details found in the attacks of tones and to the importance of the changes found in the frequencies of the harmonics of actual tones. It should be noted that much perceptual research, as well as musical synthesis, has ignored the latter parameter of sound by using constant harmonic frequencies for producing signals. This research was formally written up and published in the journal of the Acoustic Society (Grey and Moorer, 1977).

On the basis of the work done thus far, we feel that this area of research is among the most promising in timbre research. During this period, we focussed our efforts in this area on the improvement and implementation of research tools that relate to 1) the recording and cligitization of sounds, 2) the analysis of sounds, and 3) the data managernient for manipulating or simplifying the analyzed physical representations of sounds. We also 4) performed an important experiment relating the findings from research carried out on isolated **tones (above) to stimuli in musical** contexts.

1. *Recording and digitization.* Our early research was limited by the constraint of having to first record musicians using conventional tape equipment before digitizing the sounds. The tape storage which preceded digital storage was not just an unnecessary redundancy; it unfortunately added the problems of tape hiss and dropout plus amplifier noise and distortion to the signal *before* digitization. The cleanest recording process would be to *directly* digitize the output of a microphone. This would avoid altogether many of the problems encountered in our original research. For this purpose, we have constructed a recording studio within reach of the computer for the direct digitization of sound (Rush, et. al., 19761. We in fact were able to directly record into the computer several members of the San Fransisco Symphony Orchestra starting in March, 1977.

A second improvement of our conditions for direct digitization was the final debugging of our new analog-to-digital and digital-to-analog converters, thus eliminating sources of noise which had earlier hampered our research. Our present digitization capacity is 1-4 bits at up to a 51.2 kHz sampling rate; conversion from digital-to-analog allows up to 16 bits at over 51.2 kHz (note that 51.2 kHz permits a theoretical maximum signal frequency of 25.6 kHz - the aliasing filter, however, has its 3 dB point at 20 kHz). Four channel simultaneous conversion is possible.

2. Analysis of sounds. We have extended the generality and improved the theoretical accuracy of our waveform analysis process. Our former analysis technique employed the heterodyne filter; we have developed as its replacement *a phase vocoder* technique (Moorer, 19761. Several reasons account for this development. The major practical li mitation of' t-he **heterodyne filter was that it only allowed the analysis** *of isolated and harmonic tones* which had*no vibrato* or significant pitch changes. The theoretical shortcoming was that the output of the heterodyne filter did not provide information with which a waveform *identical* to the original might be constructed.

The phase vocoder is a significant improvement over the heterodyne filter. Its output can be used to resynthesize a waveform which is theoretically *identical* to the orignal. It allows the analysis of tones *in temporal contexts,* having *inharmonic partials* and any degree of *vibrato* or internal pitch shifts. These improvements make it possible to increase the range of timbres investigated both in terms of types of instruments (bells, percussion instruments and other inharmonic sounds) and types of playing styles (tones having vibrato, internal pitch changes and sharply changing attacks). Replacing the limitation of measuring responses to isolated tones, because the heterodyne analysis technique only allowed the analysis of individual signals, we now may deal with any temporal context or tonal environment. This is an important change in research potential; the study of isolated tones has a very limited relationship to real-world perception of timbre. In fact, a sizeable percentage of musicians cannot even identify the instruments producing isolated tones - a serious message for research conducted without temporal contexts.

3. Acoustic data manipulation. An extensive on-line language has been developed for the manipulation and simplification of acoustic data represenions of analyzed tones in any context. The language provides simple, interactive data management. It is display-oriented to allow the graphic examination of acoustic data in any form desired, such as a rotating three-dimensional representation of the signal in terms of ampl ' itude x frequency x time. Graphic evaluation of any data manipulations may be made. Data manipulations allow the further analysis and/or alteration of acoustic control functions with any formal techniques of signal processing. The researcher may also interactively edit functions by hand. Synthesis of tones based upon any stage of data representation is an output of this program; also output and stored is the data representation itself. Interactive flexibility is increased with the possibility of creating on-line macros or procedures to be used ~epeatedly, for example, in a set of harmonics or a set of tones. The procedures may be derived directly from past commands given to the program or may be formally programmed; storage of procedures permits their recall and editing.

The tool just described is central to all of our research. We are increasingly concerned with systematic manipulations of theform of the data representation of a signal. The simplification of the complexity of the data was certainly a design goal of our acoustic manipulation program; the alteration of the form of the representation was a design goal which originated from the orientation of our future an alysis-by-syn thesis research. An extended language which would even more efficiently handle the problems involved in on-line data manipulation was designed from the experiences gained from use of the above described language. This language includes interactive access to libraries of procedures via run-time loading, abbreviated on-line command structures geared for interactive efficiency, an expanded run-time interpretive capacity and an extensible vocabulary of item types. Graphic feedback and nested procedural control, which was found above to be a key factor in successful interactive exploratory research, is central in this language design. Implementation is currently under way, started during the above granting period.

4. Measurements in temporal contexts. An important research project initiated in this period is that our perceptual measurements - such as the measurement of timbre discrimination discussed **above - is performed** for timbres in various temporal contexts in addition to the traditional measurment of isolated tones. One of the most serious problems in attempting to relate psych oacou stical studies to the everyday perception of timbre is that researchers only study isolated tones, while professional musicians often have trouble even identifying the source of a single tone independent of musical context. This is an oversight that we are trying to correct by conducting all experimental tasks, that so lend themselves, in temporal contexts as well as in the isolated condition. This will further allow a precise assessment of the! effects of different contexts with respect to the judgment taken in isolation. By looking at tones in contexts, we will add characteristics that do not exist in isolation; there are attributes of timbre only found in the relationships between several tones - such as the overall spectral pattern across notes, that is, the resonance structure of the instrument.

An experiment was constructed to evaluate the applicablity of the above results (Grey and Moorer, 19771, in the case of more musically relevant contexts. In this case, we constructed a number of one-voice and multi-voice melodic patterns and tested the ability of musicians to detect a timbre change midway through any given pattern. The patterns may or may not have actually changed, but the two versions of a timbre used were just those very close cases shown in our previous study: one version was synthesized according to the fully complex data from the analysis and the other version was synthesized from the simplified data. Hence we tested the ability of listeners to differentiate the two timbral versions in the more musical contexts and compared the results with what was found in the case of the two-tone experiment above. To our surprize, we found that the differences between the versions of certain instrumental timbres were more easily heard in the musical contexts than in the two-tone experiment. However, for the other instruments these differences tended to disappear in the musical contexts. We eventually tr aced this directly to the types of physical differences which were caused by simplifying the functions used for synthesis. When this simplification (inadvertently) caused differences in the number of harmonics used for the tone, these differences were magnified by the more extended musical presentation. Small temporal differences caused by the simplifications were more easily heard in the rather unmusical two-tone presentation. This research was formally written up and submitted to the journal of the Acoustic Society [Grey, 19781.

Multidimensional scaling

A second research area is the analysis of perceptual similarity data using *multidimensional* scaling techniques. Spatial representations of the perceptual relationships between stimuli are constructed so that distances between the stimulus points in the space correspond to their psychological "distances" as measured by the similarity judgment. The similarity judgment is seen to be more neutral and general in nature than ratings of the stimuli upon more stimulusspecific verbal scales; in that the listeners are not instructed to attend to specific attributes of tone, there is less inherent experimenter bias introduced. The spatial representation, or solution, of the similarity data is then interpreted in terms of the physical properties of the stimuli underlying their relational arrangements.

A three-dimensional spatial representation has been found to best fit the similarity relationships for a set of 16 tones. The utilization of synthetic stimuli having the perceptual complexity and richness of natural tones, yet generated from simplified physical representations (obtained from the first phase of our research discussed above) has made the interpretation of perceptual spaces correspondingly less difficult. One dimension was interpreted to correspond to the spectral energy distribution of the stimuli: the bandwidth of the signal, the balance of energy in the low harmonics, and the existence of upper formants. The other two dimensions related to temporal attributes of the stimuli that may underly the categorization of tones into families. One dimension related to the synchronicity of the attacks and decays of tipper harmonics, also corresponding to the degree of spectral fluctuation s found in the tone through time. The other dimension appeared to relate to the existence of low-amplitude, high-frequency inharmonicity in the initial attack portion. This research was formally written up and published in the journal of the Acoustic Society [Grey, 19771, during the time period covered by NSF contracts BNS 7517715 and DCR 75-00694.

Taken in total, all (six) reported scaling studies of timbre - despite g-reat differences in the nature of their stimuli - have all found one common physical property that was important to the judgment of similarity among tones. That was the *spectral energy distribution.* We have expended a good deal of effort in this granting period in an attempt to best characterize this dimension of tone and to empirically verify its importance through the scaling of an altered set of stimuli. Further work has been done in trying to best model the other dimensions of tone uncovered in our scaling studies.

1. *Empirical verification of the spectral axis.* Recent research has been aimed at *verifying* the psychophysical interpretation of uncovered dimensions. One promising verification technique is to *alter* a subset of the original stimuli with respect to a particular physical property of tone that was interpreted to underly a subjective dimension. After certain of the stimuli are physically altered as suggested by the psychophysical interpretation, another scaling experiment follows. The newly-generated scaling solution is then examined for predicted changes of positions of stimuli along the relevant axes. This has been successfully done for the dimension relating to the *spectral*energy distribution* of signals. Four pairs of tones taken from the original 16 stimuli were altered to effect a pairwise trading of spectral envelopes. In the rescaling, all four pairs had exchanged positions on the axis relating to spectral energy distribution, verifying the psychophysical interpretation of that dimension. This research was also formally written up and submitted to the journal of the Acoustic Society [Grey and Gordon, 1978), during the time period covered by NSF contracts BNS 75-17715 and DCR 7500694.

2. *Mathematical modeling and testing of psychophysical interpretations.* In order to theoretically verify and make more explicit our psychophysical interpretations of similarity structures, we have constructed formal quantitative models of physical properties which appear to correlate to psychological dimensions. The mathematical correlation between the output of any quantitative model and the relevant axis in the spatial solution measures the relative success of that interpretation. Thus, the psychophysical interpretation becomes more explicit in form.

For example, we have constructed a number of models to quantify the spectral energy distribution of signals; correlations have been taken between the outputs of these various formalizations and the positions of stimulus points along the relevant perceptual axis. In all, over 100 specific models were tested - these were generated as parametric: combinations of **the following model** components: 1) That alternately characterized the spectral levels of the harmonics by their: a) peak amplitude values, b) averaged amplitudes, or O averaged energies; 2) That utilized the a) amplitude or b) energy pattern of the spectrum; 3) That measured the a) centroid (mean) or b) middle (adjusted median) of the spectral distribution so characterized; and 4) that characterized the spectrum by a) its physically measured levels (in the various forms above), or b) transformed the spectrum by a function which takes peripheral properties of the ear into account, including critical bands and masking, or c) the further transformation of this peripheral pattern of excitation to characterize its higher-level cognitive representation, modeled

on the loudness function.

Of these various forms for characterizing the spectral distribution, the one which gave the best results was the centroid, or balance point, of the averaged amplitudes transformed by the model for loudness. Correlations between this quantification of stimulus spectra and the locations of stimulus points on the appropriate perceptual axis turned out to be .94, .92 and .92 - for the data from three separate scaling studies. Correlations of the other models reflected the importances of the components of this best-fitting quantitative characterization of spectral distribution.

Categorical perception

A third phase of our research seeks to investigate the nature of the timbre space that exists in between the known, naturalistic tonal points of reference. This phase of research asks whether timbre is perceived in a *categorical* or continuous mode. The subject for investigation is the perceptual properties of a set of tones produced by a process of interpolation of timbres between two naturalistic "endpoint" tones. The question is whether a physically continuous interpolation would be perceived to gradually evolve from one endpoint to the other; or rather, would the tones in between be heard as identical to their nearest endpoints, with an abrupt jump somewhere in the center of the interpolation from one side to the other? The latter form of perception has been called *categorical,* where an acoustical continuum is split into groupings, or categories, with little distinction between the members within any group., With the successful simplification of physical representations of tones (from our first phase of research discussed above), various schemes for *timbral interpolations* have been constructed. Resulting synthetic continua, spanning two known tones, have been investigated for categorical perception. One~ extensive study looked at the six interpolations between 4 of the 16 tones used in a multidimensional scaling study. All interpolations were found to be smooth and non -categorical, based upon discrimination, identification and hysteresis experiments. A similarity scaling of the 6 interpolation "midpoints", 4 endpoints plus 8 other original tones yielded an interpretable three-dimensional representation, with midpoints failing between endpoints. This research (along with the above research phases) was for.mally summarized in an invited paper for the Bulletin of the Council for Research in Music Education [Grey, 1978b), during the time period covered by NSF contracts BINS 75-17715 and DCR 75-00694.

Summary and general comments

The three major areas of our research program on timbre - an alysis-by-syn thesis, multidimensional scaling and categorical perception - have been described. They are interwoven, yet provide different types of information about perception. The analysis-bysynthesis phase provides a direct access to the critical physical features of' tones, whether these features serve as *identification cues* of an instrument or relate to the *tonal quality* or *naturalness* of a sound. This area of research is primary in importance because of its extreme directness in tapping psychophysical relationships; it is also a preliminary phase! to other areas of investigation because it allows the generation of stimuli that have been reduced in their psychophysical complexity without changing their timbral properties. The multidimensional scaling research explores the salient dimensions of *relationship* heard among a set of different timbres - uncovering the stimulus variables correlating to the perceptual dimensions in an attempt to make a psychophysical model for timbral similarity. Finally, the categorical perception phase of research investigates the continuity of timbre space between the known points of reference - looking at the effects of familiarity on the perception of artificially created timbral interpolations between real tones. Our primary justification for the inclusion of diverse areas in this research program is the different sorts of perceptual information gleaned from the various phases. Timbre is a complex subject to study; no single experiment or unilateral approach will provide enough information to build a thorough model for timbre perception. Systematic investigations will generate the information needed; this is a long term project similar in many ways to the area of speech perception.

Frequency Modulation synthesis

The research accomplished in Frequency Modulation (FM) synthesis as a technique for the production of complex dynamic spectra yielded 1) an effective model for some **natural** instrument tones, 2) an important generalization of non-linear synthesis, discrete summation formulae, where FM is one of a class of very powerful techniques which are being vigorously developed, and 3) a domain of future research regarding the functional relationship between the ordering of inharmonic frequency components and tuning systems.

1. Models of Instrument Tones The extension of the basic FM algorithm of a sinusoidal modulation of a sinusoidal carrier to a complex modulation of a sinusoidal carrier was implemented in order to achieve a finer control over the spectral envelope. Most natural tones have a large number of frequency components encompassing a large bandwidth, but with a very definite weighting of the energy distribution toward the low order components. Using the single modulater algorithm one can effect either large bandwidth or spectral weighting of the energy, but not both. We discovered that by applying multiple sinusoidal modulators, where each has a different frequency ratio relative to the carrier frequency and an independent modulation index -as a function of time, both large bandwidth and spectral weighting can be obtained. This technique was first applied to the simulation of string tones, one of the most complex of natural instrument tones, with good results (Schottstaedt, 1977). The same technique was applied to the simulation of piano tones with perhaps better results. The piano is a particularly interesting case in that it's physical acoustical properties have been thouroughly studied and yet a successful simulation had remained elusive. Two aspects of the physical properties had the greatest importance in relation to the perceptual proper-ties. The first is the fact that because of the enormous tension of the piano strings their mode of vibration is as much like a vibrating rod as it is like a vibrating string.. This results in a stretching of the harmonics which partly accounts for the characteristic timbre and also results in the second important perceptual property which is the adjustment of the tuning of the strings from equal temperment in order to accomodate to the effect on pitch of the stretched harmonics. These two perceptual properties were represented in the FM simulation by two simple functions, while the spectral weighting was achieved using a three modulator one carrier algorithm (Schottstaedt, 1977).

2. Synthesis of Complex Spectra by Means of Discrete Summation Formulae The conceptual and computational simplicity of FM synthesis led to an investigation of additional non-linear techniques, which have some different attributes, but which maintain to a large extent the desirable simplicity of FM. These techniques have been described in the litrature (Moorer, 1976, 1977). With a slight increase in computational complexity there is a significant increase in the control of the spectrum. Where the spectra in FM synthesis can be specified as to bandwidth, the individual aplitudes of the frequency components can not be specified as they are determined by the Besse] coefficients. On the other hand, in the case of some of the discrete summation formulae, the amplitudes of an arbitrary number of frequency components can be precisely specified. This fact is a real advantage in that the spectral envelope is critical in the simulation of sustained instrument tones. In particular, the technique which allows specification of the spectral envelope is called 'Wave Shaping Synthesis" the theory of which has been rigorously defined (LeBrun, 1978) and which has yielded surprisingly realisitic instrument and vocal tone simulations.

3. Functional Relation Between Spectra and Tuning An interesting side issue has been raised in our research into non-linear synthesis techniques. One of the characteristics of all of these techniques is that the pattern of distribution of the frequency components in the spectrum is .determined by a single ratio. For example, in the case of FM, if a ratio of the carrier to the modulating frequencies is 1, then the pattern of distribution is in the harmonic series. If on the other hand the ratio is 1/3, the pattern of distribution will be a subset of the harmonic series where every third component is missing. A ratio not composed of simple integers, for example **1/1.414, will produce** a spectrum whose components are inharmonic as is the case with bells, gongs, drums, etc., but with a special and interesting difference: unlike inharmonic spectra produced by natural instruments, these components also fall in a recuring pattern as in the case of ratios of simple integers, The question is whether or not this property can be usefully exploited. From our initial investigations it appears that the answer is yes.,

One of the long-standing questions in musical perception is why there seems to be a special harmonic 'fusion' in tonal or common practice music which is not perceived in music based upon alternate tuning systems such as micro-tonality. The answer may lie in the fact that when primary intervals of the tonal system (octaves, fifths, thirds, minor sevenths) are played by instruments whose spectra are composed of frequencies failing in the harmonic series, there occurs common frequencies, or frequencies which are very close, in the composite spectra which may explain the harmonic fusion. For example, when two instruments play at an interval of an octave, the frequency components of the upper instrument form a subset of the lower.

Given the possibility in the digital domain of ordering inharmonic as well as harmonic spectra, we may hypothesize that the harmonic series in relation to the tempered tuning is a naturally ocurring special case which can be generalized to include alternate tuning systems where the spectral pattern is adjusted to be similarly complementary.

Summary

Extending the FM technique to include multiple modulating waves extended the degree of control over the spectrum - while maintaining the desirable conceptual and computational simplicity. The FM technique also led to the discovery of a class of non-linear synthesis techniques which has further advanced our ultimate goal of achieving the simplest mathematical models for complex perceptual events. An area of research which is dependent upon the capability of precise control of frequency components has to do with the relationship of the spectrum to the tuning system in homophonic and polyphonic musical contexts.

Reverberation Studies

Work continued on new methods of economical synthesis of cle,-,n, smooth reverberation. It is, of course, possible to just simulate a specific concert hall by measurement of the impulse response and direct convolution. The amount of computation required for such a convolution for a typical 2-second reverberation time is enormous. **For this reason, work has** been proceeding on new, simple and concise methods of reverberation. We have lately been concentrating on the all-pass or "uncolored" reverberation on the grounds that this then separates the reverberant qualities of a room from its coloration. In this manner, we can study independently these effects.

Advances in Unit Reverberator Design

The 1-multiply allpass. When the allpass unit reverberator was first introduced by Schroeder [1961a, 1961b], it required three multiplication operations for computation. This was subsequently reduced to two multiplies by Moorer [1975, 19771 and others. Moorer also introduced a higher order allpass network capable of generating an undulatory impulse response [Chowning et al., 1975; Moorer, 19771. It is now clear that the allpass network can be realized in lattice form with only one multiply, and that the higher order form can be realized with two multiplies. Figure I shows the..block diagrams of the simple allpass. Calculation will show that this is indeed an allpass network.

The frequency dependent allpass. With existing concert halls, it is clear that the reflection coefficients of the walls must be dependent on frequency. This being the case, it is interesting to ask if we can simulate this effect while still preserving the all-pass nature of the reverberator.' At first, this might seem like a contradiction in terms, but all it really means is that the poles and zeros of the transfer function should be arranged in something other than perfect circles in the complex plane. This can be accomplished by the filter structure shown in Figure 2.

It can be shown that the transfer function of this filter is the following:

TW = Z` 0 -GZm)
0 +HZ-m)

X(n) IS

Y(n) -Q

+ _M

Z

FIGURE 1. Flow diagram of the one-multiply allpass unit reverberator. It can easily be shown that this realizes an allpass transfer function. This unit uses one multiply and three additions, whereas the previous simplest form used two multiplies and two additions.

X(n)

H

Z _M +

--0.9

00. Y(n)

FIGURE 2. Flow diagram of the frequency-dependent allpass unit reverberator. If H and G are complex conjugates, this is an allpass network. For stability, the zeros of H and the poles of G must be inside the unit circle.

For this to be an allpass, G and H must be complex conjugates. This would require that the powers of Z be inverted (i.e., replacement of Z for Z-1). Reversing the coefficients of a polynomial inverts the roots, so that any roots that were inside the unit circle are now outside and vice versa. Unfortunately, it is very difficult to determine the stability of the resulting network as a whole. This can only be practically determined by trial and error, since the polynomials involved are typically of such high order (1000 to 3000 usually for realistic sounding reverberators). The network can always be stabalized by dividing each coefficient by (xi where a is greater than magnitude of the largest root. This will bring the roots inside the unit circle. The magnitude of the largest root can be determined by examining the impulse response. We can make an approximation in which the largest root dominates the impulse response in the unstable case and measure the time constant. This gives us directly the magnitude of the largest root.

Since H must be entirely realizable, we must be careful that the transfer function involves only negative powers of Z. Let us express the transfer. functions for **H and G in the following** form:

$$HW = ZM\,''' \frac{\overset{N}{Z}\ airi}{\underset{i=O}{\overset{M}{2;}}\ bi\ 'Zi} \qquad G(z) = 2\text{-}M\,':\text{-}' \frac{\overset{N}{Z}\ aizi}{\underset{i=O}{\overset{M}{Z}}\ bi\ ri}$$

From the formula for H we can see that the only ai that may be non-zero are those for which i>M and still preserve realizability. The formula for G will have positive powers of Z in the numerator for i>M. Although this would imply that we have to anticipate incoming values of X(n), we can easily simulate this by merely delaying X(n) by the proper number of samples.

Currently we are working to determine what choices of filters H and G in the frequency-dependent reverberator give pleasing and musically interesting reverberation.

Interactive Reverberator Compiler

A highly flexible reverberator compiler has been developed to aid in the simulation of reverberant environmnts. The program calculates reverberation parameters, generates and displays the impulse response of the reverberator as each unit reverberator is defined, and generates. reverberator code for the music compiler and the program that reverberats digitally recorded sound. Comparing the resultant impulse response to that of real spaces aids in the design of realistic sounding reverberation.

The CCRMA Digital Recording Studio

Each area of research and musical creation at CCRMA relies on high quality sound production and reproduction for its validity. We therefor e found it necessary during the grant period to develope an all-digital recording and sound processing system: a combination of computer software and digital **hardware which can provide high quality sound** recording and processing without the noise and handling constraints of analog tape systems.

High quality digital recording has become possible due to the increasing availibility in recent years of large high-speed storage devices and advances in the state-of-the-art engineering. The primary effort in digital recording has been from Thomas G. Stockham, Jr., and Richard B. Warnock of Soundstream, Inc., Salt Lake City [Stockham, 1972; Warnock, 19761, and CCRMA at Stanford [Rush et al., 1976; Moorer, 1977; Rush and Moorer, 19771.

With all-digital recording, the acoustical signal is transmitted from the microphone to an analog-to-digital converter and is stored in digital form on a large disc system, after which the digitized sound can be auditioned through a digital-to-analog converter and loudspeakeri or earphones. Once in digital form, the signal can be processed by a number of digital techniques, to be described further on. High quality digitized sound has applications in musical recording, music composition, preparation and presentation of psych oacou stical stimuli, and the permanent archiving of recorded sound.

Every audio signal used at CCRMA exists, at some time, in the digital domain. For example, live sounds are converted to and stored in digital form, whereas synthesized sounds are created directly in digital form. It is therefore of primary importance to acheive, effectively noise-free analog-to-digital (A/D) conversion for recording of the real sounds and digital-to-analog (D/A) conversion for playback of both the real and the synthetic sounds.

If an analog audio tape system is introduced at any point in this process, the signal degeneration is immediate and irreversable. The degeneration is caused primarily by tape hiss and dropout (irregularities in the oxide coating). If any processing of such material is desired, its use must constantly be weighed against the knowledge that, at every step, more noise will be added to the signal. Tape editing is a handicraft that can rarely match the accuracy with which acoustical signals are perceived. Equ ally discouraging is the nagging awareness that from the moment the signal is recorded -on analog tape it has begun to degenerate. These and other problems associated with standard analog recording can either be totally eliminated, reduced to inaudibility, or significantly improved upon by means of all-digital recording. Effectively noise-free recording is now a reality, the first time in the history of recorded sound.

In addition to the possibility of effectively noise-free recording, there are powerful advantages to recording an audio signal in the digital domain. The resources of the recent explosive development of digital signal processing techniques are now applicable to noise-free processing of recorded sound (Moorer, 1977). We have developed editing programs which are accurate down to the level of the individual sound sample, where modifications and new techniques can be applied as needed. The very nature of digital signals permits long-term storage without significant degradation of the signal. With appropriate coding, to be discussed later, it is now possible to preserve indefinitly those rare recorded performances, electronic music compositions, and particularily useful or significant psych oacou stical stimuli which in the past have been lost in the noise of long-term analog storage.

Having examined our digital recording goals we find that they include, in addition to functions which are unique to our research, most of the techniques that are currently available in commercial recording studios. We find ourselves, therefore, in the process of working to establish a highly automated, multi-purpose facility which will serve as a prototype for the commercial recording studio of the future.

Recording Facility and Computer Hardware

The CCRMA Digital Recording Studio consists of an acoustically treated room with microphones and line amplifiers, audio playback equipment, and a silent Computer terminal. The signal from one or more performers is picked up by the microphones, amplified, and transmitted differentially ("balanced") to the AID conversion equipment of the computer.

The conversion equipment produces a quantized sampled-data representation of the microphone signal which is then stored on a large high-speed storage device, in this case magnetic disks. The sampled data can then, under program control, be retrieved from the storage medium, fed to a D/A converter, sent differentially back to the recording studio, and played, either over speakers or earphones. The operator is responsible for starting and stopping conversion, and the ordering and labeling of the takes.

Since computers are notoriously noisy, both electrically and acoustically, the recording studio is physically isolated from the computing facility. The nearest audible source of noise is over 200 feet and several rooms away. The audio equipment in the recording studio was connected to the computer by 500 feet of foil shielded twisted pair audio cable. There is one line for each channel of the A/D and D/A converters. The output of B&K instrumentation microphones is boosted and then ballanced by Signetics T074 operational amplifiers which are configured so as to have a push-pull output that eliminates the need for * ballancing transformers. The outputs of the D/A converter are similarly treated. The ballanced signals are received at the A/D and in the studio by specially built active receivers which have very high common mode rejection, which again eliminate the need for transformers. The dynamic range of the CCRMA system was measured from the microphone input to the data in the computer memory at 75 dB. This is a significant improvement over typical recording studios. First approximation sound levels are currently taken from an L.E.D. Peak Amplitude Indicator. Subsequent levels are set using, the recording program. The monitoring equipment consists of Phillips RA544 amplifierloudspeakers.

The conversion system was designed around the Analogic MP2914A A/D converter and the Datel HR-16B D/A converter. The D/A converter was time-division multiplexed four ways to provide four channels of output. The multiplexing was done using Philbrick 4853 sample-hold units, then filtered with TTE J77B low-pass filters. A variety of sampling rates are allowed, the mos t popular being 25.6 KHz and 51.2 KHz per channel. The filters are selected automatically to correspond to the sampling rate. The power distribution system is carefully isolated for each channel so that crosstalk caused by inductive and capacitive coupling is minimized. No special digital coding scheme was used, nor was any dynamic range enhancement attempted (such as floating point conversion or post-scaling).

RecordlPlayback Program

A program was especially written for the purpose of digital recording, storage management and playback. It aids in performing actions typical of recording sessions: start a "take," abort or confirm 'a take, keep track of good takes. In order to accomplish this the program does the following:

Storage management on the mountable disk pack. Each take is given a name (called the "file" name). There is a "directory" written onto a dedicated cylinder of the pack that tells what files exist, what their names are, what cylinder they start on, and how many cylinders they occupy, as well as what cylinders are free and which are claimed. All allocation is done in terms of integral numbers of cylinders. All files occupy consecutive cylinders for maximum reading and writing speed.

Negotiations with the time-sharing monitor. Since many other people use the computer while a recording session is in progress, the program must make special arrangements with the monitor for long periods of uninterrupted service. The Stanford monitor has many provisions built in for this purpose.

Interface with the user. The program interfaces with the user, or operator, to provide creation, renaming, and deletion of files, setting up takes, calculating peak amplitudes and amplitude histograms, allowing the user to abort or truncate files at will.

At the beginning of a 'session, generally the first thing one must do is clear the directory on the disk pack. Since this is a somewhat drastic measure, it requires two different stages of verification by the user. Usually the next thing is to get a level of some kind. To do this, we just make a recording. A command is typed to start a take. The program asks for a file name and a maximum duration. It looks for enough free consecutive cylinders to contain that maximum time, enters the file name in the directory, locks itself into main memory, and waits for the "go" signal from the user. The user cues the performer and presses a key on the terminal simultaneously and the recording is begun. When the performer stops, the key is again pressed and the program ceases recording at the end of the current cylinder. The program then asks "was that a take?". If the user answers "no", then the recording location on the disk pack is repositioned and ready for the next record command. If the user answers "yes", then the directory entry is altered to reflect the true (possibly truncated) length of the file. This ensures that only the "takes" are preserved on the disk pack without wasting space on unusable starts.

Commands can then be typed which compute the peak amplitude of the file or which compute the amplitude profile (histogram) of the file. The histogram seems to be the most useful, because it tells us exactly how much time the signal spent at any given level. Once we have the level set properly, it need not be changed again (unless the program material changes drastically).

Once the recording session is complete, the user can transfer the files to the normal file system of the time-sharing monitor for further processing.

Editing, Processing and Mixing Digitized Audio Waveforms

An interactive computer-based system for the editing, mixing and processing of digitized audio waveforms has been implemented. A variety of useful techniques for manipulation of stored digital "sound files" have been developed. These include automatic segmentation (locating the desired segments in a "take" and deleting false starts, inadvertent performer's noises between phrases, etc.), editing and splicing (including "micro-surgery" down to the sample), amplitude shaping, artificial reverberation, simulated location and movement of the sound source in stereo or quadraphonic space, channel mixing and splitting, retuning and cross-syn thesis.

The Sound File Editor

With any recording system, the task of editing the "raw" recording tends to be laborious. A 11 take" will include unwanted silences before and after the desired sound. It may also include false starts, short retakes, and performer's noises between phrases. In the past, one of a good recording engineer's most prized skills has been the ability to quickly edit such a take into an acceptable master source. For digital recording, we have developed software which aids in the editing of such raw takes and automates the process wherever possible. For example, the program is given criteria for deciding which segments should be listed as of possible use. The user can then listen to the segments presented and note which should be preserved or change the criteria for a consequent location pass. The user is able to display, audition and copy any chosen portion from the file. The segmentation criteria and history is stored for use in a later editing session, or in case the criteria may prove useful for a similar take.

The ability to edit down to the sample has proven most useful. With even the finest musicians, there are times when the most musical take has, for example, a false attack which would render it useless if it could not be corrected. With the interactive sound file editor it is an easy matter to eliminate the false attack by moving pointers along the displayed waveform and listening to the edited result.

Another useful editing feature is the ability to shape the amplitude envelope of a selected portion of a sound file. This gives the user the ability to change the musical emphasis of the sound segment, either for variety, to get a more appropriate version, or because an *impossible* musical phrasing is desired (such as beginning a trombone tone from silence without any perceptible attack). . -

Retuning

Through the use of digital interpolation and decimation [Crochiere and Rabiner, 1975; Moorer, 19771, the pitch of a recorded sound can be adjusted to a very fine level. The procecure is to adjust the effective sampling rate of the signal, so that when it is played at the original sampling rate, the pitch (and duration) of the signal is changed. This technique was implemented for the purpose of the careful retuning of a recorded sound. In such cases the variances in duration are too small to be audible. The ability to retune a sound segment is very important. It allows us, for instance, to correct a mistuned note in the middle of an otherwise useful take, or to construct new or unusual tuning systems from recorded sources which used some other system.

Cross-synthesis

We have also implemented a technique which is based on linear predictive analysis [Makhoul, 19751 as developed by Tracy Peterson [1975; also see Moorer, 1977]. In this method a musical signal is used as the excitation function for a time-varying digital filter which was, for example, modeled after a speech waveform. This produces a strikingly unusual music in which the source music is heard as if it were singing the words of the speech signal.

Implementation of Existing Techniques

Artificial reverberation and the simulation of movement,of sound through illusory space have been discussed in earlier sections of this proposal as these techniques apply to synthesized sounds. We have recently implemented these techniques in the processing of recorded sound, a sumary of which is included in Moorer [19771.

We often use digital filters, including time-varying filters, as tools in a variety of processing techinques, including techniques invented for the occasion. For example, one might wish to track and emphasize the harmonic that is loudest at any point in a musical segment. This could be accomplished by putting the signal through a pitch detector set to return a map of the dominant frequencies through time, and using this map to set the peak of time-varying bandpass filters.

Another much used technique, actually a synthesis technique, is the phase-vocoder analysis-based synthesis discussed earlier. Since the an alysis-syn thesis loop in this technique is accurate enough that trained musicians cannot distinqUish between real or synthetic tones, this has proven extremely useful in accomplishing transformations in the wa.veform of" recorded instruments which would not otherwise be possible, such as a gradual timbral transformation from a violin to a clarinet [Grey, 19-751.

Digital Mixing

After the Sound file editor the next most used sound processing program is the automated diaital mixer. In addition to the advantage of the noise-free nature of digital mixing, it has the advantage of great accuracy and ease of operation. The mixer reads a mixing list which gives sound file names, starting times, amplitude factors, and channel allocations. This list can be typed in a file by the user or it can be generated by using the program which was written to translate from musical notation to sorted input for the music compiler [,Smith, 19721. Such lists may contain hundreds of occurances of sound files which are mixed in any combination.

The mixer is also used to create unusual musical textures which were previously impossible to control. By delaying and overlaying many different copies of the same sounds, we can produce Musical textures that can vary from a single instrument to masses of independently sounding instruments. Such a technique, carefully correlated, can be used to produce choral textures from a small number of digitized se ments. In addition to the much prized

automated features of the digital mixer, the timing of various voices being overlayed can be adjusted interactively using graphical display and/or audio presentation, and can be adjusted down to the sample level.

Bibliography

Crochiere, R. E., and Rabiner, L. R. *Optimum FIR Digital Filter Implementations for Decimation, Interpolation, and Narrow-Band Filtering.* IEEE Transactions on Acoustics, Speech, and Signal Processing Vol ASSP-23, 444-456, October (1975).

Makhoul, *J. Linear Prediction: A Tutorial Review.* Proceedings of the IEEE 63, 561-580, April (1975).

Schroeder, M. R. *Improved Quasi-Sterophony and Colorless Artificial Reverberation. J.* Acoust. Soc. Amer. 3 3, 1061, (196 1 a).

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

Smith, L. *C. Score, A Musician's Approach to C,;mputer Music.* J. Audio Eng. Soc., January (1972).

Stockham, T. G., Jr. *AID and DIA Converters: Their Effect on Digital Audio Fidelity. AES* preprint 834 (1971). Also available in *Digital Signal Processing.* IEEE Press, Rabiner and Rader , eds., (1972).

Warnock, R. *B. Longitudinal Digital Recording of Audio. AES* preprint 1169,(1976).

Publications and papers resulting from contract

Grey, J. M. *An Exploration of Musical Timbre.* Ph.D. Dissertation, Stanford University Report STAN-M-2, 133pp, 1975.

Grey, J. M. *Multidimensional Perceptual Scaling of Musical Timbres.* journal of the Acoustical Society of America, May, 1977.

Grey, J. M. and Moorer, *J. A. Perceptual Evaluation of Synthesized Musical Instrument Tones.* journal of the Acoustical Society of America, August, 1977.

Grey, J. M. and Gordon, J. W. *Perceptual Effects of Spectral Modifications on Musical Timbres.* journal of the Acoustical Society of America, April, 1978.

Grey, J. *M. Timbre Discrimination in Musical Patterns.* journal of the Acoustical Society of America, to be published in August, 1978.

Grey, *J. M. Experiments in the Perception of Instrumental Timbre.* invited by the Bulletin of the Council for Research in Music Education, to be published sometime in 1978.

LeBrun, *M. Wave Shaping Synthesis. To* be published in the journal of the Audio Engineering Society.

Moorer, *J. A. On the Segmentation and Analysis of Continuous Musical Sound by Digital Computer.* Stanford Univ.. Dept. of Music Tech. Rep. STAN-M-3, 1975.

Moorer, J. *A. The Use q the Phase Vocoder in Computer Music Applications.* Presented at the ~f

55th Convention of the Audio Engineering Society available as Preprint number 1146 (EI), 1976.

Moorer, J.A., *The Synthesis of Complex Audio Spectra by Means of Discrete Summation Formulae.* journal of the Audio Engineering Society, Volume 24, *9, November 1976, pp7l7727.

Moorer, J.A., *Signal Processing Aspects of Computer Music - A Survey.* Proceedings of the IEEE, July, 1977.

Invited Paper,

Rush, L., Moorer, *J. A.,* and Loy, G. All-Digital *Sound Recording and Processing.* presented at the 55th Convention of the Audio Eng. Soc., New York, 1976.

Rush, L., and Moorer, J. *A. Editing, Mixing and Processing Digitized Audio Waveforms.* presented at the 56th Convention of the Audio Eng. Soc., Los Angeles, May (1977).

Schottstaedt, *B.The Simulation of Natural Instrument Tones using Frequency Modulation with'a Complex Wave*. Computer Music journal, Vol 1, *4, 1977.

41

Theses

Moore, *F.R.,Real Time Interactive Computer Music Synthesis.* Phd. dissertation in Electrical

Engineering.

Scientific collaborators

J.M. Chowning	Adjunct Professor
J.W. Gordon	Graduate Student
J.M. Grey	Reserach Associate
M. LeBrun	Guest Researcher
D.G. Loy	Graduate Sutdent
M. MacNabb	Graduate Student
F.R. Moore	Graduate Student
J.A. Moorer	Research Associate
L. Rush	Research Associate
W. Schottstaedt	Graduate Student
L.C. Smith	Professor

Co-principal investigators

Leland C. Sinith

Avlof W

John M.Chowning

)jAA