

Timbre Space as a Musical Control Structure

Author(s): David L. Wessel

Source: *Computer Music Journal*, Vol. 3, No. 2 (Jun., 1979), pp. 45-52

Published by: The MIT Press

Stable URL: <http://www.jstor.org/stable/3680283>

Accessed: 13-02-2018 21:44 UTC

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <http://about.jstor.org/terms>



JSTOR

The MIT Press is collaborating with JSTOR to digitize, preserve and extend access to *Computer Music Journal*

Timbre Space as a Musical Control Structure

David L. Wessel
IRCAM
31, Rue St. Merri
75004 Paris, France

Introduction

Research on musical timbre typically seeks representations of the perceptual structure inherent in a set of sounds that have implications for expressive control over the sounds in composition and performance. With digital analysis-based sound synthesis and with experiments on tone quality perception, we can obtain representations of sounds that suggest ways to provide low-dimensional control over their perceptually important properties.

In this paper, we will describe a system for taking subjective measures of perceptual contrast between sound objects and using this data as input to some computer programs. The computer programs use multidimensional scaling algorithms to generate geometric representations from the input data. In the timbral spaces that result from the scaling programs, the various tones can be represented as points and a good statistical relationship can be sought between the distances in the space and the contrast judgments between the corresponding tones. The spatial representation is given a psychoacoustical interpretation by relating its dimensions to the acoustical properties of the tones. Controls are then applied directly to these properties in synthesis. The control schemes to be described are for additive synthesis and allow for the manipulation of the evolving spectral energy distribution and various temporal features of the tones.

** Taped examples of various timbral manipulation schemes discussed in this article are available from IRCAM; get cassette order information from IRCAM at the address above.*

Tests of the control schemes have been carried out in musical contexts.* Particular emphasis will be given here to the construction of melodic lines in which the timbre is manipulated on a note-to-note basis. Implications for the design of human control interfaces and of software for real-time digital sound synthesizers will be discussed.

Musical Timbre

Timbre refers to the “color” or quality of sounds, and is typically divorced conceptually from pitch and loudness. Perceptual research on timbre has demonstrated that the spectral energy distribution and temporal variation in this distribution provide the acoustical determinants of our perception of sound quality (See (Grey : 1975) for a thorough review). With one notable exception (Erickson : 1975), music theorists have directed little attention towards the compositional control of timbre. The primary emphasis has been on harmony and counterpoint. The reason for this probably lies in the fact that most acoustical instruments provide for very accurate control over pitch but provide little in the way of compositionally specifiable manipulation of timbre. With the potential of electroacoustic instruments the situation is quite different. Indeed one can now think in terms of providing accurate specifications for, by way of example, sequences of notes that change timbre one after another. It is this note-to-note change of timbre that will concern us in this paper.

Synthesis Technology

Digital technology offers powerful, general, and flexible sound synthesizers. A number of such synthesis machines have already been constructed and are producing musical results. Notable examples include the digital synthesis and processing system designed by Peter Samson (1977) now in operation at the Stanford Center for Computer Research in Music and Acoustics; the 256 digital oscillator bank designed and constructed by G. DiGiugno (1976) and the digital synthesizer of Hal Alles and DiGiugno (1977), both in operation at IRCAM; the Alles (1977 a, b) synthesizer at Bell Labs; and the Dartmouth digital synthesizer (Alonso *et al.* : 1975). Some of these devices offer the alluring possibility of using a "brute force" additive approach to the synthesis of complex and musically rich time-variant spectra. In this report we will concentrate on this form of additive synthesis, because of its generality, and with the accompanying problem of providing direct control over the perceptual properties of the synthesized sounds.

Before beginning a description of a procedure for developing controls that could facilitate the musically expressive manipulation of complex time-variant spectra, we will examine the nature of both the acoustical and perceptual data bases involved. Additive synthesis requires a considerable if not overwhelming amount of explicit information, and we shall explore ways to reduce this quantity of data without sacrificing richness in the sonic result. On the other hand, the data that we can obtain about our perceptual experience of timbre has quite a different character from the physical data of acoustics, and so we shall also examine such notions as subjective scales, perceptual dimensions, and structural representations of subjective data. We shall then see the extent to which we can give an account of the subjective experience by examining the relationship between the acoustical and the perceptual data bases. We seek a psychoacoustics of timbre that has implications for timbral control in musical contexts.

Additive Synthesis and Possibilities for Data Reduction

In the additive model for sound synthesis, a tone is represented by the sum of sinusoidal components, each of which has time-varying amplitude and frequency. Moorer (1977) gives an excellent account of the details. To synthesize a sound, one specifies a number of software or hardware sinusoidal oscillators, each with its own amplitude and frequency control envelopes. Additive synthesis of this form has two important advantages. First, it is general, that is, with a sufficient number of independently controllable oscillators a very large and highly varied class of signals can be generated. At some sacrifice in computational efficiency and with an increase in the quantity of data for specifying the envelopes, one can mimic FM synthesis (Chowning : 1973) and other non-linear techniques (Arfib : 1977; Beauchamp : 1975; Le Brun : 1979; Moorer : 1976). Of course, one can produce effects not possible with these techniques. Second, one can analyze existing sounds and obtain data that can be used to resynthesize exactly the signal that was analyzed. Moorer (1976) has described the application of the phase vocoder (Portnoff :

1976) to the analysis and synthesis of musical sounds. The phase vocoder is an advance over methods like the heterodyne filter (Beauchamp : 1969; Moorer : 1975) in that the analysis does not have to be pitch synchronous nor do the waveforms have to contain more or less harmonic components, thus permitting the analysis of tones with pitch variation as well as inharmonic tones like those produced by percussion instruments. The method also guarantees that when the analysis data is not modified, the original signal is recovered exactly.

Let us say that we want to synthesize, using data from phase vocoder analysis, a musical instrument timbre with 25 harmonics. In this case 25 amplitude and 25 frequency envelopes will be required. Storing these functions in full detail demands considerable memory, and if we are using a computer-controlled digital synthesizer like those mentioned at the start of this article, then transfer of the envelopes may exceed the bandwidth of the link between the computer and the synthesizer. Furthermore, if the shapes of the envelope functions are to be modified, the computation time required to rescale every point of the functions may easily exceed the capabilities of real-time manipulation. Clearly some form of data reduction is required, if one wants to work in real time.

A particularly attractive procedure that produces a significant reduction in the quantity of data involved is to approximate curvilinear envelope functions with functions composed of a series of straight line segments (such as those given in (Moorer : 1977)). Such straight-line-segment approximations can be stored in terms of the coordinates of the break points of the function, thus greatly reducing memory demands. Furthermore, the digital synthesizers constructed by Alles, DeGiugno, and Samson provide digital oscillators that include straight-line ramp controls for both amplitude and frequency. In working with these oscillators, one provides as data the starting value for a ramp, its slope, and a terminating value or time that indicates when a new slope is required. The actual generation of the values along the specified line segments is provided within the oscillator itself. In supplying control data from the computer's memory to these synthesizers, the break points of the line-segment approximations can be passed directly from the computer to the synthesizer, thus greatly reducing the data rate demands on the interface. Finally, the straight-line-segment approximations make possible rapid modification of the function shapes, as only the coordinates of the break points need be modified. But can we get away with such a drastic data reduction? If we approximate curvilinear functions with a small number of connected straight lines, will high audio quality and timbral richness be maintained?

Indications that the straight line segment approximations would provide satisfactory results have been obtained for brass tones described in (Risset and Mathews : 1969) and by (Beauchamp : 1969). Grey (1975) carried out a carefully controlled perceptual discrimination experiment to determine the extent to which the tones with completely detailed amplitude and frequency functions could be discriminated from those with line-segment approximations consisting of from five to seven segments per envelope (also Grey and Moorer: 1977). Grey used 16 different orchestral instrument tones, and in general he found the discriminations to be extremely difficult. His findings strongly suggest that it is not necessary to retain the highly complex temporal microstructure of the amplitude and frequency functions in order to preserve timbral quality. It would appear that the line-segment

approximations can be made with little harm. Besides the resulting data reduction, an important advantage of such approximations is that the resulting tones have more clearly defined acoustical properties. This will prove especially important when we wish to determine those physical properties that are especially important for perception.

Representation of Timbre Dissimilarities

In the next sections we will first describe a method for characterizing the structure of the relationships inherent in a set of sounds differing in timbre. We will then show how such representations of the timbres can be related to the underlying acoustical properties of the tones. We will also show that the representations can be used to compose timbral patterns with perceptual properties predictable from the structure of the representation. Finally it will be argued that when this is accomplished we have in some sense designed a systematic control scheme for the perceptually important acoustic properties of the sounds.

From a quantitative point of view, data from subjective judgments has a peculiar if not uncertain status (Luce : 1972). The notion of a unit of measurement such as the decibel or hertz is difficult if not impossible to establish for subjective scales. We can of course choose units for subjective scales like "sones" or "mels" as Stevens (1959) has done, but the so-called unit derived in one experimental context fails to remain fixed in other contexts. In fact, the "sone," the unit for subjective loudness, is not invariant across the two ears of individuals with normal hearing (Levelt, *et al.* : 1972). Such units are useful in that they provide a common language for discussing the auditory abilities of a population of listeners, but they cannot justifiably be treated with the algebra of dimensional analysis that underlies measurement in the physical sciences. I think it right to be pessimistic about the possibility of subjective scales being elevated to the same form of measurement as physical measurement. But subjective judgments, if collected over a sufficient number of objects, in this case sounds, can have a representable structure, and this structure can in turn be related to various acoustical parameters.

Perceptual judgments tend to be relative in nature. With few exceptions, we tend to judge an object in terms of the relationships it has with other objects. Relational judgments are of great interest in music, because music involves patterns composed of a variety of sounds, and it is the relational structure within and between the patterns that is of primary importance. Judgments of the extent of perceptual similarity or dissimilarity between two sounds can be made in a very intuitive fashion. One can say that sound *A* is more similar to sound *B* than to sound *C* without having to name or otherwise identify explicitly the attributes that were involved in the judgment. Research groups at IRCAM, at Michigan State University, and at the Stanford Center for Research in Music and Acoustics have been using perceptual dissimilarity judgments in a variety of musical and otherwise audio-related contexts. One of the general techniques is to represent the perceptual dissimilarities as *distances* in a spatial configuration. One begins with a set of dissimilarity judgments typically taken for all pairs of sounds that can be formed from the set. This matrix of dissimilarities is then processed by one of a variety of multidimensional scaling programs such as KYST (Kruskal : 1964 a,b). The multidimensional scaling

programs produce an *n*-dimensional spatial arrangement of points that represent the various sound objects. The programs operate to maximize a goodness-of-fit function relating the distances between the points to the corresponding dissimilarity ratings between the sounds.

Perhaps at this point it would be best to show how the multidimensional scaling experiments were carried out. At IRCAM we have recently developed a set of programs that greatly facilitate the design, execution, and interpretation of such experiments. In the following example we use the same set of sounds used in both (Grey : 1975) and (Grey and Gordon : 1978) and presented in the series "Lexicon of Analyzed Tones" (Moorer, *et al.* : 1977, 1978). This set consists of 16 synthetic orchestral instrument timbres and a group of eight hybrid instrument timbres produced by exchanging spectral envelopes between members of the original set. The goal of our experiment will be to provide a representation of these 24 timbres as points in a Euclidian space as well as an interpretation of this representation in terms of the acoustical properties of the tones.

General Method

The procedure that provides an interpreted representation of the sounds involves the following five steps: 1. Selection of materials for study; 2. Collection of the dissimilarity judgments; 3. Representation of the dissimilarity judgments with spatial and other schemes such as clusters and graphs; 4. Psychoacoustical interpretation of the structure; 5. Verification of the interpretation in musical situations.

Selecting the Materials for Study

In preparing the sounds that we wish to represent, attention must be paid to (A) the number of sound elements, (B) the problem of equalizing the sounds with respect to parameters we wish to ignore, and (C) the range of variation within the set of sounds.

A. The Number of Sound Elements

To obtain a meaningful representation, a certain minimal number of sound elements is required, in order that the dissimilarities impose a sufficient amount of constraint for fixing accurately the locations of the points in a space. Some of the multidimensional scaling programs, like KYST, operate with only qualitative or "ordinal" constraints on the distances in the space. These algorithms seek arrangements of the points such that the rank order of the interpoint distances in the space matches, in terms of a well-defined goodness-of-fit measure, the rank order of the dissimilarities between the corresponding stimuli. If we begin with interpoint distances from a known configuration of points, the programs provide a very accurate recovery of the positions of the points even when the distances are subjected to radical monotonic transformations or perturbations due to random error (Shepard: 1966). It is hard to set down a precise rule of thumb for determining the minimum number of elements to use. The choice depends on the number of dimensions and the distribution of the points in the space, an issue to which we shall return below. In most psychological research using multidimensional scaling, 10 points have seemed sufficient for two dimensions and 15 for three. In our recent research we have tended to encourage the use of between 20 to 30 points for spaces of two and three dimensions, respectively.

B. Equalizing the Tones with Respect to Extraneous Parameters

If possible, the tones should be equalized with respect to the properties that are not to influence the judgments. When studying timbre, the usual procedure is to equalize the pitch, subjective duration, loudness, and room information aspects of the tones. On the other hand, if we are studying room information (that is, reverberation structure), then we probably want to use a standard source and manipulate only the reverberation parameters. Attention should also be paid to just what is being equalized. If we are equalizing with respect to loudness, for example, and the tones in the set have different spectral shapes and attack rates, then simply matching sound pressure level (in terms of decibels) will not provide the appropriate equalization. In this case we are without a satisfactory model for the perception of the loudness of complex time-variant spectra (Moorer: 1975) and we must resort to making empirical matches. Grey (1975) provides a good example of such subjective matching procedures for pitch, duration, and loudness.

C. Controlling the Range of Variation Within the Set

The range of variation in the timbres of the sounds will certainly differ from one type of study to another. In some instances we will want to investigate a timbral domain having a broad range of variation including perhaps inharmonic percussion sounds as well as sounds with more or less harmonic components. In other situations more restrictions are imposed on the range of variation, as in, for example, the study to be described in this paper where we use only sounds derived from standard orchestral instruments played in a conventional manner. For even more refined investigations of timbral nuance one might use a very limited range of sounds.

Once a general range of variation has been determined, some attention must be paid to the homogeneity of variation within the set. Consider the following example. If we choose eight distinctively percussive timbres and eight distinctively non-percussion timbres but provide no linking elements between the two domains, then it is likely that all the subjective dissimilarities that are made for pairs that cut across the two classes of sounds are larger than all the intra-class dissimilarities. In situations like this, some of the multidimensional scaling programs give what are called degenerate solutions and the intra-class structure is not fully displayed. Shepard (1974) provides a detailed discussion of this problem and prospects for its solution. In addition, in such a situation I find that in making the judgments I have a difficult time concentrating on the relatively subtle differences for pairs within a class in the context of the much larger dissimilarities for the pairs spanning the two categories. Unfortunately, since we will most often deal with timbral domains about which we have little knowledge, clear rules for the preliminary selection of the variation in the material are hard to set down. Selection procedures depend ultimately on our specific interests and desires for control.

To facilitate the preliminary screening and selection of the sounds, we have developed an interactive random access audio playback program on IRCAM's DEC-10 computer. This program is called KEYS and was written by Bennett Smith (Wessel and Smith : 1977). One supplies KEYS with a list of the sounds written out as a list of the names of the files

in which each sound element is stored. Each sound file in the list is then related to a character that can be typed on the terminal keyboard. The relationships between the characters and the sounds are listed on the computer terminal's CRT display. When a character is typed the corresponding sound file is played through the digital-to-analog converters and the playing action is indicated by an increase in the brightness of the character-file name entry in the table displayed on the CRT. This program permits rapid auditory comparisons among a large number of different sounds. Currently the limit is 100 files of arbitrary length.

Collecting the Timbre Dissimilarity Judgements

Though there exist a variety of ways to collect perceptual dissimilarities, we have found simple ratings of the extent of dissimilarity to be the most efficient and least tedious way to make the judgments. At IRCAM we have been using direct estimates of the dissimilarity between two tones. Our listening judge, using a program written by Bennett Smith called ESQUISSE, sits before a CRT display terminal and an audio system fed by the computer's digital-to-analog converters. The two sounds in a pair are related to the terminal keys "m" and "n", allowing the listener to play the sounds at will. After listening, the judge enters a rating with one of the keys "0" through "9" on the terminal. Immediately after the judgment is entered, the next pair of tones is ready to be played from the keyboard. The sequencing of the judgment trials is random, and all of the $n(n-1)/2$ pairs are used, where n is the number of sounds under study. After all the pairs have been judged, the random sequence is unscrambled and a matrix of dissimilarities is formed.

The data collection program includes what we call a "coffee-break" feature that allows the judgment session to be interrupted either by machine failure or human fatigue. The listener can then return to the experiment at a later time and continue from the point in the sequence where the interruption occurred. With this program, one is able to listen to the sounds rapidly and freely, and the judgments can be entered with ease.

A Two-dimensional Representation of 24 Orchestral Instrument Timbres

I served as a judge using the dissimilarity data collection program on a set of 24 orchestral instrument timbres that were obtained from John Grey. These sounds were synthesized using line segment envelopes and were equalized subjectively for pitch, loudness, and duration. A two-dimensional representation of the sounds provided by the KYST program is shown in Figure 1, along with an interpretation of the dimensions of this space. The vertical axis is related to the *spectral energy distribution* of the tones, and the horizontal, to the *nature of the onset transient*. The sounds at the top of plot are bright in character, and as one moves towards the bottom the timbres become progressively more mellow. In a number of studies on timbre spaces (Weden and Goude: 1972; Wessel: 1973; Grey: 1975; Ehresman and Wessel: 1978; Grey and Gordon: 1978; Wessel and Grey: 1978) this dimension related to the spectral energy distribution has appeared. A consistent, quantitative, acoustical interpretation has also been provided by calculating an excitation pattern for the spectrum provided by Zwicker's model for loudness (Zwicker and Scharf: 1965). This trans-

formation on the acoustical spectrum compensates for certain properties of the auditory system like critical bands and the asymmetric spread of masking from low to high frequencies. The centroid or mean of this compensated spectral energy distribution is then calculated and correlated with projections of the points on the axis assumed to be related to brightness. In all of the studies these correlations have been very high.

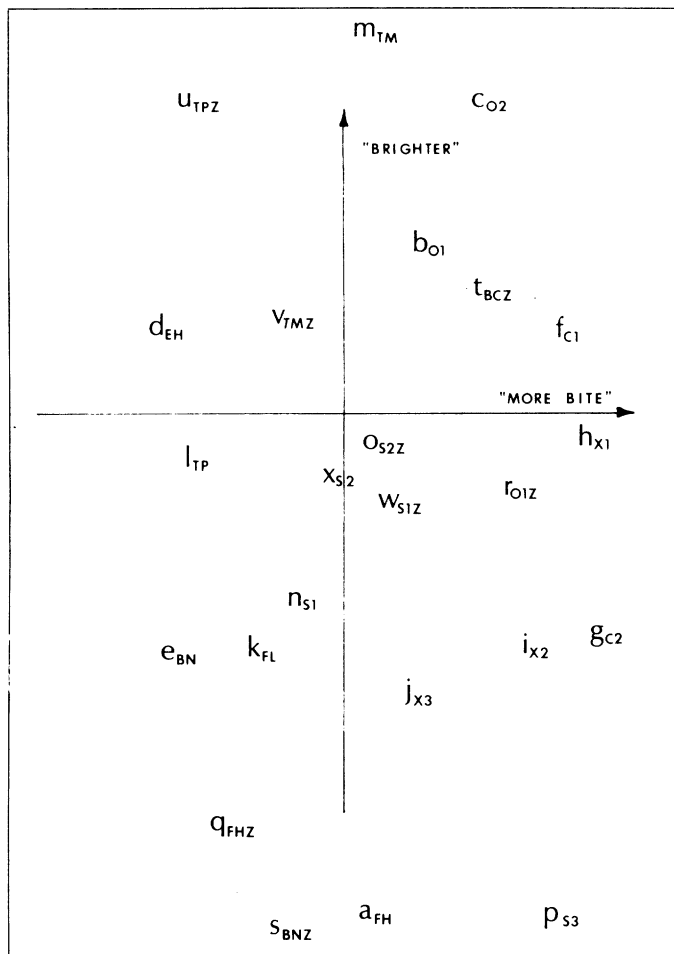


FIGURE 1. Two-dimensional timbre space representation of 24 instrument-like sounds obtained from Grey. The space was produced by the KYST multidimensional scaling program from dissimilarity judgements made by Wessel. The lower-case letters identify the sounds as recorded on a cassette-tape prepared to accompany this paper and distributed by IRCAM. The upper-case subscripts at each point identify the tones as reported in (Grey: 1975, 1977), (Grey and Gordon: 1978) and (Gordon and Grey: 1978). The original analyzed tones upon which these synthesized versions are based are being presented in the *Computer Music Journal* series "Lexicon of Analyzed Tones."

Abbreviations for stimulus points: O1, O2 = oboes, FH = French horn, BN = bassoon, C1 = E-flat clarinet, C2 = bass clarinet, FL = flute, X1, X2, X3 = saxophones, TP = trumpet, EH = English horn, S1 = cello played *sul ponticello*, S2 = cello played normally, S3 = cello played muted *sul tasto*, FHZ = modified FH with spectral envelope, BNZ = modified BN with FH spectral envelope, S1Z = modified S1 with S2 spectral envelope, S2Z = modified S2 with S1 spectral envelope, TMZ = modified TM with TP spectral envelope, BCZ = modified C2 with O1 spectral envelope, O1Z = modified O1 with C2 spectral envelope.

The horizontal dimension is related to the quality of the "bite" in the attack. Possibilities for quantification of this dimension are discussed in the next section on timbre patterns formed in the space.

Predictions About Timbre Patterns

To a large extent, music consists of syntactic patterns. It is the nature of the relationships among the elements of the patterns that is of primary importance in their perception. In the next series of examples I would like to show that when note-to-note timbral changes are organized in terms of the timbre space which we just examined, then predictable perceptual organizations of timbre patterns can be obtained.

First we will examine some auditory effects that we can relate to the dimensions of the space and the distances spanned. In the following patterns the sequence of notes will alternate between two differing timbres, but otherwise the pitch sequence and rhythmic timing will remain fixed. The pitch sequence is the simple, repeating, three-note ascending line shown in Figure 2. The alternating timbre sequence is shown by the alternating notes marked respectively with "O" and "X". When the timbral distance between the adjacent notes is small, the repeating ascending pitch lines dominate our perception. However, when the timbre difference is enlarged along the "spectral energy distribution" axis, the perceptual organization of the pattern is radically altered. The line now splits at the wide timbral intervals and for many listeners two interwoven descending lines are formed, each with its own timbral identity. This type of effect is called "melodic fission" or "auditory stream segregation" in the psychoacoustic literature (Bregman and Campbell: 1971; Van Noorden: 1975) and is a consequence of the large spectral energy distribution between the alternating timbres.

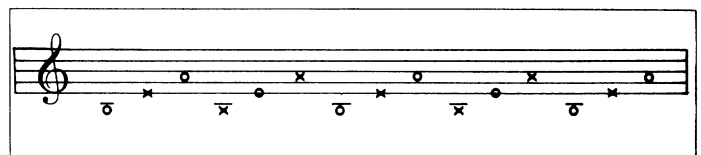


FIGURE 2. Ascending pitch patterns in "three" with two alternating timbres ("O" and "X"). If the timbral difference between adjacent notes is large, then one tends to perceive interleaved descending lines formed by the notes of the same timbral type.

A different effect is obtained by moving along the dimension we interpreted as relating to the onset characteristics of the sounds. When this is done we obtain a perceptually irregular rhythm even though the acoustical onset times of the notes are the same. This observation has some important implications for the control of sound in synthesis. When we alter the properties of the attack of the tone, we are also likely to influence the temporal location of the perceived onset of the tone. This lack of accord between physical onset time and subjective onset time has been observed with speech sounds by Morton, *et al.* (1975). Morton's experimental procedure offers the possibility of determining the relative perceived onset times for a set of notes. The procedure uses a simple *ABAB...* *AB* alternating sequence similar to those just described. The listener adjusts the shift in onset for all the *B*'s in the sequence until the sequence is perceived as regular, and the temporal displacement in the physical onset

is then noted. Perhaps with the application of such a method to musical timbres and with the employment of a good model of auditory temporal integration of complex time-varying spectra, we will be able to predict more precisely where the perceived onsets of tones with differing spectral evolutions will occur. *Nevertheless, both the fine tuning of rhythm in music and psychoacoustic research will benefit greatly if the control software of our synthesis systems allows easy and flexible adjustment of physical onset times in complex musical contexts.*

With the previous examples we have further verified the interpretation of the timbre space and have demonstrated that to some extent the properties of the space retain their validity in richer musical situations. In the next example involving timbral analogies I hope to demonstrate that other properties of the geometry of the timbre space allow us to make predictions about pattern perception as well.

Timbral Analogies

Composers frequently make transpositions of pitch patterns. It seemed natural to ask if *transpositions of timbral sequences* work as well. To get some preliminary indications regarding this possibility David Ehresman and I (Ehresman and Wessel : 1978) tested a parallelogram model of analogies developed by Rumelhart and Abramson (1973). The basic idea is illustrated in Figure 3. If we make a two-note timbral pattern, the sequence *A, B* in Figure 3, and wish to make an analogous (or, for our purposes, transposed) sequence beginning on timbre *C*, then we choose the note *D* that best completes a parallelogram in the space.

To test this idea we presented listeners with four different solutions to timbral analogies of the form $A \rightarrow B$ as $C \rightarrow D_i$. They were asked to order the alternative analogies indicating the best formed, the next best formed, and so forth. The idea was to check if the listeners' ranking of the goodness of the analogy would be inversely related to the distance between the various alternative solutions, D_i 's, and the ideal solution point specified by the parallelogram in the timbre space. In constructing the analogy problems we chose the alternatives so they would fall at graded distances from the ideal solution point. We selected 40 different analogy problems from a timbre space similar to the one just described. Our laboratory computer system allowed us to synthesize and store the tones and then to automate and analyze the experiment. One important feature of the system was that the listeners had essentially random access to the four alternative formations of an analogy problem and were able to make rapid auditory comparisons among them.

Table 1 shows the results of this experiment for nine listeners who each ordered the solutions for 40 different analogy problems. The entries in the table indicate the proportion of times, averaged over listeners and analogies, for which the *I*th closest alternative to the ideal analogy point was ranked as the *J*th best solution, where *I* is the row index and *J* is the column index. Column 1 of this table shows that the prediction was indeed fulfilled. In fact, the distance between an alternative and ideal analogy point predicts not only the best solution but the rank ordering of most of the alternatives. Though more research needs to be done, the notion of transposing a sequence of timbres by forming another sequence geometrically parallel to it in timbre space thus appears to be a reasonable and musically viable idea.

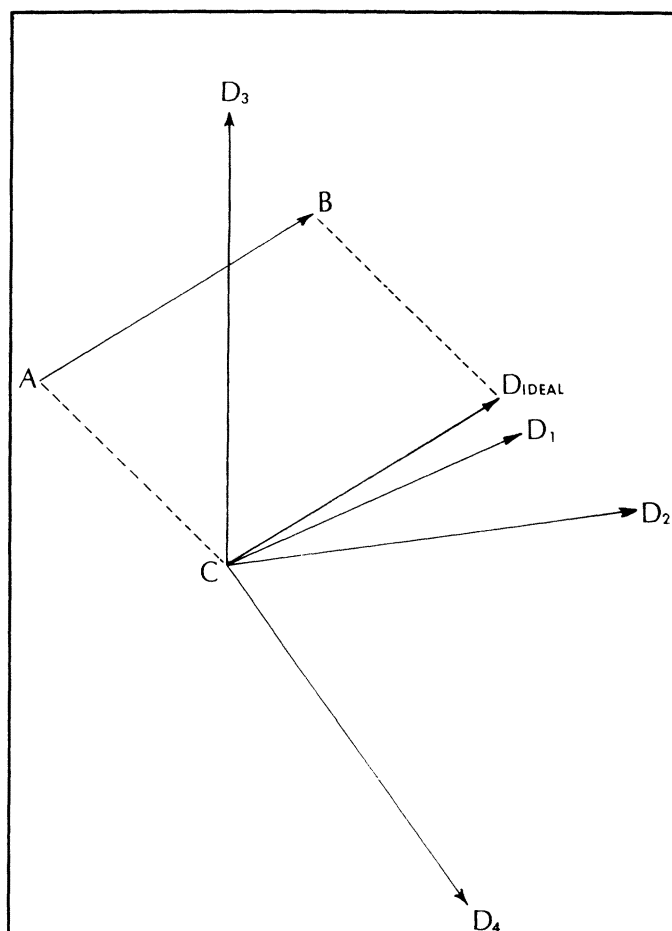


FIGURE 3. Parallelogram model of timbre analogies. $A \rightarrow B$ is a given change in timbre; $C \rightarrow D$ is a desired timbral analogy, with *C* given. *D* is the ideal solution point. D_1, D_2, D_3 , and D_4 are the actual solutions offered to the listeners.

Table 1. Rank order data averaged over nine listeners and all 40 analogies. Cf. Figure 3.

Rank Distance of the Alternative from the Ideal Solution (<i>I</i>)	Listener-Assigned Rank (<i>J</i>)			
	1	2	3	4
1	.422	.303	.156	.119
2	.322	.283	.217	.178
3	.169	.267	.358	.206
4	.086	.147	.269	.497

The idea of transposing timbral patterns suggests another procedure for representing important perceptual relationships. Using a technique called simultaneous linear equation scaling (Carroll and Chang: 1972) one can derive distance estimates directly from the analogy quality judgments and then use the standard multidimensional scaling algorithms on these distances. Though we have yet to try this scheme, it seems promising.

Designing Control Systems from the Perceptual Representations

The timbre space representation suggests relatively straightforward schemes for controlling timbre. The basic idea is that by specifying coordinates in a particular timbre space, one could hear the timbre represented by those coordinates. If these coordinates should fall between existing tones in the space, we would want this interpolated timbre to relate to the other sounds in a manner consistent with the structure of the space. Evidence that such interpolated sounds are consistent with the geometry of the space has been provided by Grey (1975). Grey used selected pairs of sounds from his timbre space and formed sequences of interpolated sounds by modifying the envelope break points of the two sounds with a simple linear interpolation scheme. These interpolated sequences of sounds were perceptually smooth and did not exhibit abrupt changes in timbre. Members of the original set of sounds and the newly created interpolated timbres were then used in a dissimilarity judgment experiment to determine a new timbre space. This new space had essentially the same structure as the original space with the interpolated tones appropriately located between the sounds used to construct them. It would appear from these results that the regions between the existing sounds in the space can be filled out, and that smooth, finely graded timbral transitions can be formed.

The most natural way to move about in the timbral space would be to attach the handles of control directly to the dimensions of the space. I examined such a control scheme in a real-time context (Wessel: 1976). A two-dimensional timbre space was represented on the graphics terminal of the computer that controlled the DiGiugno oscillator bank at IRCAM. One dimension of this space was used to manipulate the shape of the spectral energy distribution. This was accomplished by appropriately scaling the line segment amplitude envelopes according to a shaping function. The other axis of the space was used to control either the attack rate or the extent of synchronicity among the various components. Overall, the timbral trajectories in these spaces were smooth and otherwise perceptually well-behaved. To facilitate more complex forms of control, an efficient computer language for dealing with envelopes is needed. Grey and his colleagues at Stanford have developed a language for this purpose (Kahrs: 1977), and our group at IRCAM is making a similar effort. The basic idea behind such a language is to provide a flexible control structure that permits specification, sequencing, and combination of various procedures that create and modify envelopes. These procedures would include operations like stretching or shortening duration, changing pitch, reshaping spectrum, synchronizing or desynchronizing spectral components, and so forth. With such a language it will be possible to tie the operations on the envelope collections directly to the properties of the perceptual representations of the material.

Acknowledgments

I would like to thank John Grey and John Gordon for their tones and comments, and Gerald Bennett, Andy Moorer, Wayne Slawson, and John Strawn for their comments and critical reading of the manuscript.

D. Wessel: Timbre Space as a Musical Control Structure

References

- Alles, H. G. (1977a) "A Portable Digital Sound Synthesis System." *Computer Music Journal*, Vol. 1, No. 4, pp. 5-6.
- Alles, H. G. (1977b) "A Modular Approach to Building Large Digital Synthesis Systems." *Computer Music Journal*, Vol. 1, No. 4, pp. 10-13.
- Alles, H. G., and DiGiugno, G. (1977c) "A One Card, 64-Channel Digital Synthesizer." *Computer Music Journal*, Vol. 1, No. 4, pp. 7-9.
- Alonso, S., Appleton, J. H., Jones, C. (1975) "A Special-Purpose Digital System for the Instruction, Composition, and Performance of Music." *Proc. of the 1975 Conference on Computers in Undergraduate Curricula*, 6, 17-22.
- Arfib, D. (1977) "Digital Synthesis of Complex Spectra by Means of Non-Linear Distortion of Sine Waves and Amplitude Modulation." Paper presented at the 1977 International Computer Music Conference, Center for Music Experiment, University of California at San Diego, La Jolla.
- Beauchamp, J. W. (1969) "A Computer System for Time-Variant Harmonic Analysis and Synthesis of Musical Tones," in *Music by Computers*, edited by H. von Foerster and J. W. Beauchamp. John Wiley and Sons, Inc., New York.
- Beauchamp, J. (1975) "Analysis and Synthesis of Cornet Tones Using Nonlinear Interharmonic Relationships." *Journal of the Audio Engineering Society*, Vol. 23, pp. 778-795.
- Bregman, A. S., and Campbell, J. (1971) "Primary Auditory Stream Segregation and Perception of Order in Rapid Sequences of Tones." *Journal of Experimental Psychology*, Vol. 89, pp. 244-249.
- Carroll, J. D., and Chang, J. J. (1972) "Simules: Simultaneous Linear Equation Scaling." *Proceedings, 80th Annual Convention of American Psychological Association*.
- Chowning, J. M. (1973) "The Synthesis of Complex Audio Spectra by Means of Frequency Modulation." *Journal of the Audio Engineering Society*, Vol. 21, pp. 526-534. Reprinted in *Computer Music Journal*, Vol. 1, No. 2, pp. 46-54, 1977.
- DiGiugno, G. (1976) "A 256 Digital Oscillator Bank." Paper presented at the 1976 International Computer Music Conference, Massachusetts Institute of Technology, Cambridge.
- Ehresman, D., and Wessel, D. L. (1978) "Perception of Timbral Analogies." IRCAM Technical Report No. 13.
- Erickson, R. (1975) *Sound Structure in Music*. University of California Press, Berkeley and Los Angeles.
- Grey, J. M. (1975) *Exploration of Musical Timbre*. Stanford Univ. Dept. of Music Tech. Rep. STAN-M-2.
- Grey, J. M., and Moorer, J. A. (1977) "Perceptual Evaluations of Synthesized Musical Instrument Tones." *Journal of the Acoustical Society of America*, Vol. 62, pp. 454-62, 1977.
- Grey, J. M. (1977) "Multidimensional Perceptual Scaling of Musical Timbre." *Journal of the Acoustical Society of America*, Vol. 61, pp. 1270-1277, 1977.
- Gordon, J., and Grey, J. M. (1978) "Perceptual Effects of Spectral Modifications on Orchestral Instrument Tones." *Computer Music Journal*, Vol. 2, No. 1, pp. 24-31.

- Kahrs, M. (1977) "A Computer Language for Psychoacoustic Study and Musical Control of Timbre." Paper presented at the 1977 International Computer Music Conference. Center for Music Experiment, University of California at San Diego, La Jolla.
- Kruskal, J.B. (1964a) "Multidimensional Scaling by Optimizing Goodness of Fit to a Nonmetric Hypothesis." *Psychometrika*, Vol. 29, pp. 1-27.
- Kruskal, J.B. (1964b) "Nonmetric Multidimensional Scaling: A Numerical Method." *Psychometrika*, Vol. 29, pp. 115-129.
- Le Brun, M. (1979) "Waveshaping Synthesis," *Journal of the Audio Engineering Society*, Vol. 27, No. 4, pp. 250-266.
- Levelt, W.J. M., Riemersma, J.B., and Bunt, A.A. (1972) "Binaural Additivity of Loudness," *British Journal of Mathematical and Statistical Psychology*, Vol. 25, pp. 51-68.
- Luce, R. D. (1972) "What Sort of Measurement is Psychophysical Measurement?" *American Psychologist*, February, pp. 96-106.
- Moorer, J.A. (1975) "On the Loudness of Complex, Time-Variant Tones." Stanford University, Department of Music, Tech Report STAN-M-4.
- Moorer, J. A. (1976) "The Synthesis of Complex Audio Spectra by Means of Discrete Summation Formulae." *Journal of the Audio Engineering Society*, Vol. 24, pp. 717-727.
- Moorer, J. A. (1976) "The Use of the Phase Vocoder in Computer Music Applications." Presented at the 55th Convention of the Audio Engineering Society; available as Preprint No. 1146 (E1).
- Moorer, J. A. (1977) "Signal Processing Aspects of Computer Music - A Survey." *Proceeding of the IEEE*, Vol. 65, No. 8, pp. 1108-1137, and *Computer Music Journal*, Vol. 1, No. 1, pp. 4-38.
- Moorer, J.A., Grey, J.M., and Snell, J.M., (1977) "Lexicon of Analyzed Tones (Part I : Violin Tone)," *Computer Music Journal*, Vol. 1, No. 2, pp. 39-45.
- Moorer, J.A., Grey, J.M., and Strawn, J. (1977) "Lexicon of Analyzed Tones (Part II :Clarinet and Oboe Tones)," *Computer Music Journal*, Vo. 1, No. 3, pp. 12-29.
- Moorer, J.A., Grey, J.M., and Strawn, J. (1978) "Lexicon of Analyzed Tones (Part III: the Trumpet)," *Computer Music Journal*, Vol. 2, No. 2, pp. 23-31.
- Morton, J., Marcus, S., and Frankish, C. (1976) "Perceptual Centers (P-Centers)" *Psychological Review*, Vol. 83, No. 5, pp. 405-408.
- Portnoff, M.R. (1976) "Implementation of the Digital Phase Vocoder Using the Fast Fourier Transform." *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. ASSP-24, pp. 243-248.
- Risset, J. C. and Mathews, M.V. (1969) "Analysis of Musical Instrument Tones." *Physics Today*, Vol. 22, pp. 23-30.
- Rumelhart, D.E., and Abrahamson, A.A. (1973) "Toward a Theory of Analogical Reasoning." *Cognitive Psychology*, Vol. 5, pp. 1-28.
- Samson, Peter (1977) "Systems Concepts Digital Synthesizer Specifications." Available from Systems Concepts, 520 Third Street, San Francisco, California 94107.
- Shepard, R. N. (1966) "Metric Structures in Ordinal Data." *Journal of Mathematical Psychology*, Vol. 3, pp. 287-315.
- Shepard, R. N. (1972) "Psychological Representation of Speech Sounds," in *Human Communication*, edited by E.E. Davis and P.B. Denes. Mc Graw-Hill, New York.
- Shepard, R.N. (1974) "Representations of Structure in Similarity Data: Problems and Prospects." *Psychometrika* 39, 373-421.
- van Noorden, L. (1975) "Temporal Coherence in the Perception of Tone Sequences." Instituut voor Perceptie Onderzoek, Eindhoven, Holland.
- Wedin, L. and Goude, G. (1972) "Dimension Analysis of the Perception of Instrumental Timbre." *Scandinavian Journ. Psych.*, Vol. 13, pp. 228-240.
- Wessel, D.L. (1973) "Psychoacoustics and Music : A Report from Michigan State University." *PAGE: Bulletin of the Computers Arts Society*, Vol. 30.
- Wessel, D.L. and Grey, J.M. (1978) "Conceptual Structures for the Representation of Musical Material." IRCAM Technical Report #14.
- Wessel, D.L. and Smith, B. (1977) "Psychoacoustic Aids for the Musician's Exploration of New Material." Paper presented at the 1977 International Computer Music Conference, Center for Music Experiment, University of California at San Diego, La Jolla.
- Wessel, D.L. (1976) "Perceptually Based Controls for Additive Synthesis." Paper presented at the 1976 International Computer Music Conference, Massachusetts Institute of Technology.
- Zwicker, E. and Sharf, B. (1965) "A Model of Loudness Summation." *Psych. Review*, Vol. 72(1), pp. 3-26.