5

Exploration of Timbre by Analysis and Synthesis

Jean-Claude Risset

Directeur de Recherche au CNRS
Laboratoire de Mécanique et d'Acoustique
Marseille, France

David L. Wessel

Center for New Music and Audio Technologies
Department of Music
University of California, Berkeley
Berkeley, California

I. TIMBRE

Timbre refers to the quality of sound. It is the perceptual attribute that enables us to distinguish among orchestral instruments that are playing the same pitch and are equally loud. But, unlike loudness and pitch, timbre is not a well-defined perceptual attribute. Definitions tend to indicate what timbre is not rather than what it is. Take as an example the following enigmatic definition provided by the American National Standards Institute (1960, p. 45): "Timbre is that attribute of auditory sensation in terms of which a listener can judge that two sounds similarly presented and having the same loudness and pitch are dissimilar."

The notion of timbral constancy or invariance is even vaguer than that suggested in the definitions of timbre as a basis for discrimination. It would seem that a form of timbral constancy is implied by the common observation that a sound source can be reliably identified over a wide variety of circumstances. For example, a saxophone is readily identified as such regardless of the pitch or dynamic it is playing. Furthermore, the saxophone remains a saxophone whether it is heard over a distortion-ridden pocket-sized transistor radio or directly in a concert hall. Thus, the question arises as to the physical correlates of this constancy. Is there a physical invariant or a characteristic feature mediating a given timbre?

The issue is not only academic: it has musical relevance, because electronic and computer technology promises access to an unlimited world of timbres. One

must, however, know how to evoke a given timbre; that is, how to describe it in terms of the physical structure of sound.

II. TIMBRE AND THE FOURIER SPECTRUM: THE CLASSICAL VIEW

Physicists have been analyzing musical instrument tones for some time. The goal of many of these acoustical analyses is to determine the physical correlates of tone quality.

Many results of such analyses have been published (Culver, 1956; Meyer & Buchmann, 1931; Miller, 1926; Olson, 1967; Richardson, 1954). The general conclusion of such studies was that musical sounds are periodic and that the tone quality is associated solely with the waveshape, more precisely with the Fourier spectrum of the waveshape. These early analyses were strongly motivated by the theorem of Fourier, which states that a periodic waveshape is completely defined by the amplitudes and phases of a harmonic series of frequency components (see Feynman, Leighton, & Sands, 1963, chapters 21–25; Jenkins & Watts, 1968). But the claim, often known as Ohm's acoustical law, is that the ear is phase deaf. Put more precisely, Ohm's acoustical law states that if the Fourier representations of two sounds have the same pattern of harmonic amplitudes but have different patterns of phase relationships, a listener will be unable to perceive a difference between the two sounds, even though the sounds may have very different waveforms (see Figure 1).

It has been argued that the ear is not actually phase deaf. It is indeed true that under certain conditions, changing the phase relationship between the harmonics of a periodic tone can alter the timbre (Mathes & Miller, 1947; Plomp & Steeneken, 1969); however, this effect is quite weak, and it is generally inaudible in a normally reverberant room where phase relations are smeared (Cabot, Mino, Dorans, Tackel, & Breed, 1976; Schroeder, 1975). One must remember, though, that this remarkable insensitivity to phase, illustrated by Figure 1, holds only for the phase relationship between the harmonics of periodic tones.[1]

Thus, it would appear that timbre depends solely on the Fourier spectrum of the sound wave. The most authoritative proponent of this conception has been Helmholtz (Helmholtz, 1877/1954). Helmholtz was aware that "certain characteristic particularities of the tones of several instruments depend on the mode in which they begin and end," yet he studied only "the peculiarities of the musical tones which continue uniformly," considering that they determined the "musical quality of the tone." The temporal characteristics of the instruments were averaged out by

[1]A varying phase can be interpreted as a varying frequency. Also, dispersive media (for which the speed of propagation is frequency dependent) cause inaudible phase distortion for periodic tones and objectionable delay distortion for nonperiodic signals (e.g., the high frequencies can be shifted by several sounds with respect to the low ones in a long telephone cable: this makes speech quite incomprehensible).

FIGURE 1 The waves 1 to 4 correspond to tones generated with the same spectrum but with different phase relations between the components: these tones with quite different waveforms sound very similar (Plomp, 1976).

the early analyses (Hall, 1937); but because different instruments had different average spectra, it was believed that this difference in average spectrum was utterly responsible for timbre differences. This view is still widely accepted: a reputed and recent treatise like the *Feynmann Lectures on Physics* gives no hint that there may be factors of tone quality other than "the relative amount of the various harmonics."

Actually, even a sine wave changes quality from the low to the high end of the musical range (Köhler, 1915, Stumpf, 1926). In order to keep the timbre of a periodic tone approximately invariant when the frequency is changed, should the spectrum be transposed so as to keep the same amplitude relationship between the harmonics or should the absolute position of the spectral envelope be kept invariant? This question produced a debate between Helmholtz and Herman (cf. Winckel, 1967, p. 13). In speech, a vowel corresponds approximately to a spectrum with a given formant structure. A formant is a peak in the spectral envelope that occurs at a certain frequency and is often associated with a resonance in the sound source. This is the case for voice sounds, and the formants can be related to resonances in the vocal tract.

Indeed, in many cases, a fixed formant structure (Figure 2) gives a timbre that varies less with frequency than a fixed spectrum—a better invariance for "sound color," as Slawson (1985) calls timbre for nonchanging sounds (Plomp, 1976, pp. 107–110; Plomp & Steeneken, 1971; Slawson, 1968).

Certain characteristics of the spectrum induce certain timbral qualities. This can easily be demonstrated by modifying the spectrum with filters. Brightness (or sharpness) relates to the position of the spectral envelope along the frequency axis (see Section XVI). Presence appears to relate to strong components around 2000 Hz.

The concept of critical bandwidth,[2] linked to the spectral resolution of the ear (Plomp, 1966), may permit a better understanding of the correlation between spectrum and timbre. In particular, if many high-order harmonics lie close together, that is, within the same critical bandwidth, the sound becomes very harsh.

[2]The critical bandwidth around a certain frequency roughly measures the range within which this frequency interacts with others. The width of a critical band is about one third of an octave above 500 Hz and approximately 100 Hz below 500 Hz (cf. Scharf, 1970). This important parameter of hearing relates to spectral resolution (Plomp, 1964, 1976).

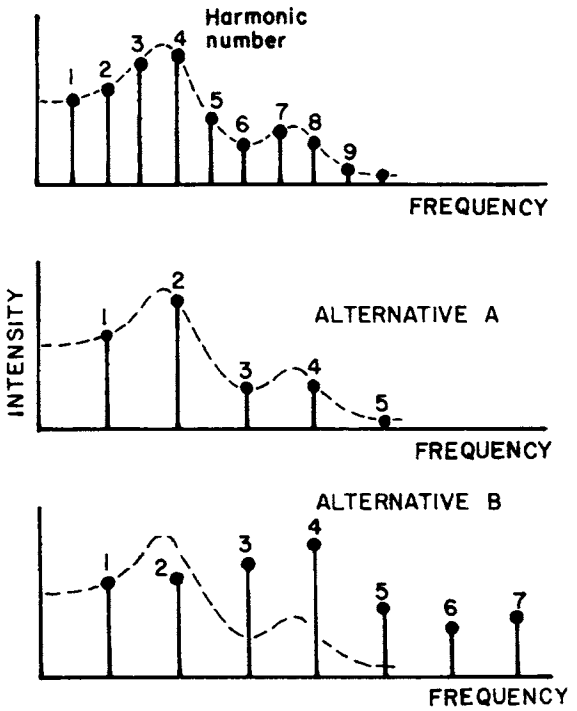

FIGURE 2 This figure refers to an experiment by Slawson (1968) comparing alternative predictions of invariance in timbre under octave increases in fundamental frequency. The experiment rules out alternative B, that of the relative pitch or overtone theory, in favor of alternative A, that of the fixed-frequency or formant theory.

Hence, for instance, antiresonances in the frequency response of string instruments play an important part in diminishing the roughness of the tones. It may be more significant to display spectra modified so as to take critical bands into account. This was done in some studies: the frequency axis is converted into so-called Bark units: 1 Bark corresponds to the width of one critical band over the whole frequency range (cf. Grey & Gordon, 1978; Moore & Glasberg, 1981; Scharf, 1970; Zwicker, 1961).

III. THE SHORTCOMINGS OF THE CLASSICAL CONCEPTION

So, for periodic tones, timbre depends upon spectrum. It has long been thought that musical tones are periodic, at least for most of their duration. Musical tones are often thought of as comprising three sections: attack, steady state, and decay. Helmholtz and his followers considered timbre to be determined by the spectrum of the steady state. However, this conception suffers from serious difficulties. As

we noted at the beginning of this chapter, musical instruments can be recognized even from a very poor recording, despite the fact that their spectra are radically changed by such distortion (Eagleson & Eagleson, 1947).

In fact, a normally reverberant room has an incredibly jagged frequency response, with fluctuations up to 20 dB, and this frequency response is different at every point in the room (Wente, 1935). Hence, spectra are completely changed in ways that depend on the specific location. However, when one moves in the room, the corresponding timbres are not completely upset as one would expect them to be if they depended only on the precise structure of the frequency spectrum.

Also, various methods of sound manipulation show that temporal changes bear strongly on tone quality. Removing the initial segment of notes played by various instruments impairs the recognition of these instruments, as noted by Stumpf as early as 1910 (Stumpf, 1926). Subsequently, tape-recorder manipulation (George, 1954; Schaeffer, 1966) has made it easy to demonstrate the influence of time factors on tone quality. For instance, playing a piano tone backwards gives a non-piano-like quality, although the original and the reversed sound have the same spectra. However, temporal factors were not taken into account in most early analyses (cf. Hall, 1937): the analysis process could not follow fast temporal evolutions.

Recently, computer sound synthesis (Mathews, 1963, 1969) has made it possible to synthesize virtually any sound from a physical description of that sound. Efforts have been made to use the results of analyses of musical instrument tones that are to be found in treatises on musical acoustics as input data for computer sound synthesis. In most cases, the sounds thus obtained bear little resemblance to the actual tones produced by the instrument chosen; the tones thus produced are dull, lacking identity and liveliness (Risset & Mathews, 1969). Hence, the available descriptions of musical instrument tones must be considered inadequate, because they fail to pass the foolproof synthesis test. This failure points to the need for more detailed, relevant analyses and for a more valid conception of the physical correlates of timbre. Clearly, one must perform some kind of "running" analysis that follows the temporal evolution of the tones.

IV. ATTACK TRANSIENTS

A few attempts have been made since 1930 to analyze the attack transients of instrument tones (Backhaus, 1932; Richardson, 1954). These transients constitute an important part of the tones—in fact, many tones like those from the piano or percussion instruments have no steady state—yet their analysis has not produced much progress. The transients are intrinsically complex, and they are not reproducible from one tone to another, even for tones that sound very similar (Schaeffer, 1966). Most analyses have been restricted to a limited set of tones, and the researchers have tended to make generalizations that may be inappropriate even for different samples collected from the same instruments. These shortcomings

have produced many discrepancies in the literature and cast doubt on the entire body of acoustic data.

V. COMPLEXITY OF SOUNDS: IMPORTANCE OF CHARACTERISTIC FEATURES

Sounds are often intrinsically complex. Musical instruments have a complex physical behavior (Benade, 1976); often the damping is low, and transients are long compared with note duration. Also, the tones are not generated by a standardized mechanical player, but by human musicians who introduce intricacies both intentionally and unintentionally. Even if a human player wanted to, a human being could not repeat a note as rigorously as a machine does. If the musician has good control of the instrument, he or she should be able to play two tones sounding nearly identical, but these tones can differ substantially in their physical structure. More often the performer will not want to play all notes the same way, and the performer's interpretation of some markings depends on the performer's sense of style and technique. All these considerations, which involve different disciplines—physics, physiology, psychology, esthetics—certainly make it difficult to isolate characteristic invariants in musical instrument sounds.

This points out the need to extract significant features from a complex physical structure. Also, one must be able to control through synthesis the aural relevance of the features extracted in the analysis—to perform *analysis by synthesis.* Only recently has this been possible, thanks to the precision and flexibility of the digital computer.

We shall now review pioneering work on the exploration of timbre by computer analysis and synthesis.

VI. INSTRUMENTAL AND VOCAL TIMBRES: ADDITIVE SYNTHESIS

The study of trumpet tones performed in the mid-1960s by one of the authors (Risset, 1966; Risset & Mathews, 1969) illustrates some of the points just made. We chose trumpet tones because we were experiencing difficulties in synthesizing brasslike sounds with the computer. The tones synthesized with fixed spectra derived from the analysis of trumpet tones did not evoke brass instruments.

To obtain more data, we recorded musical fragments played by a professional trumpet player in an anechoic chamber. Sound spectrograms suggested that, for a given intensity, the spectrum has a formant structure; that is, the spectrum varies with frequency so as to keep a roughly invariant spectral envelope. The spectrograms gave useful information, although it was not precise enough. Thus, selected tones were converted to digital form and analyzed by computer, using a pitch-synchronous analysis (PISA program, Mathews, Miller, & David, 1961). Pitch-

synchronous analysis assumes that the sound is quasi-periodic; it yields displays of the amplitude of each harmonic as a function of time (one point per fundamental pitch period). The curved functions resulting from the analysis program were approximated with linear segments (Figure 3). These functions were then supplied to the MUSIC IV sound-synthesis program, and the resulting synthetic tones were indistinguishable from the originals, even when compared by musically skilled listeners. Hence, the additive synthesis model, with harmonic components controlled by piecewise linear functions, captures the aurally important features of the sound.

Conceptually, the model is simple. The pitch-synchronous analysis yields a string of snapshot-like spectra, hence a kind of time-variant harmonic analysis that is further reduced by fitting the linear segments to the amplitude envelope of each component. Computationally, however, this model is not very economical. Figure 3 shows that the functions can be quite complex, and the parameters must be estimated for every tone. So further simplifications of the model were sought. By systematic variation of the various parameters— one at a time—the relative importance of the parameters was evaluated. Whereas some parameters were dismissed as aurally irrelevant—for example, short-term amplitude fluctuations—a few physical features were found to be of utmost importance. These include the following: the attack time, with faster buildup of the low-order harmonics than the

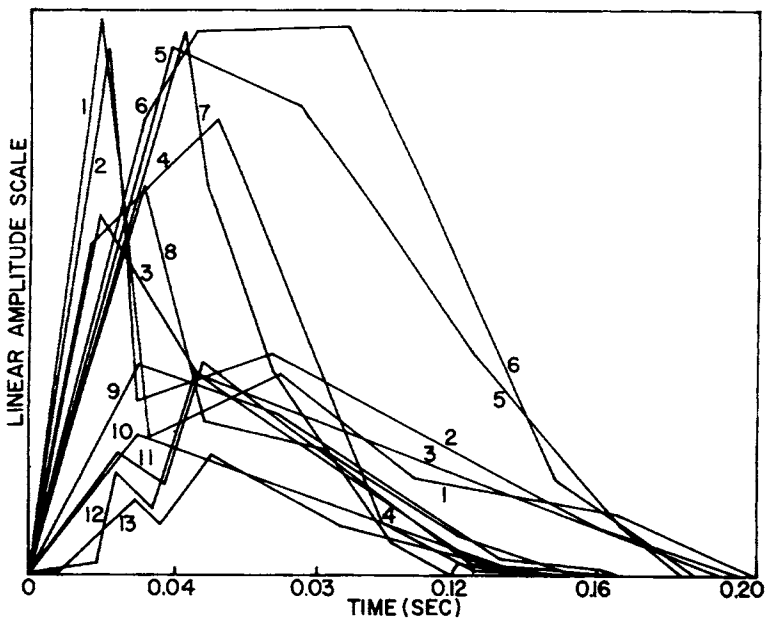

FIGURE 3 Line-segment functions that approximate the evolution in time of 13 harmonics of a D_4 trumpet tone lasting 0. 2 sec. Functions like these, obtained by analysis of real tones, have been used to control the harmonic amplitudes of synthetic tones (Risset & Mathews, 1969).

high-order ones; for certain tones, a quasi-random frequency fluctuation; and, most importantly, a peak in the frequency spectrum between 1000 and 1500 Hz and an increase in the proportion of high-order harmonics with intensity.

In fact, the latter property permitted us to abstract a simplified model of brass-like tones. Here only the amplitude function for the first harmonic was provided, and the amplitude functions for the other harmonics were deduced as fixed functions of this first harmonic amplitude, such that they increased at a faster rate (Risset, 1969). The specification was much more economical than the previous one and did not need to be precisely adjusted to yield the brasslike quality. Hence this property of an increase in spectral width with amplitude seems to be the most salient physical correlate of brasstone quality. This shows that, in addition to the way spectra vary over the pitch range, the variation in time of the spectrum can be critical to determine timbre. In the case of the brass, it can be described in terms of a nonlinear characteristic that enriches the spectrum when the amplitude increases. During the short attack of a brass tone, lasting less than 50 msec, the ear interprets the increase of spectral width as a brassy onset, even though it cannot describe what happens.

Beauchamp (1975) studied nonlinear interharmonic relationships in cornet tones and ascribed the brasslike character to the type of nonlinear relationship between the different harmonics, which are all functions of the first one regardless of the general level. This relationship has been found to have an acoustical basis (Backus & Hundley, 1971; Benade, 1976, pp. 439–447). This nonlinear property was used in the late sixties by Moog to produce brasslike sounds with his analog synthesizers: the cutoff frequency of a low-pass filter was made to go up with the amplitude, which was easy to achieve through voltage control. This characteristic has also been implemented in a very simple, satisfying way, using Chowning's powerful technique of spectral generation by frequency modulation described later (see Section X; Chowning, 1973; Morrill, 1977).

It was found in the trumpet-tone study that some factors may be important in some conditions and inaudible in others. For instance, details of the attack were more audible in long sustained tones than in brief tones. Also, it appeared that some listeners, when comparing real and synthetic tones, made their decision about whether a tone was real or synthetic on the basis of some particular property. For instance, they often assumed that the real tones should be rougher, more complex than the synthetic ones. This suggests that by emphasizing roughness in a synthetic tone, one could cause the listeners to believe it was a real tone. In his striking syntheses of brassy tones, Morrill (1977) has simulated intonation slips that greatly enhance the realistic human character of the tones. Similarly, in their study of string tones, Mathews, Miller, Pierce, and Tenney (1965, 1966) had included an initial random-frequency component, which simulates the erratic vibration that takes place when the string is first set in motion by the bow. When exaggerated, this gives a scratchy sound strikingly characteristic of a beginning string player. Such idiomatic details, imperfections, or accidents (Schaeffer, 1966) are characteristic of the sound source, and the hearing sense seems to be quite sensi-

tive to them. Taking this into account might help to give stronger identity and interest to synthetic sounds. Indeed, a frequency skew imposed on even a simple synthetic tone can help strongly endow it with subjective naturalness and identity. The pattern of pitch at the onset of each note is often a characteristic feature of a given instrument: the subtle differences between such patterns (e.g., a violin, a trombone, a singing voice) act for the ear as signatures of the source of sound.

The paradigm for the exploration of timbre by analysis and synthesis followed in the latter study has been much more thoroughly pursued by Grey and Moorer (1977) in their perceptual evaluation of synthesized musical instrument tones. Grey and Moorer selected 16 instrumental notes of short duration played near E♭ above middle C. This pitch was selected because it was within the range of many instruments (e.g., bass clarinet, oboe, flute, saxophone, cello, violin); thus, the tones represented a variety of timbres taken from the brass, string, and woodwind families of instruments. The tones were digitally analyzed with a heterodyne filter technique, providing a set of time-varying amplitude and frequency functions for each partial of the instrumental tone. Digital additive synthesis was used to produce a synthetic tone consisting of the superposition of partials, each controlled in amplitude and frequency by functions sampled in time. Each of the 16 instrumental notes could appear in at least four of the five following conditions: (a) original tone; (b) complex resynthesized tone, using the functions abstracted from the analysis; (c) tone resynthesized with a line-segment approximation to the functions (4 to 8 line segments); (d) cut-attack approximation for some of the sounds; and (e) constant-frequencies approximation. In order to evaluate the audibility of these types of data reduction, systematic listening tests were performed with musically sophisticated listeners. The tones were first equalized in duration, pitch, and loudness. An AA AB discrimination paradigm was used. On each trial four tones were played, three of them identical and the fourth one different; the listeners had to detect whether one note was different from the others, to tell in which pair it was located, and to estimate the subjective difference between this note and the others. The judgments were processed by multidimensional scaling techniques.

The results demonstrated the perceptual closeness of the original and directly resynthesized tones. The major cue helping the listeners to make a better than chance discrimination was the tape hiss accompanying the recording of the original tones and not the synthetic ones. The results also showed that the line-segment approximation to the time-varying amplitude and frequency functions for the partials constituted a successful simplification, leading to a considerable information reduction while retaining most of the characteristic subjectivity (see Figure 4). This suggests that the highly complex microstructure in the time-varying amplitude and frequency functions is not essential to the timbre and that drastic data reduction can be performed with little harm to the timbre. The constant-frequencies approximation (for tones without vibrato) was good for some tones but dramatically altered other ones. The importance of the onset pattern of the tones was confirmed by the cut-attack case.

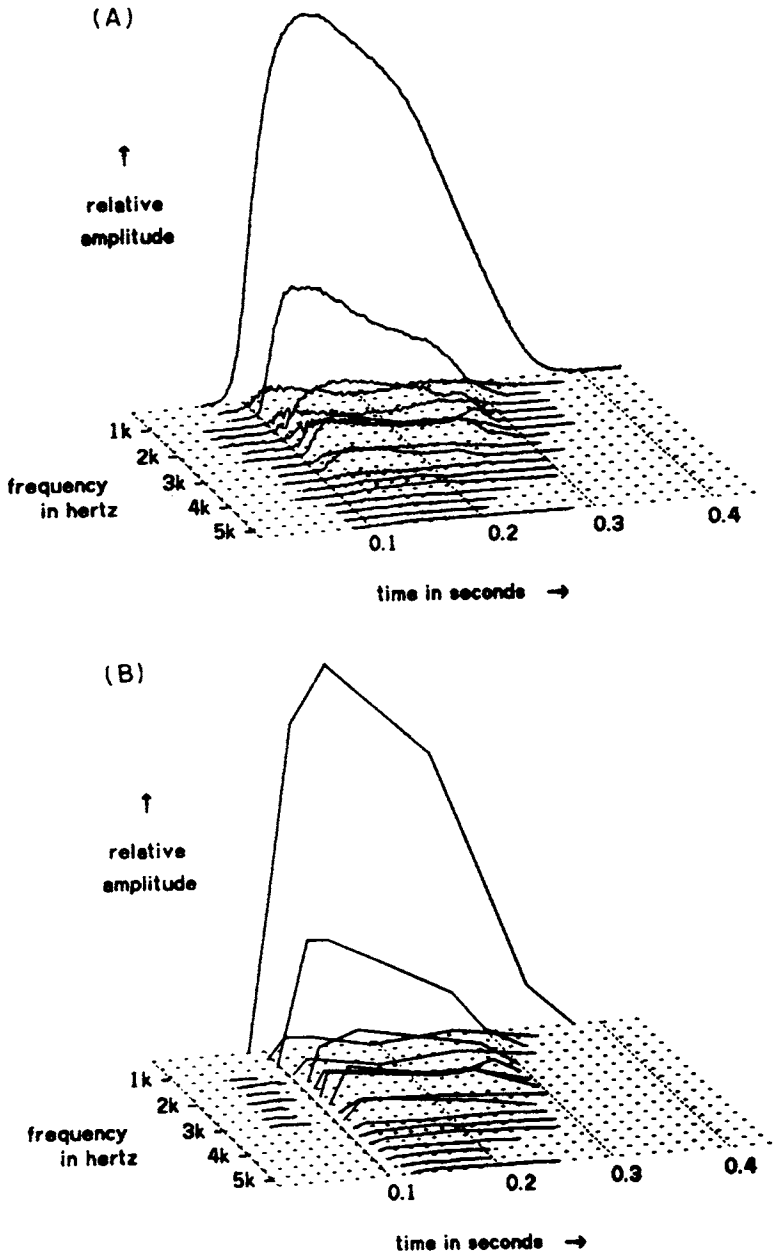

FIGURE 4 (A) Time-varying amplitude functions derived from heterodyne analysis from a bass clarinet tone, shown in a three-dimensional perspective plot. (B) Line-segment approximation to the function plotted in A. Both of these functions have been used to resynthesize the tone. Form B gives a considerable information reduction (Grey & Moorer, 1977).

A later study by Charbonneau (1979) has demonstrated that the simplification can go even further for most of the tones studied by Moorer and Grey (i.e., short tones of nonpercussive instruments). The various envelopes controlling each harmonic are replaced by a single averaged envelope; for each harmonic, this curve is weighted in order to preserve the maximum amplitude for this harmonic; it is also warped in time in order to preserve the times of appearance and extinction of the various harmonics. Although this is not a proper model for flute tones, it permits a good imitation for most of the other instruments. We shall mention other examples of simplified resynthesis in the following paragraphs.

Fletcher and his collaborators (Fletcher & Bassett, 1978; Fletcher, Blackham, & Christensen, 1963; Fletcher, Blackham, & Stratton, 1962; Fletcher & Sanders, 1967) studied the timbre of several instruments by analysis and synthesis, using an additive synthesis model. (The earlier of these studies did not use a computer but ad hoc analysis and synthesis devices). A study of the quality of piano tones (Fletcher et al., 1962) indicated that the attack time must be less than 0.01 sec, whereas the decay time can vary from 20 sec for the lowest notes to less than 1 sec for the very high ones. The variation of partial level versus time during the decay was highly complex and not always monotonic—the partials at times increase in intensity rather than decrease. However, the complexities of the decay pattern did not appear to be very relevant to the ear because the much simplified syntheses could sound similar to the original sounds—although it appeared in later studies that the dissimilarity between the behavior of different partials is often linked to liveliness. The piano study provided a major insight. It ascribed subjective warmth to the inharmonicity of the partials. The frequencies of the successive partials of a low piano tone are close to, but higher than, the frequencies of the harmonic series, to the extent that the 15th partial frequency can be 16 times that of the lowest one (Young, 1952). Now this slightly inharmonic pattern gives rise to a complex pattern of beats that induces a peculiar lively and warm quality. This is an important feature for low piano tones (and also for organ tones; cf. Fletcher et al., 1963).

Many analyses have been performed on piano sounds (Martin, 1947). They have been used to devise electronic pianos (Dijksterhuis & Verhey, 1969) whose tone quality (although not fully satisfying) depends on the simplified model abstracted from the analyses. Acoustic piano tones are extremely complex: low notes may comprise 100 or more significant spectral components, and the spectra fluctuate considerably. The quality of the acoustic piano may be hard to approach by synthesis. The issue is of practical significance, however, because digital pianos can be made much cheaper than acoustic ones, and, even more important, they are much easier to insulate acoustically. Current digital pianos obtained by sampling—that is, by recording of actual pianos tones—are not satisfactory: they fail to emulate the extremely responsive character of the acoustic piano, where the spectrum changes drastically with loudness, and also the interaction between strings occurring thanks to the soundboard.

In a study of violin tones, Fletcher and Sanders (1967) investigated the slow frequency modulation (around 6 Hz) known as vibrato, showing that it also modulates the spectrum of the tone. They also pointed to two features that enhance naturalness if they are simulated in the synthetic tones: the bowing noise at the onset of the tone and the sympathetic vibrations coming from the open strings (the latter occur substantially only when certain frequencies are played).

Clark, Luce, and Strong have also performed significant research on wind instrument tones by analysis and synthesis. In a first study (Strong & Clark, 1967a) wind instrument tones were synthesized as the sum of harmonics controlled by one spectral envelope (invariant with note frequency) and three temporal envelopes. (A more specific model was also sought for brass instruments, cf. Luce & Clark, 1967). Listeners were tested for their capacity to identify the source of the tones. Their identification was nearly as good as for real instrument tones, which indicates that this model grasps the elements responsible for the difference between the sounds of the different instruments. Incidentally, the probability of confusion between the tones of two instruments gives an indication of the subjective similarity between these tones; it has been used to ascertain the perceptual basis of the conventional instrument families (cf. Clark, Robertson, & Luce, 1964). The results suggest that some conventional families represent fairly well the subjective differentiations, especially the string and the brass family. A double reed family also emerged, comprising a tight subfamily (oboe and English horn) and a more remote member (the bassoon).

VII. ADDITIVE SYNTHESIS: PERCUSSION INSTRUMENTS

The aforementioned studies of timbre resorted to models of additive synthesis, whereby the sound was reconstituted as the superposition of a number of frequency components, each of which can be controlled separately in amplitude and frequency. Such models require much information specifying in detail the way each component varies in time: hence, they are not very economical in terms of the amount of specification or the quantity of computations they require. However, as was stated, the information on the temporal behavior of the components can often be simplified. In addition, the development of the digital technology has made it possible to build special processors with considerable processing power, for instance, digital synthesizers that can yield in real time dozens of separate voices with different envelopes (Alles & Di Giugno, 1977); so additive synthesis is a process of practical interest, considering its power and generality. It is not restricted to quasi-periodic tones; in fact, it can be used to simulate the piano and percussion instruments (Fletcher & Bassett, 1978; Risset, 1969).

In percussion instruments, the partials are no longer harmonics: their frequencies, found from the analysis, are those of the modes of vibration excited by the percussion and can sometimes be predicted from consideration of theoretical

acoustics. The synthesis can correspond to a considerably simplified model and still be realistic, provided it takes into account the aurally salient features. Fletcher and Bassett (1978) have simulated bass drum tones by summing the contribution of the most important components detected in the analysis—these were sine waves decaying exponentially, with a frequency shift downward throughout the tone. The simulation was as realistic as the recorded bass drum tones. The authors noted, however, that the loudspeakers could not render the bass drum tones in a completely satisfactory way.

Timbre can often be evoked by a synthesis that crudely takes into account some salient properties of the sound. Bell-like tones can be synthesized by adding together a few sine waves of properly chosen frequencies that decay exponentially at different rates—in general, the higher the frequency, the shorter the decay time. The frequency tuning of the first components is often critical, as it is in church bells, for instance: the frequencies of the first components approximate frequencies falling on a harmonic series, so that a distinct pitch (the strike tone) can be heard, even though there is no component at the corresponding frequency. The lowest component, which rings longer, is called the hum tone (cf. Rossing, 1990). Chinese bells have modes that tend to occur in pairs; these bells can emit two distinct notes depending on where they are struck (Rossing, Hampton, Richardson, Satoff, & Lehr, 1988). The Chinese gong has a very rich sound, which can strongly vary in the course of the sound, owing to a nonlinearity that is strengthened by the hammering of the metal during construction. This nonlinearity induces a chaotic behavior (Legge & Fletcher, 1989).

Realism is increased by introducing slow amplitude modulation for certain components of the spectrum. Such modulations exist for real bells; they can be ascribed to beats between closely spaced modes because the bell does not have perfectly cylindrical symmetry. Beats in bell-like and gong-like sounds can produce an effect of warmth. Snare drums can also be imitated with additive synthesis: the decays are much faster than for bells, and the effect of the snares can be evoked by adding a high-pitched noise band (Risset, 1969). Bell-like or drum-like sounds synthesized this way can also be transformed morphologically by changing the envelopes controlling the temporal evolution of the components. Thus, for instance, bells can be changed into fluid textures with the same harmonic (or rather inharmonic)[3] content yet with a quite different tone quality (Figure 5).

Additive synthesis can be performed with components less elementary than sine waves—for instance, groups of sinusoidal components (Kleczkowski, 1989). This can simplify synthesis, and in many cases it will not affect the aural result much. Percussive sounds often contain a great number of modes of vibration: it is easier to simulate the resulting signal as a noiselike component than by using many sinusoids (Risset, 1969). X. Serra and Smith (1990) have implemented a very effective technique to emulate percussive sounds: the idea is to separate a "deterministic" part—which can rendered through additive synthesis, as the sum

[3]cf. *Inharmonique*, on CD INA C1003.

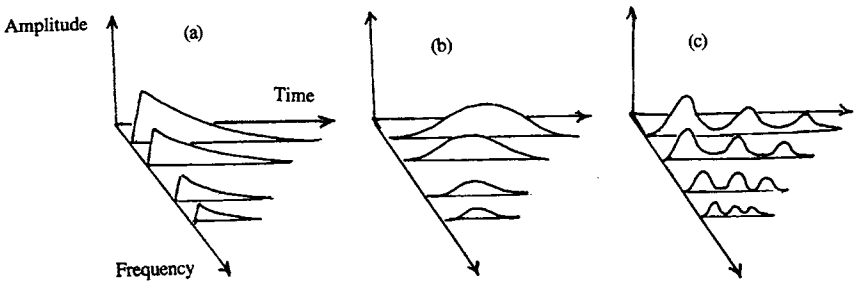

FIGURE 5 Perspective plots of synthetic *inharmonic* tones (i.e., with unequally spaced frequency components). Time runs from left to right, frequency increases from back to front, and amplitude increases from bottom to top. In (a), the sharp attack followed by a decay yields a bell-like tone. In (b), the time-varying amplitude function yields a fluid nonpercussive tones; because of differences in duration, the components are skewed in time, and they can be heard out much better than in the fused bell-like tone. In (c), the amplitude function has bounces: the components wax and wane, which causes a kind of timbral scintillation.

of a few sine-wave components controlled by amplitude and frequency envelopes—and a "stochastic" part, which is irregular but can be imitated by a proper time-varying noise component. In addition to this synthesis model, X. Serra and Smith developed an almost automatic analysis scheme, which extracts from a given sound the parameters to emulate this sound with the model. This approach has also been implemented by Depalle and Rodet in their fast Fourier transform (FFT) method (Cf. Section XIII). Rodet (1994) explores the possibility of controlling the proportion of chaos in sound synthesis: it can be low, for sounds perceived as harmonic, or high, for very noisy sounds. This applies to other instruments: for instance, flute sounds can be more or less breathy, and the shakuhachi flute sounds contain noise. It may also permit the synthesis of novel timbres.

VIII. CROSS-SYNTHESIS AND VOICE SYNTHESIS

In order to evaluate the relative significance of spectral and temporal envelopes, Strong and Clark (1967b) resorted to an interesting process: they exchanged the spectral and temporal envelopes among the wind instruments and asked listeners to attempt to identify these hybrid tones. The results indicated that the spectral envelope was dominant if it existed in a unique way for the instrument (as in the oboe, clarinet, bassoon, tuba, and trumpet); otherwise (as in the flute, trombone, and French horn), the temporal envelope was at least as important.

It should be noted that the above conclusions apply to wind instruments, which can have different temporal characteristics, although not very drastic ones. On the other hand, it is easy to verify by synthesis that a sharp attack followed by an

exponential decay gives a plucked or percussive quality to any waveform. In this case, temporal cues tend to dominate over spectral ones.

One often speaks of cross synthesis to characterize the production of a sound that compounds certain aspects of a sound A and other aspects of a sound B. Instruments such as the Australian aboriginal didgeridoo and the Jews' harp function through a kind of cross-synthesis between the instrument alone and the human vocal tract. There are interesting possibilities for cross-synthesis when sound production can be modeled as the combination of two relatively independent processes. In particular, a sound source can often be thought of as comprising an excitation that is transformed in ways that can be characterized in terms of a stable response (Huggins, 1952)—think of someone hitting a gong or blowing into a tube. The temporal properties of the sound are often largely attributable to the excitation insofar as the response depends on the structural properties of a relatively stable physical system; the spectral aspects result from a combination of those of the excitation and those of the response. (Huggins suggests that the hearing mechanism is well equipped to separate the structural and temporal factors of a sound wave). A good instance is that of voice production (cf. Fant, 1960): the quasi-periodic excitation by the vocal cords is fairly independent of the vocal tract response, which is varied through articulation. Thus, the speech waveform can be characterized by the *formant* frequencies (i.e., the frequencies of the vocal tract resonances) and by the *fundamental* frequency *(pitch)* of the excitation— except when the excitation is noiselike (in unvoiced sounds like *s* or *f*).

A considerable amount of research on speech synthesis has demonstrated the validity of this physical model. It is possible to synthesize speech that sounds very natural. It remains difficult, however, to mimic with enough accuracy and suppleness the transitions in spectrum and frequency that occur in speech. In fact, although one can faithfully imitate a given utterance by analysis and synthesis, it is still difficult to achieve a satisfactory "synthesis by rule," whereby the phonetic elements (phonemes or dyads) would be stored in terms of their physical description and concatenated as needed to form any sentence, with the proper adjustments in the physical parameters performed automatically according to a set of generative rules. We cannot dwell at length here on this important problem; we can notice that the correlates of a speaker's identity are multiple: the spectral quality of the voice as well as the rhythmic and intonation patterns are significant. At this time, one cannot reliably identify speakers from their voiceprints as one can from their fingerprints (cf. Bolt, Cooper, David, Denes, Pickett, & Stevens, 1969, 1978), but recent research shows that a careful investigation of the prosodic parameters (cf. Section XIII) can provide good cues for identification.

The notion of independence between the vocal tract and the vocal cords is supported by an experiment by Plomp and Steeneken (1971); however, it has to be qualified for the singing voice. For high notes, sopranos raise the first formant frequency to match that of the fundamental in order to increase the amplitude (Sundberg, 1977, 1987): hence certain vowels (for instance i, as in *deed*) are so

distorted that they cannot be recognized when sung on high pitches. Specific features detected in the singing voice have been confirmed by synthesis in the work of Sundberg, Chowning, Rodet, and Bennett (1981). Through certain processes of analysis (like inverse filtering or linear predictive coding; cf. Flanagan, 1972), one can decompose a speech signal to separate out the contributions of the vocal cords and the vocal tract. These processes made it possible for Joan Miller to synthesize a voice as though it were produced with the glottis of one person and the vocal tract of another one (cf. Mathews et al., 1961).[4] Actually, the source signal (because of the vocal cords) can be replaced by a different signal, provided this signal has enough frequency components to excite the vocal tract resonances (between, say, 500 and 3000 Hz). It is thus possible to give the impression of a talking (or singing?) cello or organ. Composers are often interested in less conspicuous effects, for instance in producing timbres from the combination of two specific tone qualities, using processes other than mere mixing or blending. This can be achieved through processes of analysis and synthesis, like the phase vocoder or the predictive coding process, or also through the reconstitution of the sounds through a certain model, like frequency modulation or additive synthesis. By physically interpolating the envelopes of the harmonics, Grey and Moorer (1977) have been able to gradually transform one instrumental tone into another one (e.g., a violin into an oboe) through monodic intermediary stages that do not sound like the superposition of a violin and a oboe.

IX. SUBTRACTIVE SYNTHESIS

Most of the work quoted in the preceding section on cross-synthesis and voice synthesis uses some form of subtractive synthesis. This method consists of submitting a spectrally rich wave to a specific type of filtering, thus arriving at the desired tone by eliminating unwanted elements rather than by assembling wanted ones. Subtractive synthesis is better adapted to certain types of sounds. As was mentioned, the process of speech articulation consists of shaping the vocal tract so that it filters in a specific way the spectrally rich source signal produced by the vocal cords. In fact, linear prediction coding consists of adjusting the parameters of a time-variant recursive filter so as to minimize the difference between the original speech signal and the signal obtained by filtering a single, quasi-periodic pulse wave by this recursive filter (see later).

Another instance in which subtractive synthesis has proven most useful is the case of violin tones, as demonstrated by Mathews' electronic violin. Mathews and Kohut (1973) have studied the aural effect of the resonances of the violin box

[4]Impressive examples of voice synthesis and processing for musical uses have been demonstrated in particular by Bennett, Chowning, Decoust, Dodge, Harvey, Lansky, Moorer, Olive, Petersen, Wishart, and others (Cf. records New Directions in Music, Tulsa studios; CRI SD 348; and CDs Wergo 2012-50, 2013-50, 2024-50, 2025-2, 2027-2, 2031-2, Elektra 9 60303-2, Neuma 450-73, New Albion Records NA 043, GMEM EI-06, Computer Music Journal Sound Anthology: 15-19, 1991-1995, and 20, 1996).

through electronic simulation. They have approximated the complex frequency response of a violin (which exhibits many peaks and minima—as many as 20 or more in the audible frequency range) with a set of electrical resonant filters (between 17 and 37). In this experiment, the vibration of the violin string near the bridge was converted into an electric signal by a magnetic pickup. This signal was approximately a triangular wave, as predicted by Helmholtz (Kohut & Mathews, 1971); hence, it consisted of a number of significant harmonic components whose amplitudes decay regularly with the rank. This signal was then subjected to the complex filtering approximating the response of the box. It was possible to change the characteristics of that filtering by changing both the damping of the resonances and their distribution along the frequency axis. It was found that a violin-like tone could be achieved with 20 or 30 resonances distributed in the frequency range of 200–5000 Hz, either randomly or at equal musical intervals (Figure 6). The best tone was obtained with intermediate values of damping, corresponding to a peak-

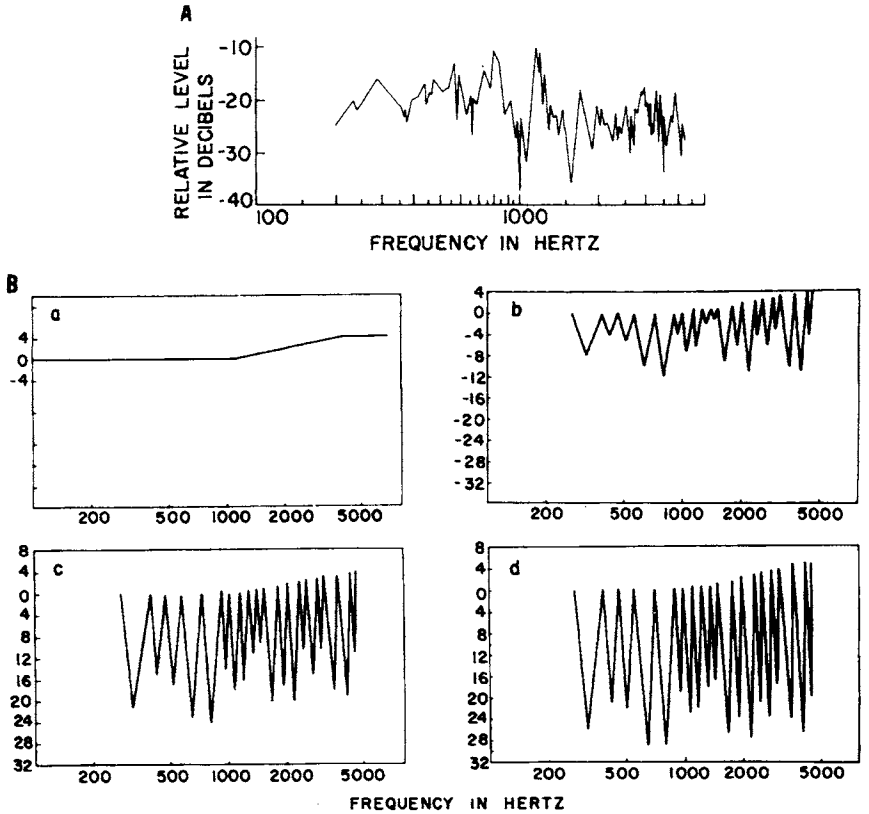

FIGURE 6 Relative frequency response: (A) as measured in a real violin from sine-wave excitation; (B) as simulated in the electronic replication of the violin tone: from (a) to (d), the Q of the simulated resonances increases from a value that is too low to a value that is too high (Mathews & Kohut, 1973).

to-valley ratio of about 10 dB in the response curve of the resonant filter. With too small a damping, the sound was even but dull; with too great a damping, the sound was hollow and uneven in intensity for various pitches.

The experimental equipment constitutes an electronic violin, which has been used musically to obtain either violin-like tones (e.g., in a quartet by M. Sahl) or sounds of very different qualities, by changing the filter settings (e.g., in pieces by V. Globokar, M. Urbaniak, or R. Boulanger).

This experiment has suggested to us that the specific quality of violin vibrato could be due to the interaction of the frequency modulation with the resonant peaks, producing a complex pattern of spectral modulation. Different harmonics are modulated in different ways, depending on the slope of the frequency response at the position of these harmonics. This can be verified by synthesis: the effect is not very sensitive to the parameters of a jagged frequency response. Imitative synthesis of violin tones (Schottstaedt, 1977) indicates that a good vibrato quality may be obtained in apparently simpler ways, not ensuring a truly fixed spectral envelope, but modulating the spectrum in a complex way. Actually the latter imitation was not obtained by subtractive synthesis, but through a variant of Chowning's audio frequency modulation, a powerful nonlinear technique evoked in the next section.

X. GLOBAL (OR NONLINEAR) SYNTHESIS

The additive synthesis method has been compared to the building a house by piling bricks, and subtractive synthesis has been likened to the sculpture of a block of stone by removing the unwanted material. There are ways to perform synthesis in a global way, by taking sound material and altering its shape, distorting it as when modeling clay. Such methods are nonlinear and hence difficult to understand and control. Global synthesis is more complex and less intuitive than additive or subtractive synthesis: but it is powerful and important.

Changing the shape of a sine wave while keeping the wave periodic will produce a new wave; in the Fourier analysis of this wave, one will find other components than the fundamental: hence this distortion has enriched the spectrum. This process is called nonlinear distortion or waveshaping. It can be implemented by taking a prescribed nonlinear function of a sine wave. When the amplitude of the sine wave increases, the amount of distortion also increases, so the spectrum gets richer. It is possible to determine a function that yields a desired spectrum through the distortion of a sine wave of a specific amplitude (cf. Arfib, 1979; Le Brun, 1979; Roads, 1996).

Another process uses frequency modulation (FM). FM is commonly used to emit radio waves. In the domain of musical sounds, FM is known as vibrato: for instance, the violin player modulates the audio frequency he plays (the carrier frequency) by slow periodic motions of his left hand, at a modulating frequency of about 6 Hz. This is heard as a specific tone quality rather than a frequency change.

Chowning had the idea to frequency modulate a sine tone by another sine tone, both tones being in the audio range. This results in a complex spectrum that is harmonic if the frequencies of the two tones are the same or in simple harmonic relation. The width of the produced spectrum—a quite salient parameter for the ear—depends on the amount of modulation—the so-called modulation index. By carefully controlling such parameters and their evolution in the course of the sound, Chowning (1973) has been able to produce a huge variety of timbres with dynamic spectra, including convincing simulation of many musical instruments and of the singing voice. Frequency modulation (FM) makes it possible to produce rich dynamic spectra including as many as 50 components with only three or four oscillators, whereas additive synthesis would require at least 100 oscillators. The string simulation mentioned earlier (Schottstaedt, 1977) used an elaboration of this technique, called multiple-carrier frequency modulation: this permits independent control of several formant regions.

Thus FM permits easy specification of certain aspects that are important for specifying timbre, particularly the global control of dynamic spectra. It is often said that FM sounds have an unmistakable sound of their own, like a trademark; but this is not necessarily so. Most users apply FM in a crude and lazy way, often with a high modulation index and very simple controls, which does result in stereotyped effects. But FM can be used with subtlety to achieve extremely fine results, as Chowning demonstrated with his imitations of the human voice (1980), or his interpolations between different timbres—*morphings* (Tellman, Haken, & Holloway, 1995) similar to the gradual metamorphosis, often performed today in image processing, of a shape into another shape.

The FM synthesis technique was implemented in the Yamaha DX-7 commercial digital synthesizers, which appeared in 1983: they met considerable success because they were affordable, Musical Instrument Digital Interface (MIDI)- compatible—this was the beginning of the MIDI standard (Loy, 1985)—and above all, because they provided a variety and quality of timbres unprecedented for commercial synthesizers. In addition to a host of ready-made "factory voices" imitating acoustic instruments or providing other timbres, users could do their own "voicing," that is, develop their own personal timbres by using the programming capabilities of the synthesizers (Chowning & Bristow, 1987). The control of the parameters has to be performed through MIDI. It is important to recall that the variety and quality of the timbres were not due simply to technological advances, but primarily to the know-how developed in the research of Chowning on timbre: technology enabled this know-how to be implemented at low cost.

XI. PHYSICAL MODELING AS A SYNTHESIS TECHNIQUE

Synthesis techniques described so far attempt to imitate the sound of the instrument: but the processes involved in the sound generation have little or no similar-

ity with the working of the instrument. Hence it is often difficult to go from a signal produced physically to a set of synthesis parameters. When one alters these parameters, the heard changes are related in a complex way to the alterations performed: thus the auditory result is hard to control. On the other hand, instead of approximating the final result (the wave), one can try to emulate the physical processes at work in the instrument that physically produced this result. It is indeed possible to mathematically model the motions of the vibrating parts in a musical instrument.

This ambitious approach was first implemented by Hiller and Ruiz (1971) in their work on the use of physical models for string sounds. The analytical study of the acoustic behavior of an instrument can lead to differential equations governing the motion of the vibrating elements. One can try to synthesize the sound by solving these differential equations. This approach is in a way the reciprocal of the approach used in analog computers, in which one assembles a physical system with parameters governed by the equations to be solved. In the latter case, the measurement of these parameters gives solutions to these equations. In the study of Hiller and Ruiz, the resolution of the differential equations gives an approximation to the sound of the instrument. This approximation may be a good one if the differential equations embody a good physical model. Actually time represented on a digital computer is quantized: the differential equations of motion are approximated by finite difference equations. Ruiz could produce convincing demonstrations of the behavior of the violin strings. He also demonstrated the intriguing possibility of choosing completely unrealistic values for the physical parameters, for instance, negative stiffness. The numerical approximations, however, produced some spurious effects. Also, the considerable amount of computation required seemed prohibitive at the time.

Despite these difficulties, physical modeling has great virtues. The model is controlled through physical parameters, for instance length, tension, and stiffness of a string, which tend to be much more basic and intuitive than signal processing parameters such as component envelopes or modulation index. In fact, the early experience of sound synthesis has shown that timbres tend to have stronger identity for sounds that the listener can plausibly ascribe to a specific acoustic generation process—hitting, scraping or blowing—even though nothing hits, scrapes, or blows within a computer program (Freed,1990; Freed & Martens, 1986; also see Section XV,A). Physical modeling synthesis provides parameters that are closely related to physical features of the sound source, and so, by controlling them, one is likely to produce salient effects on the ear. This is indeed confirmed by experimenting with physical sonic models.

Thus physical modeling synthesis has gained popularity, especially with computers getting faster and faster. The basic technique used by Hiller and Ruiz was considerably developed by Cadoz, Luciani, and Florens (1984, 1993). To implement synthesis via physical modeling, they have written a modular compiler called Cordis, similar to synthesis programs such as MusicV (Mathews, 1969), except that the modules, instead of being signal synthesis or processing elements,

emulate material elements (masses) and link elements (springs and dampers). These modules can be assembled into a network, and driving forces can be applied at various places. This permits simulation of vibrating strings or other mechanical objets, from a bouncing ball to a piano's double action. The simulation restores the sensory aspects of the objects (acoustic, but also visual, gestural, and tactile) by using specially devised transducers. The simulation can be done in real time, so that the user can operate in an "instrumental situation" (cf. footnote 5, later)—one would now speak of virtual reality. This requires a powerful dedicated computer installation with a specific architecture and house-built gestural-tactile interfaces: thus it has only recently begun to be used by musicians.

Morrison and Adrian (1993) have developed physical modeling synthesis using a modal approach, well known in mechanics: vibrating objects are represented as a collection of resonant structures that vibrate and interact together. Each structure is characterized by a set of natural modes of vibration. In the modal approach, the control parameters bear some similarity to those of additive and subtractive synthesis—a mode is similar to a resonant filter.

Smith (1992, 1996) has proposed simplified and powerful physical models. His method uses so-called waveguide filters—digital filters emulating propagation in a vibrating medium (e.g., along a string or within a pipe). Vibrating media, divided into sections of constant impedance, can be represented by pairs of delay lines. Several sections are coupled: waves propagate forward and backward through each section, which causes time delays. At interfaces between sections, absorption, transmission, and reflection occur. The method was first efficiently implemented for plucked strings by Jaffe and Smith (1983), from a modification of the so-called Karplus-Strong technique (Karplus & Strong, 1983). The imposed perturbation propagates back and forth along the string: here the model is very close to physical reality. Jaffe produced in his piece *Silicon Valley Breakdown* a tone sounding like a huge plucked string, bigger than the cables of the Golden Gate Bridge. Later, waveguide synthesis was adapted to wind instruments (cf. Cook, 1992; Smith,1992), to the vocal tract (cf. Cook, 1993), to the piano (Smith & Van Duyne, 1995) and to two-dimensional structures such as plates, drums, and gongs (cf. Van Duyne & Smith, 1993). The control of synthesis applies to parameters that have physical significance. Based on the digital waveguide technique, a "virtual acoustics" commercial synthesizer, the VL1, was introduced by Yamaha in 1993. Although its polyphonic resources are limited, the user can interactively act upon quasi-physical parameters and achieve robust and lively timbres.

Besides synthesis, physical modeling provides insight on the working of acoustic instruments. For instance, Weinreich (1977, 1979) has shown the contribution to the tone of the piano of the coupling between strings that are not exactly tuned to the same frequencies (this ensures the prolongation of the tone as well as a specific quality): he is currently applying this model successfully to the synthesis of piano-like tones. McIntyre, Schumacher, and Woodhouse (1983) have given a general formulation of oscillations in musical instruments, stressing the importance of nonlinear interaction between the exciter and the resonator. Keefe (1992)

has applied this model to the synthesis of wind instruments. A recorder flute model has been developed by using advanced notions of fluid dynamics (Verge, Caussé, & Hirschberg, 1995). One should mention the growing attention given to nonlinearities in physical models (cf. Smith, 1996).

Physical modeling is clearly not limited to known acoustic sounds, because one can give unrealistic values to the physical parameters, vary them in ways that are not feasible in actual acoustic devices, and even program arbitrary physical laws. However, physical modeling is unlikely to make other methods obsolete. It is very complex and very demanding in terms of computational power. The necessary approximations can introduce artifacts of their own. It is quite difficult to find the proper physical model to emulate a given sound. Physical modeling has problems producing certain variants or metamorphoses of the original sound (cf. Section XV,C), because the implemented physical model implies some robust timbral characteristics. Physical modeling may not be able to produce certain sounds that can be obtained otherwise. For instance, one does not at this point know how to produce with physical models the auditory illusions or paradoxes demonstrated by Shepard (1964) and Risset (1971, 1989) using additive synthesis. This, however, might be feasible through physical models relying on different laws of physics: perception could be puzzled by sounds from a separate, different reality.

XII. SAMPLING

Given their overwhelming popularity, it seems appropriate to discuss "samplers." In the mid-1980s, a number of firms introduced musical instruments that stored sounds in memory and played them back on demand at different transpositions. Ironically, some of these devices are called sampling synthesizers. Actually, no synthesis and certainly no analysis is involved at all. The desired sounds are simply recorded into memory and played back in performance with little modification other than a typically small transposition, a bit of filtering, and some waveform segment looping to prolong the duration. Modern samplers are sophisticated sound playback devices and can be used to layer a large mixture of sounds.

More than just a popular means of producing electronic musical sound, samplers have come to dominate the electronic musical instrument industry. By the mid-1990s, the synthesis techniques described in the previous sections actually lay claim to only a tiny percentage of the commercial electronic musical instrument market. The FM synthesis technique has been all but abandoned by Yamaha, and their efforts to introduce a synthesizer based on the waveguide acoustic modeling technique has not been particularly successful commercially so far.

Why this present decline in the commercial interest in synthesis and the domination of sampling? We believe that there are several central reasons. Samplers provide a simple means of reproducing sounds accurately no matter how complex the sounds may be. By contrast, parameter settings for global and acoustic modeling synthesis methods are difficult to make and it turns out that only a small num-

ber of persons have mastered these synthesis voicing techniques. As we pointed out earlier, these global and acoustic modeling synthesis methods are limited to certain types of sound material and lack the generality or "take-on-all-sounds" character of the sampling approach. One cannot ignore familiarity as a factor. Many musicians like to reproduce familiar sounds such as those of the traditional instruments, be they acoustic instruments or vintage electronic, and it seems apparent that the demand for acoustic accuracy is higher for such familiar sounds.

Another important factor is impoverished control. Acoustic instruments are controlled by carefully learned gestures. For synthesizers, the MIDI keyboard is by far the most dominant controller (Loy, 1985; McConkey, 1984). These keyboard controllers start sounds, stop them, and control at the start of the sound the intensity at which the sound will be played. With few exceptions, like globally applied pitch bend and modulation, keyboards provide little in the way of control over the evolution of sounds once they have been initiated. Keyboards do not seem well adapted to the production of the expressive phrasing and articulation one hears in a vocal line or in that produced by a solo wind or bowed-sting instrument. Because almost no commercially available alternatives to the trigger-oriented keyboards and percussion controllers exist, there has been little demand for synthesizers that accommodate alternative forms of musically expressive control. It would seem then that because of the lack of alternatives to the trigger-oriented controller, the sampler may continue in its role as a satisfactory means of producing sound. However this could change with the commercial implementation of analysis-based synthesis methods (cf. Sections XIV and XV,D).

XIII. THE IMPORTANCE OF CONTEXT: MUSICAL PROSODY, FUSION, AND SEGREGATION

The importance of a given cue depends on context. For instance, details of the attack of trumpet-like tones (especially the rate at which various partials rise) are more significant in long sustained tones than in brief or evolving tones (Risset, 1965, 1966). In the case of a very short rise time (as in the piano), the subjective impression of the attack is actually more determined by the shape of the beginning of the amplitude decay (Schaeffer, 1966). The acoustics of the room may also play an important role (Benade, 1976; Leipp, 1971; Schroeder, 1966). The sound of an organ, for instance, depends considerably upon the hall or church in which it is located.

Most of the exploration of timbre by analysis and synthesis has focused on isolated tones, but music usually involves musical phrases. Throughout these phrases, the physical parameters of the tones evolve, and this evolution can obscure the importance of certain parameters that are essential for the imitation of isolated tones. Similarly, in the case of speech, the parameters of isolated acoustic elements (e.g., phonemes) undergo a considerable rearrangement when the ele-

ments are concatenated to form sentences. The specification of simple and valid models of this rearrangement is the problem of speech synthesis by rule. The importance of prosodic variations throughout the sentence is obvious in speech; pitch bends and glides—even subtle ones—are also essential in music. In a musical context, the evolution of various parameters throughout a musical phrase can be significant. The prosodic variation of one parameter may subjectively dominate other parameters. The difficulty in controlling phrasing is a major limitation of keyboard sampling synthesizers (cf. Section XII), which perform mere frequency transposition of the recorded sounds but do not make it easy to perform changes from one sound to the next. So it is essential to study musical prosody by analysis and synthesis. Actually, this appears to be the new frontier for exploration of analysis and synthesis. Of course, one had to first understand the parameters of isolated tones to be able to describe how they evolve in a musical phrase. The attempts made by Strawn (1987) to characterize note-to-note transitions show their variability, depending on performance techniques, and the importance of the context.

Currently, musical prosody studies appear difficult because the phrasing is likely to depend on the musical style. Its importance seems greater, for instance, in Japanese shakuhachi flute playing than in Western instrumental playing. In the latter, the music is built from fairly well defined and relatively stable notes from which the composer can make up timbres by blending, whereas in the former, the state of the instrument is constantly disrupted. Hence, a prosodic study on the shakuhachi is interesting, even necessary, because the sound can be described properly only at the level of the phrase. As initially suggested by Bennett (1981), Depalle has performed analysis and synthesis of musical phrases played by Gutzwiller, a shakuhachi master. Rodet, Depalle, and Poirot (1987) have been able to achieve proper resynthesis by concatenating elements similar to spoken diphones (in speech diphone synthesis, the word *Paris* is obtained by chaining together the diphones *Pa*, *ar*, *ris*) (cf. Rodet, Depalle, & Poirot, 1987): thus data on transitions are stored as well as data on sustained notes.

Mathews has used the GROOVE hybrid synthesis system (Mathews & Moore, 1970), which permits the introduction of performance nuances in real time, to explore certain correlates of phrasing, for instance, the role of overlap and frequency transition between notes in achieving a slurred, legato effect. Using his algorithms for trumpet synthesis, Morrill has looked for correlates of phrasing in the trumpet. Grey (1978) has studied the capacity of listeners to distinguish between recorded instrumental tones and simplified synthetic copies when the tones were presented either in isolation or in a musical context (single or multivoiced). He found that whereas multivoice patterns made discrimination more difficult, single-voice patterns seemed to enhance spectral differences between timbres, and isolated presentation made temporal details more apparent. This finding may relate to the phenomenon of *stream segregation* (Bregman, 1990; Bregman & Campbell, 1971; McAdams & Bregman, 1979); see also Chapter 9, this volume),

an important perceptual effect that can be described as follows: if a melodic line is made up of rapidly alternating tones belonging to two sets that are sufficiently separated, the single stream of sensory input splits perceptually into segregated lines. (Baroque composers, such as Bach, resorted to this interleaving of lines to write polyphonic structures for instruments capable of playing only one note at a time.) This segregation is helped by increasing the frequency separation between the lines. Studies by van Noorden (1975) and by Wessel (1979) indicate that the influence of frequency separation on melodic fission has more to do with brightness—that is, with spectral differences—than with musical pitch per se, which appears to be linked with Grey's finding on single-voice patterns. Note that timbre differences without much brightness disparity (for instance, trumpet and oboe, or bombard and harpsichord) do not induce stream segregation in rapidly alternating tones.

Chowning (1980) has performed syntheses of sung musical phrases that sound supple and musical. In addition to carefully tuning the tone parameters for each note, specifically a vibrato (frequency-modulation) partly quasi-periodic and partly random, he has given due care to the change of musical parameters throughout the phrase. He has found that the parameters had to vary in ways that are to some extent systematic and to some extent unpredictable. These changes seem to be essential cues for naturalness. In fact, as in keyboard samplers, the musical ear may be "turned off" by a lack of variability in the parameters, which points to an unnatural sound for which even complex details may be aurally dismissed.

In musical performance, phrasing and expressivity are taken care of by the performer, who makes slight deviations with respect to a mathematically accurate rendering of the musical score. Performance practice has been studied by analysis (Sloboda, 1988; Sundberg, 1983), but also by synthesis: rules have been proposed to change parameters throughout the phrase, according to the musical context, so as to make the synthesis less machine-like and more musical (Friberg, 1991, Friberg, Frydén, Bodin, & Sundberg, 1991; Sundberg & Frydén, 1989). Some of these rules depend on the style of music—for instance, the notion of "harmonic charge" is mostly significant in tonal music—but many seem universal. Basically, they aim at making the musical articulation clearer to the listener. Timbral aspects intervene in leading the phrases. One hopes that such performance rules will be used to expressive ends in electroacoustic and computer music.

In his study of the singing voice, Chowning has given strong evidence that the addition of the same vibrato and jitter to several tones enhances the fusion of these tones, a fact that was investigated by Michael McNabb and Stephen McAdams (1982). Chowning's syntheses strongly suggest that the ear relies on such micromodulations to isolate voices among a complex aural mixture such as an orchestral sound. By controlling the pattern of micromodulations, the musician using the computer can make an aggregate of many frequency components sound either as a huge sonic mass or as the superposition of several voices. As Bregman indicates, the more two voices are fused, the more dissonant they can sound: preparation of

a dissonance attenuates it by separating the dissonant voice from the rest of the chord. Fusion or segregation of two voices depends on their timbres and at the same time influences the perceived timbre(s).

Simultaneous fusion or fission and stream segregation are examples of perceptual organization, whereby hearing analyses the auditory scene and parses it into voices assigned to different sources (Bregman, 1990). Another aspect of perception is worth mentioning here: in a complex auditory situation, it often appears that one dominant feature can eradicate more subtle differences. The most striking aspect according to which the stimuli differ will be taken into consideration rather than the accumulation of various differences between a number of cues. Lashley (1942) has proposed a model of such behavior in which the dominant feature masks the less prominent features. This often seems to hold for perception in a complex environment. Certainly, in the case of musical timbre, which can depend on many different cues, context plays an essential role in assessing whether or not a given cue is significant.

XIV. ANALYSIS-SYNTHESIS AS FITTING PHYSICAL AND PERCEPTUAL MODELS TO DATA

Having described a number of significant studies of timbre by analysis and synthesis, we shall pause here to put these studies in a conceptual framework that will help us to understand possible applications of the analysis-synthesis approach.

A general scheme that we have found useful is shown in Figure 7. The analysis-synthesis process begins with a sound that is to be modeled. In these general terms, the analysis of a sound involves estimating the parameters of a model (e.g., in the Fourier analysis model the frequencies, amplitudes, and phases of a set of sine-wave components must be estimated). Once the parameters of the model have been estimated, the model can be driven with them to generate a synthetic version of the original sound. For our purposes, the appropriate goodness-of-fit evaluation technique is to make auditory comparisons between the original sound and its synthetic replica. If the analysis-synthesis model captures the essential perceptual features of the sound in a thorough way, then the listener should be unable to distinguish the difference between the original and the synthetic version.

The above criterion of validity characterizes what we call a *perceptual model*, as opposed to a *physical model*: the latter would mimic the physical mechanisms that give rise to the sound whereas the former simulates the sound through processes that may well not reflect the way the sound is really produced, provided the aural result comes close enough to the original. As we have seen, a good acoustic model can also be a good perceptual model; but the physical behavior of the sound-emitting bodies is very complex, and acoustic simulations require simplifications such that they can rarely sound faithful to the ear. Although hearing is very

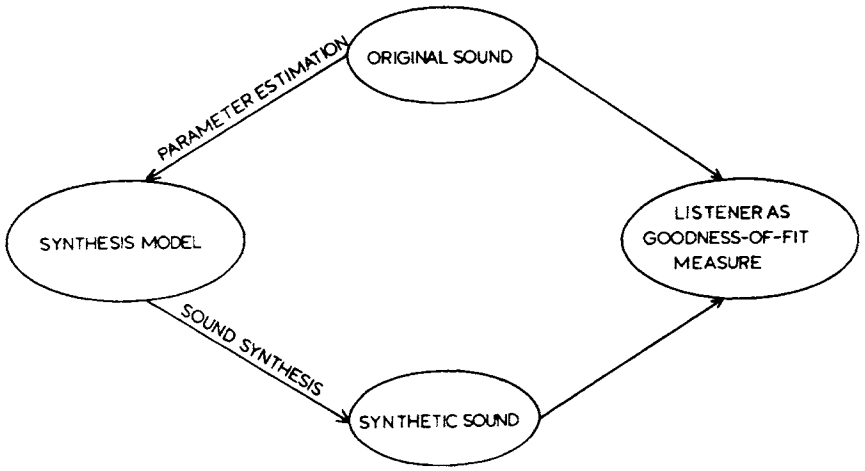

FIGURE 7 Conceptual framework of the analysis-synthesis process.

demanding in some respects, it is also very tolerant in other respects: perceptual models can concentrate on those features to which the ear is most sensitive. Recently, the expression *signal model* has been introduced to designate a model that mainly attempts to approximate the sound signal. Hence a signal model is in principle not a physical model. Is it a perceptual model? In general, a signal model calls for sophisticated signal processing techniques chosen so as to represent sounds in a way that can satisfy the listener: the auditory judgment is the final justification of the model. Yet the goodness-of-fit criteria used in the process of analysis are usually objective, especially when parameter estimation is programmed as an automatic procedure.

Physical and perceptual models often represent waveshapes in terms of certain mathematical functions. In the appendix, readers will find some general notions about representations of signals and their utility. A workable and reasonably general perceptual model is also described.

A lot of effort is devoted to the development of analysis procedures capable of automatically estimating parameters for a given synthesis model. This is a difficult problem, but one that is worth tackling. If this problem were solved, it would bypass the need for psychoacoustic know-how to recreate an existing timbre or to combine several timbres by cross-synthesis or morphing. It would turn synthesis techniques into powerful means of sound transformation. It would provide an easy way to inject the richness of arbitrarily recorded sounds, including acoustic sounds, into synthesizers based on the given synthesis model: such synthesizers have much more potential than samplers for sound transformation (see Section XV,D).

The need for parameter estimation was discussed by Tenney (1965), who called it demodulation. Justice (1979) was among the first to try to find an automatic

procedure to approximate a given spectrum using FM. This is still a research problem, although partial solutions are under way (Delprat, Guillemain, & Kronland-Martinet, 1990; Horner, 1996a; Kronland-Martinet & Guillemain, 1993; Tan, Gam, Lim, & Tang, 1994).

The problem of parameter estimation is quite difficult for physical modeling—it is compounded by the need to construct the model capable of producing a given sound, a difficult task. When one has chosen a physical model, procedures to extract the parameters have been developed for certain situations—they can apply only to sounds that are within the possibilities of the selected model. This research, however, is still quite preliminary (cf. Smith, 1996).

It is generally easier to predict the potential of various perceptual or signal models. For instance, additive synthesis is well adapted to the simulation of quasi-periodic tones, or of sounds consisting of a limited number of frequency components (spectral lines), and it has trouble with noisy sounds: it takes a lot of randomly spaced sine waves to give the ear the impression of white noise (about 15 per critical bandwidth, according to unpublished tests by Andrew Gerszo). A number of procedures for analysis-synthesis have been developed: they often attempt a rather crude approximation to the signal to be reproduced, yet, insofar as certain aurally salient features are properly emulated, the resynthesis may be acceptable to the ear. The most crucial feature seems to be the profile of spectral evolution—the *spectromorphological gait.* Thus the evolution of spectra should be approximated. There are various methods to do that. In the technique of analysis and synthesis of tones by spectral interpolation (M.-H. Serra, Rubine, & Dannenberg, 1990), a substantial amount of information is dropped, yet this method can produce faithful reproductions of the original timbres.

Various methods have been used to solve the identification problem necessary to extract the parameters automatically: classical least square methods, multidimensional scaling (Kramer & Mathews, 1956; Laughlin, Truax & Funt, 1990; Sandell & Martens, 1995; Stapleton & Bass, 1988; Zahorian & Rothenberg, 1981), neural networks (Lee, Freed, & Wessel, 1991, 1992; Lee & Wessel, 1992), genetic algorithms (Cheung & Horner, 1996; Horner, Beauchamp, & Haken, 1993). The latter techniques are recent and advanced, yet they are quickly becoming widely available.

Various processes of analysis-synthesis are available. Models of resonance (Barrière, Potard, & Baisnée, 1985) extract parameters of the instrument response to feed Rodet's CHANT synthesis program (Rodet, Potard, & Barrière, 1984); they work very well for percussive sounds such as bells and marimbas. This is also true of the Prony method, which analyses the signal in terms of a sum of decaying sine waves (Dolansky, 1960; Huggins, 1957; Laroche, 1989). For additive synthesis, variants of the phase vocoder have been developed that track the evolution of the signal (Dolson, 1986; McAulay & Quatieri, 1986). The wavelet transform has been elaborated to extract precisely the amplitude and frequency of spectral lines (Guillemain & Kronland-Martinet, 1992). X. Serra and Smith have developed an analysis/synthesis program for nonharmonic sounds; they later improved it by

separating the "deterministic" part, which can be emulated by a few additive synthesis components, and the "stochastic part," better evoked by using controlled random signals.

FFT^{-1} (Depalle & Rodet, 1992), as its name indicates, uses inverse fast Fourier transform for resynthesis: it could be said that it uses spectral frames, much like movie frames. It also implements the separation between deterministic and stochastic contributions to the sound. Implementation of FFT^{-1} has been specially optimized. This is a powerful method: its limitation may be spectral resolution and emulation of noise.

Spectral interpolation (M.-H. Serra et al., 1990) and wavetable interpolation (Cheung & Horner, 1996; Horner et al., 1993) are akin to cross-fading between fixed images: these are cruder techniques that can nonetheless be useful. To determine the optimal spectra or waveforms between which the synthesis process will interpolate, two techniques were used: principal component analysis and genetic algorithms.

We also mention below (cf. Section XVI) a fairly special technique implemented on pitched sounds using multidimensional scaling: it permits reconstitution of sounds, to perform morphings (interpolations between two sounds), to modify their duration, to alter them by keeping fewer dimensions from the representation (Hourdin, 1995).

XV. THE USE OF ANALYSIS-SYNTHESIS MODELS OF TIMBRE

The models drawn from analysis-synthesis of timbre can be useful for several purposes: (a) to provide insight and understanding, (b) for information reduction, (c) potentially to produce variants or modifications, and (d) in particular, to control musical prosody in real-time synthesis.

A. INSIGHT

Analysis-synthesis provides insight into the perception of timbre, which displays highly specific features. Many of these features can perhaps be better understood from an evolutionary perspective, considering the ways in which hearing has adapted to provide useful information about the environment. For instance, hearing is very sensitive to changes: it is well equipped to be on the alert, which makes sense because sounds propagate far and around obstacles. Perhaps this is why the musical ear tends to reject steady sounds as dull and uninteresting. Hearing is very sensitive to frequency aspects, which are only rarely modified between the sound source and the listener. Conversely, the ear is quite insensitive to the phase relations between the components of a complex sound, which is fortunate because these relations are smeared in a reverberant environment. Timbre is related to rather elaborate patterns that resist distortion (e.g., the relationship be-

tween spectrum and intensity in the brass). From these elaborate patterns, hearing has intricate ways of extracting information about loudness and distance (cf. Chowning, 1971, 1989): the listener can recognize that a sound has been emitted far away and loudly or close-by and softly. To do so, the listener resorts to fine cues: the ratio of direct to reverberated sound to appreciate the distance of the source, and also timbral cues to evaluate the energy of emission (e.g., for brassy tones, the spectral width, which increases with loudness). Segregation and fusion, mentioned in Section XIII, play an important role in helping the listener to disentangle a complex sonic mixture into components he or she can assign to different simultaneous sources of sound. Such extraordinary capacities have probably developed during the long evolution of animal hearing in the physical world: hearing plays an essential role for survival, because it provides warning and information about external events, even if they happen far away and out of view.

It appears that the listener tends to make an unconscious inquiry on the way the sound could have been produced in the physical world[5]: he or she tends to make a gross categorization of sounds in terms of their assumed origin. Schaeffer recommended hiding the origin of recorded sounds used in *musique concrète,* to prevent listeners from attaching labels to sounds, a reductionist trend that Smalley (1986, 1993) calls "source bonding." Hearing is adapted to performing "scene analysis" (Bregman, 1990): it can make fine identifications and discriminations with acoustic sounds, whereas it has more trouble differentiating sounds produced using straightforward electroacoustic or digital techniques—unless these latter sounds can be interpreted in terms of physical generation.

When such fine mechanisms are not called for, something may be lacking in perception. If electroacoustic or computer music plays only with amplification, it does not necessarily convey the feeling of sounds produced with energy. If the ear is deprived of fine cues for distance, the auditory scene seems to come from the plane of the loudspeakers, it sounds flat and lacks depth. On the contrary, by taking advantage of the specific mechanisms of hearing, one can produce a wide range of robust effects and vivid simulacra.

Models of timbre shed light on our capacity to assign different sounds to the same source, for instance, recognition of a note as such regardless of the register in which it is playing. The models help us to understand what properties form the basis of such categorization.[6] This understanding can be important in the fields of experimental music: a composer may want to confer some distinctive identity to certain artificial sounds.

[5]For instance, presented with recordings of metal objects struck with percussion mallets, the listener evaluates the hardness of the mallet (Freed, 1990). Such features make sense within the so-called ecological point of view on perception (Gibson, 1966; Neisser, 1976; Warren & Verbrugge, 1984).

[6]It seems clear that the identity of the timbre of an instrument such as the clarinet, whose high notes and low notes are physically very different, must be acquired through a learning process. It has been proposed that this learning process involves senses other than hearing. The experiments of Cadoz et al. (1984, 1993) aim at better understanding "motor" aspects of timbre perception, in particular how the gestural experience of producing a sound in a physical world interacts with its perception.

B. INFORMATION REDUCTION

Usually, we require that there should be many fewer parameters in the analysis-synthesis model than there are degrees of freedom in the data of the original signal. This is a form of data reduction. For example, consider a digitally sampled sound of 1-sec duration. If the sampling rate is 40,000 samples per second and if we wish to account for all these sample values in our model, then we could trivially simulate this signal with a model containing 40,000 parameters; however, a model with a reduced amount of information would be more practical.

In fact, much research on speech analysis-synthesis (e.g., the channel vocoders) has been performed to try to find a coding of speech that would reduce the bandwidth necessary to transmit the speech signal (Flanagan, 1972). Such a coding would in fact be an analysis-synthesis model because the speech would be analyzed before transmission and resynthesized at the other end (see Appendix). Such systems have only occasionally been put into practical use because it is difficult to preserve good speech quality and because the price of the transmission bandwidth has gone down substantially, so that the devices implementing analysis and synthesis at the ends of the transmission line would be more costly than the economized bandwidth. However, information reduction can work very well for certain types of sound, as we have already seen (Grey & Moorer, 1977): linear predictive coding is an economical way to store speech and is now used in portable speaking machines. In certain techniques for information reduction, specific properties of hearing are being taken advantage of, especially masking: it is useless to carefully transmit portions of the signal that are masked to the ear by the signal itself (cf. Colomes, Lever, Rault, Dehery, & Faucon, 1995; Schroeder, Atal, & Hall, 1979). This calls for sophisticated signal processing, as do a number of recent methods that can perform information reduction: simplified additive synthesis (Freedman, 1967; Grey & Moorer, 1977; Sasaki & Smith, 1980), multidimensional techniques (Kramer & Mathews, 1956; Sandell & Martens, 1995; Stapleton & Bass, 1988; Zahorian & Rothenberg, 1981), spectral interpolation (M.-H. Serra et al., 1990), multiple wavetable synthesis (Horner et al., 1993), and adaptive wavelet packets (Coifman, Meyer, & Wickerhauser, 1992).

The availability of powerful worldwide networks opens new possibilities in the domain of audio and music. Digital networks not only permit information about sounds to be conveyed but can transmit the sounds themselves. Digital networks even enable sound processing to be done on remote computers. This may give a new impetus to timbral manipulation in musical practice. However, sound uses a lot of bits compared with text, which can saturate the networks. Hence the growing success of networking gives a new motivation to research on information reduction. With the ever-increasing power of general-purpose computers, one can hope to use even complex coding techniques to transmit digital sound without needing special hardware to perform coding and decoding.

One should be aware that coding that permits substantial information reduction often takes advantage of specific properties of the signal to be compressed: hence

it will not work for all types of signals. In addition, such coding tends to be highly nonlinear, which makes it difficult or even impossible to perform even simple transformations such as amplification or mixing on the coded version of the sounds.

C. POSSIBILITY OF PRODUCING VARIANTS

If one manipulates the parameters before resynthesis, one will obtain modifications of the original sound; such modifications can be very useful. For instance, starting with a recording of a spoken sentence, one can change the speed by playing it on a variable-speed tape recorder; however, the pitch and the formant frequencies will also be changed, completely distorting the original speech. Now if one analyzes this sentence according to an analysis-synthesis process, which separates glottal excitation and vocal tract response (e.g., channel vocoder, phase vocoder, linear predictive coding [Arfib, 1991; Dolson, 1986; Flanagan, 1972; Moorer, 1978]), the resynthesis can be performed so as to alter the tempo of articulation independently of pitch. Using the Gabor transform, Arfib has been able to slow down speech or music excerpts by a factor of 100 or more without losing quality or intelligibility. Rodet, Depalle, and Garcia (1995) took advantage of analysis-synthesis to recreate a castrato's voice for the soundtrack of the widely distributed movie *Farinelli*. Through processing with their FFT^{-1} method, they achieved smooth timbral interpolations—*morphings*—between the recorded voices from a countertenor at the low end and a coloratura soprano at the high end. These examples show the usefulness of analysis-synthesis in obtaining variants of the original sounds.

We shall distinguish between two uses of sound modification: classical musical processing and expanding timbral resources. In classical musical processing, the goal is to transform the sound so as to maintain timbral identity while changing pitch and/or duration (also possibly articulation and loudness). For instance, as mentioned in Section VIII, linear predictive coding or phase vocoder analysis-synthesis permits the changing of pitch and speed independently. Also, as was discussed at the beginning of this chapter (see Figure 2), it is often improper to keep the same spectrum as one changes pitch. It may also be necessary to change the spectrum as one changes loudness. Such changes are essential if one wants to use digitally processed real sounds (e.g., instrumental sounds) for music. Without resorting to analysis-synthesis processes, one can perform only rather superficial and often unsatisfying modifications of the sound. On the other hand, one should be aware that these processes are complex and difficult to implement, especially in real time. Even a fast digital processor can have difficulty in coping with the demands of real time if it has to perform analysis-synthesis processes. The analysis part is especially difficult, whereas processors or even fast general-purpose computers can now cope with real-time resynthesis. However analysis can be performed in advance for a corpus of sounds: then one can work live to perform intimate transformations on those sounds. Such a process was put to work using

the FFT^{-1} method, implemented in an especially efficient way: powerful intimate transformations could be controlled interactively.

In expanding timbral resources, the goal is different: to change certain aspects of the tone so as to modify the timbre while preserving the richness of the original model. Here again, analysis-synthesis processes are essential for allowing interesting timbral transformations (like cross-synthesis), interpolation between timbres (Grey & Moorer, 1977), extrapolation beyond an instrument register, "perversion" of additive synthesis to produce sound paradoxes and illusions (Deutsch, 1975, 1995; Risset, 1971, 1978a, 1978b, 1978c, 1986, 1989; Shepard, 1964; Wessel & Risset, 1979), transformation of percussive sounds into fluid textures while preserving their frequency content, imposing a given harmony on sounds while preserving their dynamic characteristics[7] (see Figure 5). The extension of models can thus lead to the synthesis of interesting unconventional timbres, which is a fascinating area open to musicians.

D. CONTROLLING MUSICAL PROSODY IN REAL-TIME SYNTHESIS

As described earlier, analysis-synthesis permits transformation of the timbre and interpolation between different timbres. The intimate variations allowed by analysis-based synthesis permit the production of sound to be controlled so as to allow satisfactory prosodic control, even in real time.

As was mentioned in the discussion of samplers (Section XII), trigger-oriented control is not satisfactory to introduce prosodic specifications. It is often necessary to control the evolution of the sound once it has been initiated. Although MIDI keyboards are not adequate for this, the evolution of parameters could be controlled by means of other controllers (e.g., pressure-sensitive keys, breath controllers, bowlike sensors) or by functions of time. These functions could either be stored ahead of time and simply edited in real time or they could be generated by rule in real time. In any case, the synthesizing device should be able to perform the appropriate sonic changes. It is thus conceivable to design an "algorithmic sampler" (cf. Arfib, Guillemain, & Kronland-Martinet, 1992) by using analysis-based synthesis.

To date, no commercially available instrument has been introduced that applies the analysis-synthesis method to user-provided sound material. Although completely automatic "take-on-all" sound-analysis methods have yet to be fully developed, it would seem that they are within reach. Analysis-synthesis methodology

[7]Boyer has used a chromatic wavelet analysis to impose chords onto a signal (cf. Risset, 1991, p. 34). The musical transformations mentioned earlier can be heard in pieces like Morrill's *Studies for Trumpet and Computer*, Chowning's *Sabelith, Turenas* and *Phoné*, Risset's *Songes, Sud, Echo, Invisible*, Harvey's *Mortuos Plango*, Truax's *Pacific*, Reynold's *Archipelago* and *Transfigured Wind*, in many of Lansky or Dodge's pieces, on CDs quoted in footnotes 3 and 4, and on CDs Chenango CD1, CCRMA, Cambridge Street Records CSR-CD 9101, IRCAM CD0002, ICMA PRCD1300, Bridge Records BCD 9035, Neuma 450-74.

could become a core technology for electronic musical instruments. With just a little more engineering innovation, analysis-synthesis technology should accomplish for virtually all sounds the acoustic accuracy of recording playback as used in samplers. Thus research in analysis-synthesis is perhaps now more important than ever.

The advantages of decomposing musical sounds into controllable elements are significant. True, the traditional MIDI keyboard may not be able to use all potential for control but the users of sequencers and other composing and control software will be provided with many new expressive possibilities. In Figure 8, the relationships among sampling and synthesis techniques are presented to show the potentially important role for the analysis-synthesis methodologies. These methods should provide for both the acoustic accuracy of sampling and the musically expressive control as afforded by synthesis and yet be generally applied to a broad class of sounds.

Although Figure 8 indicates the value of analysis-based additive synthesis, in order to ensure musical expressivity, the analysis methodology should be informed by the musical controls that will be applied to the analysis-produced data.

XVI. TIMBRAL SPACE

We have discussed perceptual models; we have also said that analysis-synthesis is useful in modifying timbres. In this respect, it would be useful to have a good

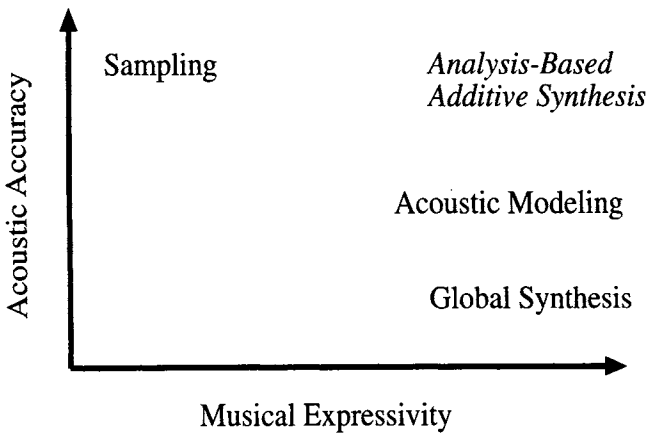

FIGURE 8 This figure displays the characteristics of various synthesis methods with regard to both acoustic accuracy and musical expressivity. Acoustic accuracy refers to the extent the electronically produced sound resembles the original sound, particularly with respect to the details. Musical expressivity refers to the extent to which the sound can be shaped and linked to other sounds. Here musical expressivity embodies the idea of control of phrasing and articulation as, for example, in vocal, wind, or bowed-string lines. (This diagram was first presented to us by Dana Massie, 1998.)

notion of the structure of the perceptual relationship between different timbres. This can be greatly eased by geometrical models provided by multidimensional techniques, which in effect provide displays of this structure. As was stated earlier by one of the authors: "A timbre space that adequately represented the perceptual dissimilarities could conceivably serve as a kind of map that would provide navigational advice to the composer interested in structuring aspects of timbre" (Wessel, 1973).

One can indeed propose geometric models of subjective timbral space such that individual sounds are represented as points in this space: sounds judged very dissimilar are distant, and sounds judged similar are close. The models are not constructed arbitrarily, but by asking subjects to rate for many pairs of sounds the dissimilarities between the sounds of each pair and by submitting the dissimilarity data to multidimensional scaling programs. These programs—strictly devoid of preconceptions about the data—provide a geometrical model that best fits these data. The dimensions of the model can then be interpreted (e.g., by investigating the stimuli that are least, or most, differentiated along these dimensions).[8] Wessel (1973, 1978) and Grey (1975) have thus already provided models of timbral space for string and wind instrument tones. These models unveil two dimensions—one that differs within the instruments of a same family (e.g., cello, viola, violin) and appears to relate to the spectral distribution of energy in the sound (cf. von Bismarck, 1974a, 1974b) and one that is the same within a family of instruments and seems to be linked to temporal features like the details of the attack (Grey, 1977; Wessel, 1973).

The corresponding representations of timbral space tempt one to fill the space, to draw trajectories through it, like the timbral interpolations mentioned earlier. According to Grey (1975), "The scaling for sets of naturalistic tones suggest a hybrid space, where some dimensions are based on low-level perceptual distinctions made with respect to obvious physical properties of tones, while other dimensions can be explained only on the basis of a higher level distinction, like musical instrument families." The intervention of cognitive facets, such as familiarity and recognition, indicates that a fully continuous timbre space may not be obtainable. Nevertheless, subjective space models can propose new paths and new intriguing concepts, such as that of analogies between timbral transitions (Wessel, 1979), which may permit one to do with timbres something similar to melodic transposition with pitches. Resolving timbre, that "attribute" defined as neither pitch nor loudness, into dimensions may uncover new features or parameters susceptible to precise differentiation and appropriate for articulating musical structures.

For instance, multidimensional scaling of timbre often unveils a dimension correlated with the frequency of the centroid of the spectral energy distribution, hence with that aspect of timbre termed *brightness.* As Wessel (1979) has shown, this

[8]From only quantitative judgments of dissimilarities between sounds, multidimensional scaling in effect unveils in what ways these sounds differ. Schaeffer failed to realize this in his criticism of the process as described by Babbitt (1965) (cf. Music and Technology (1971), pp. 77–78).

dimension is the one that can best articulate stream segregation (McAdams & Bregman, 1979). Here, isolating dimensions of timbres permits one to make predictions about the behavior of these timbres in context.

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 timbral space would be to attach the handles of control directly to the dimensions of the space. One of the authors examined such a control scheme in a real-time context (Wessel, 1979). A two-dimensional timbre space was represented on the graphics terminal of the computer that controlled the Di Giugno oscillator bank at the Institut de Recherche et Communication Acoustique/Musique (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, we need an efficient computer language for dealing with envelopes. 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.

As we have mentioned, multidimensional techniques have been used for information reduction. Hourdin, Charbonneau, and Moussa have used multidimensional scaling to provide representations of pitched sounds that permit resynthesis of those sounds by additive synthesis. This work seems quite promising (Hourdin,

1995). Data from the analysis of sounds are submitted to scaling to reduce their dimensionality. Each sound is represented by a closed curve in the resulting objective timbre space. The curve is similar to a timbral signature, because the perception of the timbre is associated with the shape of this curve. From the curve, one can go back to additive synthesis data. Operations on the curves permit the reconstitution of sounds, interpolation between sounds, modification of their duration, and alteration of the sounds by keeping fewer dimensions from the representation. Further work providing ad hoc graphic tools could permit us to realize other interesting timbral operations from these curves.

Another field that may yield interesting possibilities is the objective representation of musical sounds as trajectories in phase space (or state space), a representation used for dynamic systems (Gibiat, 1988). This field is specially informative for multiphonic sounds, which are an instance of chaotic behavior (Lauberhorn, 1996). In this spatial representation, time is not available; however, it appears to be possible to use this representation to reconstruct sounds with the same morphology but with different time and frequency characteristics.

Representations of timbral space are topographic. It has also been attempted to provide verbal descriptions of timbre (Bismarck, 1974a, 1974b; Kendall & Carterette, 1993a, 1993b) that could even be operational, in the sense that musical timbre synthesis or transformation could be based on descriptive terms rather than numerical data (Ethington & Punch, 1994). In the latter work, listening experiments were performed in order to map verbal descriptions to synthesis or processing operations, then the verbal descriptions were mapped to timbral features using a space with a fixed set of dimensions. Although this work is preliminary, early results indicate that musicians may be able to control effectively synthesis programs or computer-controlled synthesizers by describing in words desired timbral characteristics or transformations.

XVII. CONCLUSION

As we explained before, the exploration of timbre by analysis and synthesis can serve several purposes: it provides insight into the physical parameters of the sound and the relevance of these parameters to the resulting timbre; it leads to simplified models that permit data reduction in the synthetic replication of the sound; and it uses models to perform transformations on the original sound, either from the point of view of classical musical processing (e.g., by independently changing pitch, duration, articulation, and loudness) or by expanding timbral resources (rearranging at will the complex variations abstracted from the analysis to obtain new and rich sounds).

Exploration of timbre by analysis and synthesis is difficult but rewarding. Since the development of analysis and synthesis devices, in particular the digital computer and its descendants, understanding of the physical correlates of timbre has improved and recipes for new musical resources have been developed.

Although much remains to be done, these new possibilities available to musicians will probably increase the musical role of timbre. In classical Western music, timbres were used mostly to differentiate musical lines. Later, this linear organization was disrupted. Debussy often used chords for their timbral effect rather than for their harmonic function—he was followed by Messiaen, who coined the expression "accord-timbre" ("timbre-chord"). Varèse (1983) longed for novel sonic materials that would lend themselves to novel architectures. In one of his *Five Orchestra Pieces* Op. 16 composed in 1909 and subtitled "Farben" ("Colors"), Schoenberg kept pitch constant and varied only the instruments—according to his instructions, only the changes in tone color should be heard by the listener (cf. Erickson, 1975, p. 37). Schoenberg wrote in 1911: "the sound becomes noticeable through its timbre and one of its dimensions is pitch ... The pitch is nothing but timbre measured in one direction. If it is possible to make compositional structures from timbres which differ according to height (pitch), structures which we call melodies, ... then it must be also possible to create such sequences from the other dimension of the timbres, from what we normally and simply call timbre." In the 1950s, Boulez submitted successions of timbre to serial organization. With the control of timbre now made possible through analysis and synthesis, composers can compose not only with timbres, but they can also compose timbres: they can articulate musical compositions on the basis of timbral rather than pitch variations. It has been argued that timbre perception is too vague to form the basis of elaborate musical communication; however, as Mathews has remarked, there already exists an instance of a sophisticated communication system based on timbral differentiation, namely human speech.[9]

Indeed, the exploration of timbre by analysis and synthesis has become an important musical tool. It requires the acute ear and judgment of the musician, some psychoacoustic know-how, and a good interactive environment helping him or her to achieve fine tunings and manipulations. The functional role of timbre in musical composition was already a central point in computer-synthesized pieces such as Risset's *Little Boy* (1968), *Mutations* (1969), *Inharmonique* (1977), Chowning's *Turenas* (1972), *Stria* (1977) and *Phone* (1981), Erickson's *Loops* (1976), Branchi's *Intero* (1980), Murail's *Désintégrations* (1983), Barrière's *Chréode* (1984), and Saariaho's *Jardin Secret I* (1985). The use of timbre in contemporary musical composition is discussed at length in the references by Erickson (1975), McAdams and Saariaho (1985), Emmerson (1986), Barrière (1991), and Risset (1994). The timbral innovations of electroacoustic and computer music have influenced the instrumental music of Varèse, Ligeti, Penderecki, Xenakis, Scelsi, Crumb, Dufourt, Grisey, Murail.

Hence the 20th century is that of the "eternal return of timbre" (Charles, 1980): from an ancillary function of labeling or differentiating, the role of timbre has extended to that of central subject of the music. Then, paradoxically, the very no-

[9]As Moorer demonstrated by analysis and synthesis, speech can remain intelligible under certain conditions after removal of pitch and rhythmic information.

tion of timbre, this catchall, multidimensional attribute with a poorly defined identity, gets blurred, diffuse, and vanishes into the music itself.

APPENDICES

A. SIGNAL REPRESENTATIONS AND ANALYSIS-SYNTHESIS PROCESSES

Analysis-synthesis according to a given process implies estimating the parameters of a model of the sound. This model may or may not be adequate; it may or may not lend itself to a good imitation of the sound. For instance, Fourier series expansion is a useful tool for periodic tones, and Fourier synthesis, using the data of Fourier analysis, indeed permits one to synthesize a faithful copy of a periodic sound. However, as was explained earlier, most sounds of interest are not periodic; hence, Fourier series expansion is inadequate to replicate, for instance, a sound whose spectrum varies with time.

A sound can be mathematically described by the waveshape function $p(t)$, giving the acoustic pressure as a function of time. Mathematics tells us that reasonably regular functions can be analyzed in a number of ways, that is, in terms of one or another set of basic functions. This set is said to be complete if an arbitrary function can indeed be obtained as the proper linear combination of these basic functions. (This proper combination is unveiled by the analysis process that consists of estimating the parameters of the corresponding model.) For instance, Fourier's theorem states that any periodic function (of frequency f) can be expanded as a linear combination of the sine and cosine functions of frequencies f, $2f$, $3f$, ..., so that this linear combination can be arbitrarily close to the periodic function. Hence, the set of sine and cosine functions of frequencies f, $2f$, $3f$, and so on is "complete" over the space of periodic functions of frequency f (cf. Panter, 1965; Rosenblatt, 1963).

Actually, the representation of nonperiodic signals in terms of basic functions usually requires an infinite number of basic functions so that the series expansion turns into a transformation. For instance, nonperiodic signals can be represented in terms of the so-called Fourier transform or Fourier integral, in which the discrete spectral components are replaced by a continuous amplitude spectrum; the discrete phases are also replaced by a phase spectrum. There are other transformations used for analysis-synthesis (e.g., the Walsh-Hadamard, the Karhunen-Loève, the Gabor, and the wavelet transforms). Such linear expansion in terms of a basic set of signals is similar to the expansion of a vector in terms of a set of basic vectors; it is practical (although not necessary for all purposes) to use orthogonal transforms—that is, to use functions that form an orthonormal (and complete) set (cf. Harmuth, 1972). For instance, using orthogonal wavelets as basic functions is a strong limitation, because it imposes an interval of 1 octave between two adjacent wavelets—whereas choosing an interval of a semitone can be musically use-

ful (cf. footnote 6). In addition, this choice does not allow an easy evaluation of energy (Guillemain & Kronland-Martinet, 1996).

The application of a given transform to a sound signal provides a representation of the signal that may be revealing and should make it possible to restore the signal by means of the inverse transform. Hence, the representation of signals is closely linked to analysis-synthesis processes. Actually, the representation of signals purports both to characterize the information (bearing elements in the signal) and to describe in a simple way the effect of modifications of the signals (like those introduced by an imperfect transmission system or by a deliberate simplification of the signal).

Although we cannot go into much detail here, we would like to make several points:

1. Certain analysis-synthesis processes and the corresponding representation are intrinsically limited to certain classes of signals. Others can be transparent if they are complete in the above sense—for instance, the Fourier, the Walsh-Hadamard, the Gabor and the wavelet transforms, the phase vocoder, and the linear predictive coding scheme. However, the two latter schemes allow reproduction of the original signal only at the expense of a considerably detailed analysis, an information explosion instead of an information reduction. This can be substantially simplified only for certain classes of signals (quasi-periodic signals with relatively independent excitation and response mechanisms, such as speech; for instance, linear predictive coding is efficient in simulating oboe sounds but poor for low clarinet sounds because eliminating the even harmonics is taxing for the filter). Indeed, much work on analysis-synthesis and signal transformation was originally directed toward efficient coding of speech information for economical transmission over technical channels (Campanella & Robinson, 1971; Flanagan, 1972; Schafer & Rabiner, 1975). It is also for certain types of signals that the representation of the signal will be most enlightening (but, for instance, phase vocoders' programs implemented by Moorer, 1978, have permitted Castellengo to obtain useful information on nonharmonic "multiphonic" tones). Certain sounds, especially percussive sounds, can often be modeled efficiently as a sum of decaying exponentials (Dolansky, 1960 ; Huggins, 1957; Laroche, 1989).

Similarly, Gabor's expansion of a signal into gaussian elementary signals—"sonic grains" (sine waves with a gaussian amplitude envelope; see Figure 9)—has been proven to be complete (Bacry, Grossman, & Zak, 1975; Bastiaans, 1980), and so has the wavelet transform (Grossman & Morlet, 1984). Hence, these methods can in principle produce exactly what Fourier or other types of synthesis can produce (cf. Gabor, 1947; Roads, 1978 ; Xenakis, 1971). The idiosyncrasies of different complete analysis-synthesis methods appear only in the modifications they permit—or suggest—in a simplified, archetypal use. There are strong differences, however, between the representations and the transformations they allow by altering the analysis data before resynthesis. For instance, Gabor's grains are the sonic elements for the so-called granular synthesis, explored by Xenakis, Roads, and Truax to create novel textures *ex nihilo* without performing any analy-

sis to extract parameters, but they have also been used by Arfib as the elementary signals for implementing the analysis-synthesis of existing sounds and for stretching them in time without altering the frequency content (cf. Arfib, 1991; Roads, 1996).

2. The Walsh-Hadamard transform seems promising because it leads to operations that are easy to implement with digital circuits. However, from a psychoacoustical standpoint, this transform is quite inappropriate. The basic functions do not sound elemental to the ear; they are spectrally very rich, and an approximated representation in those terms would lead to aurally unsatisfying results. The analysis-synthesis process does not deteriorate gracefully for the ear, and it has great difficulty in producing timbres that are not rich and harsh (for instance, it has trouble approaching a sine wave).

3. Fourier-type analysis (and synthesis) has been much criticized, often in a poorly documented way. Whereas Fourier series expansion is indeed inadequate for nonperiodic sounds, more elaborate variants of Fourier analysis have great utility. The Fourier transform provides complete information of an amplitude spectrum and a phase spectrum; however, the latter characterizes the evolution of the signal in time in a way that is unintuitive and very hard to use. Because this evolution in time is very significant to the ear, one needs some kind of running analysis to provide so-called sonagrams, that is, diagrams in the time-frequency plane showing the evolution of the frequency components as a function of time—the amplitude of the components being indicated by the blackness and thickness of the lines. This is obtained by calculating, as a function of time, the spectrum of the signal viewed through a specified *time window* (also called *weighting function*), which at any time only shows the most recent part of the past values of the signal. Such representations are very useful: they have been used in several of the studies previously described. The sound spectrograph (Koenig, Dunn, & Lacey, 1946) implements this type of running analysis: its windows are appropriate for a useful portrayal of speech sounds, but it often displays significant features of music as well (Leipp, 1971), even though the analysis is often too crude to provide data for a proper synthesis. Sonagrams give revealing and useful "photographs" of the sound (Cogan, 1984). They can now be produced by computer programs. If these programs preserve the phase information, the sound can be reconstituted: the short-term Fourier transform is also complete—it bears strong similarity to the Gabor transform.

The significance of Fourier analysis has a multiple basis. Clear evidence exists that the peripheral stages of hearing, through the mechanical filtering action of the basilar membrane, perform a crude frequency analysis with a resolution linked to the critical bandwidth (Flanagan, 1972; Plomp, 1964). The distribution of activity along the basilar membrane relates simply to the Fourier spectrum. Thus features salient in frequency analysis are of importance to perception. Sonagrams reveal much more about sounds than oscillograms do: Figure 1 shows that the inspection of waveforms is not very informative and can be misleading, even though the waveform of course contains all the information about the sound. Also, when the

sound is quasi-periodic, the phase deafness of the ear (Figure 1) permits a substantial reduction of information. One can also in this case take advantage of the concentration of energy at the harmonic frequencies to describe the sounds by the evolution in time of the amplitude of few harmonics. We have seen that such additive synthesis was a very useful model (cf. Grey & Moorer, 1977; Guillemain & Kronland-Martinet, 1996; Keeler, 1972; Risset & Mathews, 1969).

Yet it may seem bizarre to try to represent a finite signal, such as a tone of a musical instrument, as a sum of sine waves, which never begin or end: this is only possible if the sum is infinite. Thus the Fourier transform of an unlimited sine wave consists of a single spectral line; but if this "sine wave" has a beginning and an end, its Fourier transform includes a continuous frequency-band, which is not easy to deal with. The time windows used for the so-called short-time Fourier transform deal with this problem through a kind of ad hoc mathematical treatment. As Gabor initially proposed, it may be more logical to try to represent a finite signal as a sum of basic functions that are limited in time—as well as in frequency. The basic functions proposed by Gabor are sine waves multiplied by a Gaussian time window of fixed length: the center frequency of such a function corresponds to the frequency of the sine wave. Both the time and the frequency resolution are determined by the time and spectral extension of the Gaussian window (see Figure 9). Theoretically this function is not limited in time, because the Gaussian function goes to zero only at infinity, however, this function is limited in practice, because sounds are represented with a finite number of bits. The expansion into Gabor's gaussian functions has been implemented; it permits faithful reconstitutions and useful transformations (Arfib, 1991).

A new representation, initially derived from Gabor's ideas by Morlet, is based on the "wavelet," an elementary function of constant shape rather than with a constant envelope: to vary the frequency, this function is shrunk or stretched in time (Figure 9). Thus the frequency resolution does not correspond to a constant frequency amount, but rather to a constant interval (e.g., one octave); conversely, the time resolution is finer at higher frequencies—a so-called constant Q or multiresolution process. Grossman generalized this representation by showing that one can expand well-behaved functions in terms of more general wavelets of constant shape, the requirements for a function to qualify as a wavelet being the following: the function must be limited in time, it must oscillate, and its average must be zero. This provides a quite general *time-scale* representation, rather than time-frequency, because the successive frequencies of the basic wavelets are scaled by the same factor. Thus the wavelet representation is well adapted to the investigation of fractal phenomena. It has already found applications in various domains, including sound and music (cf. Combes, Grossman, & Tchamitchian, 1989; Grossman & Morlet, 1984; Guillemain & Kronland-Martinet, 1992; Kronland-Martinet, 1989; Kronland-Martinet, Morlet, & Grossman, 1987). Such a universal method of representation cannot work miracles in itself, it has to be adapted to the specific problem at hand; however, it has some computational and practical virtues, including making it possible to choose specific basic wavelets. The wavelet representation

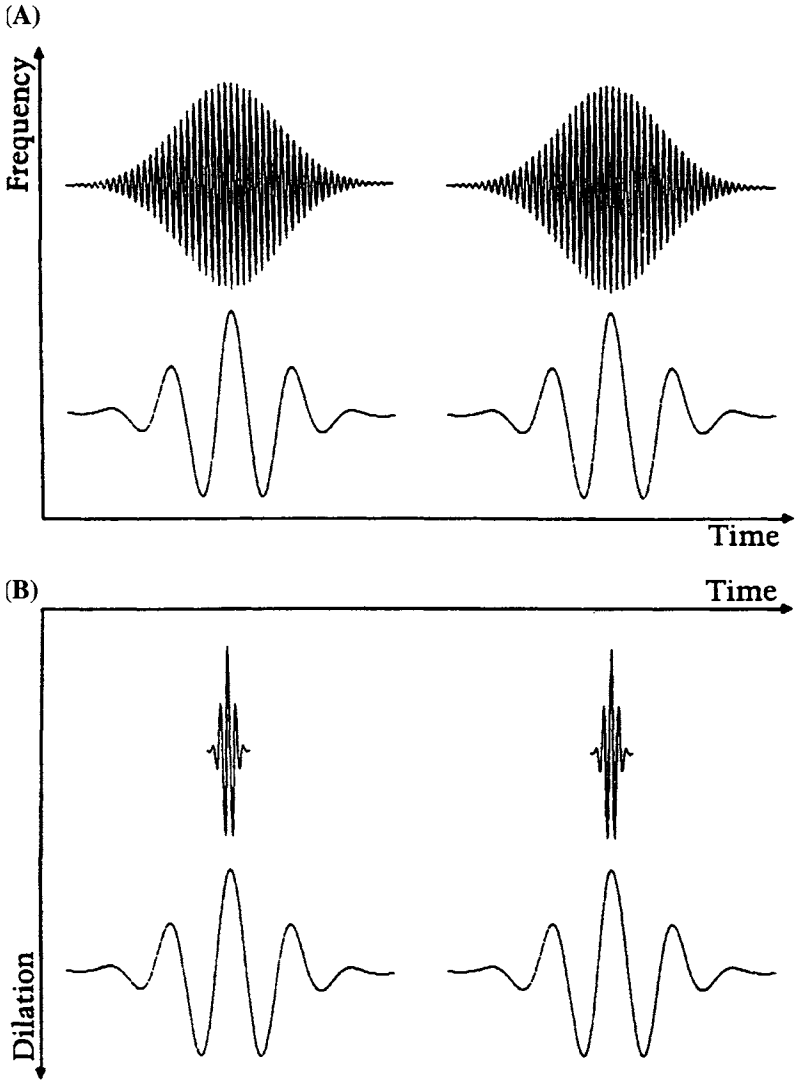

FIGURE 9 (A) Elementary grains used for the Gabor expansion in the time-frequency domain. (B) Elementary Morlet wavelets used for the wavelet expansion in the time-dilation domain.

has already be shown to permit faithful reconstitutions and to offer intriguing possibilities.

4. Auditory models have been exploited in the analysis-synthesis process, and their application shows promise. Auditory periphery models or "ear" models have been implemented in the form of computer programs and are able to reproduce via simulation a significant body of psychoacoustical phenomena. The goal is to simulate the various stages in the auditory periphery and supply as output a neural

code similar to that of the auditory nerve (Flanagan, 1962; Schroeder, 1975). Both physiologically inspired models (Duda, Lyon & Slaney, 1992; Slaney & Lyon, 1993) and more psychoacoustically inspired models (Licklider, 1951; Moore & Glasberg, 1981; Patterson, 1976) involve a filter bank at an early stage, followed by processing modules that implement a form of autocorrelation or periodicity analysis in each of the frequency bands. Thus they can track pitch, but also spectral variations, which often play an important role in the recognition of timbre.

These models have been shown to be useful as preprocessors to speech recognizers as well as in the determination of the pitch of complex time-variant tones. There is also a significant body of work in the newly emerging field called computational auditory scene analysis (Brown & Cooke, 1994a, 1994b), a follow-up to auditory scene analysis (Bregman, 1990); cf. Sections XIII and XV). The scheme for auditory organization proposed incorporates the lower level auditory or ear models into more elaborate models of auditory cognition. These models are being applied for the most part to the problem of auditory source separation, also known as the "cocktail party effect" in reference to the practice of selective listening in a din of conversation.

In a theoretical paper, Yang, Wang and Shamma (1992) demonstrated that a physiologically plausible model of the auditory periphery can be inverted and used to regenerate the original input completely. This is in an interesting result in that it suggests that the auditory system does not lose information about the acoustic signal in its early stages of processing. Slaney has inverted the Slaney and Lyon ear model and synthesized signals using data coming from the analysis of the signal by the ear model. He has applied aspects of this inversion method to sound morphing or timbral interpolation (Slaney, 1996; Slaney, Covell, & Lassitier, 1996).

5. Global methods like frequency modulation and nonlinear distortion, described in Section X, are appealing because they provide powerful control over salient features of the spectrum in terms of few parameters: the amount of specification and processing is much reduced as compared with additive synthesis. However, strength is at the expense of generality. Certain specific results are hard to achieve with the global methods, unless one uses them in refined ways that can quickly become quite complex (cf. Schottstaedt, 1977). It is difficult to develop analysis techniques extracting the parameters for resynthesis using a global model (see Section XIV).

B. A SYNTHESIS MODEL BASED ON PERCEPTUAL PRINCIPLES

We now give a brief account of how synthesis procedures can provide for direct control over some essential perceptual attributes of timbre. The essential principle underlying these synthesis schemes is the decomposition of a musical signal into perceptual attributes that are, for the most part, perceptually independent of each other. The motivation is to provide a reasonably general but simple control scheme

for additive synthesis as this form of synthesis is becoming more and more practical with advances in the development of high-speed digital-synthesis hardware.

1. Pitch Versus the Global Impression of the Spectral Envelope

Several studies (Plomp & Steeneken, 1971; Risset, 1978b, 1978c) suggest that musical pitch and the global spectral energy distribution as perceptual attributes are reasonably independent of each other. This is true to a large extent for harmonic tones that tend to produce clear pitch percepts, but it is not true for inharmonic spectra whose ambiguous and otherwise multiple-pitch content depends on the spectral balance of the components. What we mean by independence is that it is possible to manipulate, for example, the placement and shape of formants without influencing the perceived pitch, and conversely, it is possible to manipulate the pitch while keeping the perceived shape of the spectrum constant. The voice provides an example of such an independent control scheme that operates over a reasonably wide range of pitches and spectral shapes. A singer can sing the same pitch with a large variety of vowel qualities and can likewise maintain a constant vowel quality over a substantial range of pitches.

2. Roughness and Other Spectral Line-Widening Effects

Terhardt (1978) has provided evidence that our impression of roughness in sounds depends on an additive combination of independent spectrally distributed amplitude fluctuations. Consider the following example using a tone consisting of the three components: 400, 800, and 1600 Hz. Here the components are widely distributed (i.e., more than a critical bandwidth between them) and amplitude fluctuations of say 10% of the component amplitude at frequencies between 10 and 35 Hz contribute independently to the overall impression of roughness. The implication for synthesis is to provide for independent control of the amplitude fluctations in different regions of the spectrum.

By *spectral line widening,* we mean the spreading or smearing of energy around a spectral line. Such spectral line widening can be obtained by amplitude or frequency modulation of a sinusoid. Many instrumental timbres have noiselike effects in their attack transients and most often their spectral placement is essential to the timbre. For example, in the synthesis of stringlike attacks, the middle to upper spectral regions require more noise than the lower regions. It is to the synthesis model's advantage to allow for the independent placement of noiselike effects in separate spectral regions, which can be accomplished by widening the spectral lines in those regions.

3. Vibrato and Frequency Glides

Our impression of timbre is often strongly dependent on the presence of a vibrato or frequency glide, and the synthesis procedure should provide for an easy application of these effects without disrupting the global spectral energy distribution. A frequency glide of an oscillator with a fixed spectrum results as well in a glide of the spectral energy distribution and thus violates the desired indepen-

dence. Such independence has been accomplished in the glissando version of Shepard's illusion (Shepard, 1964) produced by Risset (1969, 1978b, 1978c, 1986, 1989; cf. Braus, 1995).

In our additive synthesis procedure, we should be able to provide an overall spectral envelope that remains constant in spite of changes in the specific frequencies of the components. In addition, the model should provide for the independent placement of roughness and noiselike effects in separate regions of the spectrum again without violating the overall spectral envelope. These kinds of control can be accomplished fairly easily in most sound synthesis languages by the use of table-look-up generators, such as the VFMULT of the MUSIC V language. These generators allow one to store a spectral envelope function that is used to determine the sample-by-sample amplitude of a given component that could be executing a frequency glide. This technique works similarly for control of the spectral distribution of roughness or other line-widening effects. To obtain time-variant effects with these attributes, the spectral envelopes and roughness distributions are defined at successive and often closely spaced points in time, and interpolation is carried out between successive pairs of these functions.

REFERENCES

Allen, J. B., & Rabiner, L. R. (1977). A unified approach to short-time Fourier analysis and synthesis. *Proceedings of the IEEE, 65,* 1558–1564.

Alles, H. G., & Di Giugno, P. (1977). A one-card 64-channel digital synthesizer. *Computer Music Journal, 1*(4), 7–9.

American National Standards Institute. (1960). *USA standard acoustical terminology.* New York: American National Standards Institute.

Appleton, J. H., & Perera, R. C. (Eds.). (1975). *The development and practice of electronic music.* Englewood Cliffs, NJ: Prentice Hall.

Arfib, D. (1979). Digital synthesis of complex spectra by means of multiplication of non-linear distorted sine waves. *Journal of the Audio Engineering Society, 27,* 757–768.

Arfib, D. (1991). Analysis, transformation, and resynthesis of musical sounds with the help of a time-frequency representation. In G. De Poli, A. Picciali, & C. Roads (Eds.), *The representation of musical signals* (pp. 87–118). Cambridge, MA: MIT Press.

Arfib, D., & Delprat, N. (1993). Musical transformations using the modifications of time-frequency images. *Computer Music Journal, 17*(2), 66–72.

Arfib, D., Guillemain, P., & Kronland-Martinet, R. (1992). The algorithmic sampler: An analysis problem? *Journal of the Acoustical Society of America, 92,* 2451.

Babbitt, M. (1965). The use of computers in musicological research. *Perspectives of New Music, 3*(2).

Backhaus, W. (1932). Einschwingvorgänge. *Zeitschrift für Technische Physik, 13,* 31.

Backus, J. (1969). *The acoustical foundations of music.* New York: Norton.

Backus, J., & Hundley, J. C. (1971). Harmonic generation in the trumpet. *Journal of the Acoustical Society of America, 49,* 509–519.

Bacry, A., Grossman, A., & Zak, J. (1975). Proof of the completeness of lattice states in the kq-representation. *Physical Review, B12,* 1118.

Balzano, G. J. (1986). What are musical pitch and timbre? *Music Perception, 3*(3), 297–314.

Barrière, J. B. (Ed.). (1991). *Le timbre: Une métaphore pour la composition.* Paris: C. Bourgois & IRCAM.

Barrière, J. B., Potard, Y., & Baisnée, P. R. (1985). Models of continuity between synthesis and processing for the elaboration and control of timbre structure. *Proceedings of the 1985 International Music Conference* (pp. 193–198). San Francisco: Computer Music Association.

Bastiaans, M. J. (1980). Gabor's expansion of a signal into Gaussian elementary signals. *Proceedings of the IEEE, 68,* 538–539.

Beauchamp, J. W. (1975). Analysis and synthesis of cornet tones using non linear interharmonic relationships. *Journal of the Audio Engineering Society, 23,* 778–795.

Benade, A. H. (1976). *Fundamentals of musical acoustics.* London and New York: Oxford University Press.

Bennett, G. (1981). Singing synthesis in electronic music. In *Research aspects of singing* (pp. 34–40). Stockholm: Royal Academy of Music, Publication 33.

Berger, J., Coifman, R., & Goldberg, M. J. (1994). A method of denoising and reconstructing audio signals. *Proceedings of the 1994 International Music Conference* (pp. 344–347). San Francisco: Computer Music Association.

Berger, K. W. (1964). Some factors in the recognition of timbre. *Journal of the Acoustical Society of America, 36,* 1888–1891.

Bismarck, G. von. (1974a). Sharpness as an attribute of the timbre of steady sounds *Acustica, 30,* 159–172.

Bismarck, G. von. (1974b). Timbre of steady sounds: a factorial investigation of its verbal attributes. *Acustica, 30,* 146–158.

Blackman, R. B., & Tukey, J. W. (1958). *The measurement of power spectra from the point of view of communications engineering.* New York: Dover.

Bolt, R. H., Cooper, F. S., David, E. E., Jr., Denes, P. B., Pickett, J. M., & Stevens, K. N. (1969). Identification of a speaker by speech spectrogram. *Science, 166,* 338–343.

Bolt, R. H., Cooper, F. S., David, E. E., Jr., Denes, P. B., Pickett, J. M., & Stevens, K. N. (1978). *On the theory and practice of voice identification.* Washington, DC: National Research Council.

Boomsliter, P. C., & Creel, W. (1970). Hearing with ears instead of instruments. *Journal of the Audio Engineering Society, 18,* 407–412.

Braus, I. (1995). Retracing one's steps: an overview of pith circularity and Shepard tones in European music:1550–1990. *Music Perception, 12*(3), 323–351.

Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound.* Cambridge, MA: MIT Press.

Bregman, A. S., & Campbell, J. (1971). Primary auditory stream segregation and perception of order in rapid sequences of tones. *Journal of Experimental Psychology, 89,* 244–249.

Bregman, A. S., & Pinker, S. (1978). Auditory streaming and the building of timbre. *Canadian Journal of Psychology, 32,* 19–31.

Brown, G. J., & Cooke, M. (1994). Computational auditory scene analysis. *Computers Speech Language (UK), 8*(4), 297–336.

Brown, G.J., & Cooke, M. (1994). Perceptual grouping of musical sounds: a computational model. *Journal of New Music Research (Netherlands), 23*(2), 107–132.

Cabot, R. C., Mino, M. G., Dorans, D. A., Tackel, I. S., & Breed, H. B. (1976). Detection of phase shifts in harmonically related tones. *Journal of the Audio Engineering Society, 24,* 568–571.

Cadoz, C., Luciani, A., & Florens, J. L. (1984). Responsive input devices and sound synthesis by simulation of instrumental mechanisms. *Computer Music Journal, 14*(2), 47–51.

Cadoz, C., Luciani, A., & Florens, J. L. (1993). CORDIS-ANIMA: a modeling and simulation system for sound and image synthesis—the general formalism. *Computer Music Journal, 17*(1), 19–29.

Campanella, S. J., & Robinson, G. S. (1971). A comparison of orthogonal transformations for digital speech processing. *IEEE Transactions on Communication Technology, COM-19,* 1045–1050.

Charbonneau, G. (1981). Timbre and the perceptual effect of three types of data reduction. *Computer Music Journal, 5* (2), 10–19.

Charles, D. (1980). L'eterno ritorno del timbro. In *Musica e Elaboratore* (pp. 94–99). Venice: Biennale.

Cheung, N. M., & Horner, A. (1996). Group synthesis with genetic algorithms. *Journal of the Audio Engineering Society, 44,* 130–147.

Chowning, J. (1971). The simulation of moving sound sources. *Journal of the Audio Engineering Society, 19,* 2–6.

Chowning, J. (1980). Computer synthesis of the singing voice. In *Sound generation in winds, strings, computers* (pp. 4–13). Stockholm: Royal Swedish Academy of Music.

Chowning, J. (1973). The synthesis of complex audio spectra by means of frequency modulation. *Journal of the Audio Engineering Society, 21,* 526–534.

Chowning, J. (1989). Frequency modulation synthesis of the singing voice. In M. V. Mathews & J. R. Pierce (Eds.), *Current directions in computer music research* (with a compact disk of sound examples) (pp. 57–63). Cambridge, MA: MIT Press.

Chowning, J., & Bristow, D. (1987). *FM theory and applications: By musicians for musicians.* Tokyo: Yamaha Foundation.

Clark, M., Robertson, P., & Luce, D. (1964). A preliminary experiment on the perceptual basis for musical instrument families. *Journal of the Audio Engineering Society, 12,* 194–203.

Cogan, R. (1984). *New images of musical sound.* Cambridge, MA: Harvard University Press.

Coifman, R, Meyer, Y., & Wickerhauser, V. (1992). Wavelet analysis and signal processing. In M. B. Ruskai, G. Beylkin, R. Coifman, I. Daubechies, S. Mallat, Y. Meyer, & L. Raphael (Eds.), *Wavelets and their applications* (pp. 153–178). Boston: Jones and Bartlett.

Colomes, C, Lever, M., Rault, J. B., Dehery, Y. F., & Faucon G. A.. (1995). Perceptual model applied to audio bit-rate reduction. *Journal of the Audio Engineering Society, 43,* 233–239.

Combes, J. M., Grossman, A., & Tchamitchian, P. (Eds.) (1989). *Wavelets.* Berlin: Springer-Verlag.

Cook, P. R. (1992). A meta-wind-instrument physical model controller, and a meta-controller for real-time performance control. *Proceedings of the 1992 International Music Conference* (pp. 273–276). San Francisco: Computer Music Association.

Cook, P. R. (1993). SPASM, a real-time vocal tract physical model controller; and Singer, the companion software synthesis system. *Computer Music Journal, 17*(1), 30–44.

Cook, P. R. (1996). Singing voice synthesis: history, current work, and future directions. *Computer Music Journal, 20*(3), 38–46.

Cosi, P., de Poli, G., & Lauzzana, G. (1994). Auditory modeling and self-organizing neural networks for timbre classification. *Journal of New Music Research (Netherlands), 23*(1), 71–98.

Culver, C. A. (1956). *Musical acoustics.* New York: McGraw Hill.

De Poli, G. (1983). A tutorial on sound synthesis techniques. *Computer Music Journal, 7*(4), 8–26.

Delprat, N., Guillemain, P., & Kronland-Martinet, R. (1990). Parameter estimation for non-linear resynthesis methods with the help of a time-frequency analysis of natural sounds. *Proceedings of the 1990 International Music Conference* (pp. 88–90). San Francisco: Computer Music Association.

Depalle, P., & Poirot, G. (1991). SVP: a modular system for analysis, processing, and synthesis of the sound signal. *Proceedings of the 1991 International Music Conference* (pp. 161–164). San Francisco: Computer Music Association.

Depalle, P., & Rodet, X. (1992). A new additive synthesis method using inverse Fourier transform and spectral envelopes. *Proceedings of the 1992 International Music Conference* (pp. 161–164). San Francisco: Computer Music Association.

Desain, P., & Honing, H. (1992). *Music, mind and machine.* Amsterdam: Thesis Publishers.

Deutsch, D. (1975). Musical illusions. *Scientific American, 233,* 92–104.

Deutsch, D. (1995). *Musical illusions and paradoxes* [CD]. La Jolla, CA: Philomel.

Dijksterhuis, P. R., & Verhey, T. (1969). An electronic piano. *Journal of the Audio Engineering Society, 17,* 266–271.

Dodge, C., & Terse, T. A. (1985). *Computer music: synthesis, composition and performance.* New York: Schirmer Books, McMillan.

Dolansky, L. O. (1960). Choice of base signals in speech signal analysis. *IRE Transactions on Audio, 8,* 221–229.

Dolson, M. (1986). The phase vocoder: A tutorial. *Computer Music Journal, 10*(4), 14–27.

Dowling, J., & Harwood, D. (1986). *Music cognition.* New York: Academic Press.

Duda, R. O., Lyon, R. F., & Slaney, M. (1990). Correlograms and the separation of sounds. *24th Asilomar Conference on Signals, Systems and Computers* (pp. 457–461), Pacific Grove, CA.

Eagleson, H. W., & Eagleson, O. W. (1947). Identification of musical instruments when heard directly and over a public address system. *Journal of the Acoustical Society of America, 19,* 338–342.

Emmerson, S. (Ed.) (1986). *The language of electroacoustic music.* London: McMillan Press.

Erickson, R. (1975). *Sound structure in music.* Berkeley, CA: University of California Press.

Ethington, R., & Punch, B. (1994). Seawave: a system for musical timbre description. *Computer Music Journal, 18*(1),30–39.

Fant, G. (1960). *Acoustic theory of speech production.* Gravenhage: Mouton.

Feynman, R. B., Leighton, R. B., & Sands, M. (1963). *The Feynman lectures on physics.* Reading, MA: Addison-Wesley.

Flanagan, J. L. (1962). Models for approximating basilar-membrane displacement. *Bell System Technical Journal, 41,* 959–1009.

Flanagan, J. L. (1972). *Speech analysis, synthesis and perception.* New York: Academic Press.

Fletcher, H., & Bassett, I. G. (1978). Some experiments with the bass drum. *Journal of the Acoustical Society of America, 64,* 1570–1576.

Fletcher, H., & Sanders, L. C. (1967). Quality of violin vibrato tones. *Journal of the Acoustical Society of America, 41,* 1534–1544.

Fletcher, H., Blackham, B. D., & Christensen, D. A. (1963). Quality of organ tones. *Journal of the Acoustical Society of America, 35,* 314–325.

Fletcher, H., Blackham, B. D., & Stratton, R. (1962). Quality of piano tones. *Journal of the Acoustical Society of America, 34,* 749–761.

Fletcher, N. H., & Rossing, T. D. (1991). *The physics of musical instruments.* New York: Springer-Verlag.

François, J. C. (1995). Fixed timbre, dynamic timbre. *Perspectives of New Music, 13*(1), 112–118.

Freed, A., Rodet, X., Depalle, P. (1993). Synthesis and control of hundreds of sinusoidal partials on a desktop computer without custom hardware. In *Proceedings of the 1993 International Computer Music Conference* (pp. 98–101) San Francisco: Computer Music Association.

Freed, D. J, & Martens, W. L. (1986). Deriving psychoacoustic relations for timbre. In *Proceedings of the 1986 International Music Conference* (pp. 393–405). San Francisco: Computer Music Association.

Freed, D. J. (1990). Auditory correlates of perceived mallet hardness for a set of recorded percussive sound events. *Journal of the Acoustical Society of America, 87,* 311–322.

Freedman, M. D. (1967). Analysis of musical instrument tones. *Journal of the Acoustical Society of America, 41,* 793–806.

Friberg, A. (1991). Generative rules for music performance: a formal description of a rule system. *Computer Music Journal, 15*(2), 56–71.

Friberg, A., Frydén, L., Bodin, L. G., & Sundberg, J. (1991). Performance rules for computer-controlled contemporary keyboard music. *Computer Music Journal, 15* (2), 49–55.

Gabor, D. (1947). Acoustical quanta and the nature of hearing. *Nature, 159*(4).

George, E. B., & Smith, M. J. T. (1992). Analysis-by-synthesis: Overlap-add sinusoidal modeling applied to the analysis and synthesis of musical tones. *Journal of the Audio Engineering Society, 40,* 497–516.

George, W. H. (1954). A sound reversal technique applied to the study of tone quality. *Acustica, 4,* 224–225.

Gibiat, V. (1988). Phase space representations of acoustic musical signals. *Journal of Sound and Vibration, 123,* 529–536.

Gibson, J. J. (1966). *The senses considered as perceptual systems.* Boston: Houghton Mifflin.

Goad, P., & Keefe, D. H. (1992). Timbre discrimination of musical instruments in a concert hall. *Music Perception, 10*(1), 43–62.

Gregory, A. H. (1994). Timbre and auditory streaming. *Music Perception, 12*(3), 161–174.

Grey, J. M. (1975). *An exploration of musical timbre.* Doctoral dissertation, Stanford University, Stanford, CA.

Grey, J. M. (1977). Multidimensional perceptual scaling of musical timbres. *Journal of the Acoustical Society of America, 61,* 1270–1277.

Grey, J. M. (1978). Timbre discrimination in musical patterns. *Journal of the Acoustical Society of America, 64,* 467–472.

Grey, J. M., & Gordon, J. W. (1978). Perceptual effect of spectral modifications in musical timbres. *Journal of the Acoustical Society of America, 63,* 1493–1500.

Grey, J. M., & Moorer, J. A. (1977). Perceptual evaluation of synthesized musical instrument tones. *Journal of the Acoustical Society of America, 62,* 454–462.

Grossman, A., & Morlet, J. (1984). Decomposition of Hardy functions into square integrable wavelets of constant shape. *SIAM Journal of Mathematical Analysis, 15,* 723–736.

Guillemain, P., & Kronland-Martinet, R. (1992). Additive resynthesis of sounds using continuous time-frequency analysis techniques. *Proceedings of the 1992 International Music Conference* (pp. 10–13). San Francisco: Computer Music Association.

Guillemain, P., & Kronland-Martinet, R. (1996). Characterization of acoustic signals through continuous linear time-frequency representations. *Proceedings of the IEEE, 84,* 561-585.

Hall, H. H. (1937). Sound analysis. *Journal of the Acoustical Society of America, 8,* 257–262.

Handel, S. (1990). *Listening: An introduction to the perception of auditory events.* Cambridge, MA: MIT Press.

Harmuth, H. (1972). *Transmission of information by orthogonal functions.* New York: Springer.

Helmholtz, H. von. (1954). *Sensations of tone* (E. J. Ellis, Trans.). New York: Dover. (Original work published 1877)

Hiller, L. & Ruiz, P. (1971). Synthesizing musical sounds by solving the wave equation for vibrating objects: Part I. *Journal of the Audio Engineering Society, 19,* 463–470.

Horner, A. (1996a). Double-modulator FM matching of instrument tones. *Journal of the Audio Engineering Society, 44,* 130–147.

Horner, A. (1996b). Computation and memory tradeoffs with multiple wavetable interpolation. *Journal of the Audio Engineering Society, 44,* 481–496.

Horner, A., & Beauchamp, J. (1996). Piecewise-linear approximation of additive synthesis envelopes: a comparison of various methods. *Computer Music Journal, 20*(2), 72–95.

Horner, A., Beauchamp, J. , & Haken, L. (1993). Methods for multiple wavetable synthesis of musical instrument tones. *Journal of the Audio Engineering Society, 41,* 336–356.

Hourdin, C. (1995). *Etude psychophysique du timbre: Application au codage et à la synthèse des sons en musique.* Dissertation, l'Université Paris XI, Orsay.

Huggins, W. H. (1952). A phase principle for complex frequency analysis and its implication in auditory theory. *Journal of the Acoustical Society of America, 24,* 582–589.

Huggins, W. H. (1957). *Representation and analysis of signals: Part I. The use of orthogonalized exponentials.* Baltimore, MD: Johns Hopkins University, Report AF 19 (604)-1941, ASTIA n) AD133741.

Huron, D., & Sellmer, P. (1992). Critical bands and the spelling of vertical sonorities. *Music Perception, 10,* 129–150.

Iverson, P., & Krumhansl, C. K. (1993). Isolating the dynamic attributes of timbre. *Journal of the Acoustical Society of America, 94,* 2595–2603.

Jaffe, D. A., & Smith, J. O. (1983). Extensions of the Karplus-Strong Plucked String Algorithm. *Computer Music Journal, 7*(2), 56–69.

Jansson, L., & Sundberg, J. (1975/1976). Long-term average spectra applied to analysis of music. *Acustica, 34,* 15–19, 269–274.

Jenkins, G. M ., & Watts, D. G. (1968). *Spectral analysis and its applications.* San Francisco: Holden-Day.

Justice, J. (1979). Analytic signal processing in music computation. *IEEE Transactions on Speech, Acoustics and Signal Processing, ASSP-27,* 670–684.

Karplus, K., & Strong, A. (1983). Digital synthesis of plucked string and drum timbres. *Computer Music Journal, 7*(2), 43–55.

Keefe, D. H. (1992). Physical modeling of wind instruments. *Computer Music Journal, 16*(4), 57–73.

Keeler, J. S. (1972). Piecewise-periodic analysis of almost-periodic sounds and musical transients. *IEEE Transactions on Audio & Electroacoustics, AU-1O,* 338–344.

Kendall, R., & Carterette, E. C. (1993). Identification and blend of timbres as basis for orchestration. *Contemporary Music Review, 9,* 51–67.

Kendall, R., & Carterette, E. C. (1993a). Verbal attributes of simultaneous instrument timbres: I von Bismarck adjectives. *Music Perception, 10*(4), 445–467.

Kendall, R., & Carterette, E. C. (1993b). Verbal attributes of simultaneous instrument timbres: II. Adjectives induced from Piston's orchestration. *Music Perception, 10*(4), 469–502.

Kleczkowski, P. (1989). Group additive synthesis. *Computer Music Journal, 13*(1), 12–20.

Koenig, W., Dunn, H. K., & Lacy, L. Y. (1946). The sound spectrograph. *Journal of the Acoustical Society of America, 18,* 1949.

Köhler, W. (1915). Akustische Untersuchungen. *Zeitschrift fur Psychologie, 72,* 159.

Kohut, J., & Mathews, M. V. (1971). Study of motion of a bowed string. *Journal of the Acoustical Society of America, 49,* 532–537.

Kramer, H. P., & Mathews, M. V. (1956). A linear coding for transmitting a set of correlated signals. *IRE Transactions on Information Theory, IT2,* 41–46.

Kronland-Martinet, R. (1989). The use of the wavelet transform for the analysis, synthesis and processing of speech and music sounds. *Computer Music Journal, 12*(4), 11–20 (with sound examples on disk).

Kronland-Martinet, R., & Guillemain, P. (1993). Towards non-linear resynthesis of instrumental sounds. *Proceedings of the 1993 International Music Conference* (pp. 86–93). San Francisco: Computer Music Association.

Kronland-Martinet, R., Morlet, J., & Grossman, A. (1987). Analysis of sound patterns through wavelet transforms. *International Journal of Pattern Recognition and Artificial Intelligence, 11*(2), 97–126.

Laroche, J. (1989). A new analysis/synthesis system of musical signals using Prony's method. Application to heavily damped percussive sounds. *Proc. IEEE ICASSP-89,* 2053–2056.

Laroche, J. (1993). The use of the matrix pencil method for the spectrum analysis of musical signals. *Journal of the Acoustical Society of America, 94,* 1958–1965.

Lashley, K. S. (1942). An examination of the "continuity theory" as applied to discriminative learning. *Journal of General Psychology, 26,* 241–265.

La synthèse sonore [Special issue]. (1993). *Cahiers de l'IRCAM, 2.*

Lauberhorn, W. (1996). Nonlinear dynamics in acoustics. *Acustica (Acta Acustica), 82,* S46–S55.

Laughlin, R. G., Truax, B. D., & Funt, B. V. (1990). Synthesis of acoustic timbres using principal component analysis. *Proceedings of the 1990 International Music Conference* (pp. 95–99). San Francisco: Computer Music Association.

Le Brun, M. (1979). Digital waveshaping synthesis. *Journal of the Audio Engineering Society, 27,* 250–266.

Lee, M., & Wessel, D. (1992). Connectionist models for real-time control of synthesis and compositional algorithms. In *Proceedings of the 1992 International Music Conference* (pp. 277–280). San Francisco: Computer Music Association.

Lee, M., & Wessel, D. (1993). Real-time neuro-fuzzy systems for adaptive control of musical processes. In *Proceedings of the 1993 International Music Conference* (pp. 172–175). San Francisco: Computer Music Association.

Lee, M., Freed, A., & Wessel, D. (1992). Neural networks for classification and parameter estimation in musical instrument control. In *Proceedings of the SPIE Conference on Robotics and Intelligent Systems,* Orlando, FL.

Lee, M., Freed, A., Wessel, D. (1991). Real-time neural network processing of gestural and acoustic signals. In *Proceedings of the 1991 International Music Conference.* San Francisco: Computer Music Association.

Legge, K. A., & Fletcher, N. H. (1989). Non-linearity, chaos, and the sound of shallow gongs. *Journal of the Acoustical Society of America, 86,* 2439–2443.

Leipp, E. (1971). *Acoustique et musique.* Paris: Masson.

Licklider, J. C. R. (1951). A duplex theory of pitch perception. *Experientia, Suisse, 7,* 128–133.

Loy, G. (1985). Musicians make a standard: the MIDI phenomenon. *Computer Music Journal, 9*(4), 8–26.

Luce, D., & Clark, M., Jr. (1967). Physical correlates of brass-instrument tones. *Journal of the Acoustical Society of America, 42,* 1232–1243.

Maganza, C., Caussé, R., & Laloë, F. (1986). Bifurcations, period doublings and chaos in clarinet-like systems. *Europhysics Letters, 1,* 295–302.

Martin, D. W. (1947). Decay rates of piano tones. *Journal of the Acoustical Society of America, 19,* 535.

Massie, D. C. (1998). Wavetable sampling synthesis. In M. Kahrs & K. Brandenberg (Eds.), *Applications of digital signal processing to audio and acoustics.* Norwell, MA: Kluwer Academic Publishers.

Mathes, R. C., & Miller, R. L. (1947). Phase effects in monaural perception. *Journal of the Acoustical Society of America, 19,* 780–797.

Mathews, M. V. (1963). The digital computer as a musical instrument. *Science, 142,* 553–557.

Mathews, M. V. (1969). *The technology of computer music.* Cambridge, MA: MIT Press.

Mathews, M. V., & Kohut, J. (1973). Electronic simulation of violin resonances. *Journal of the Acoustical Society of America, 53,* 1620–1626.

Mathews, M. V., & Moore, F. R. (1970). Groove—a program to compose, store and edit functions of time. *Communications of the ACM, 13,* 715 –7 21.

Mathews, M. V., Miller, J. B., & David, B. B., Jr., (1961). Pitch synchronous analysis of voiced sounds. *Journal of the Acoustical Society of America, 33,* 179–186.

Mathews, M. V., Moore, F. R., & Risset, J. C. (1974). Computers and future music. *Science, 183,* 263–268.

Mathews, M. V. , Miller, J. B., Pierce, J. R., & Tenney, J. (1965). Computer study of violin tones. *Journal of the Acoustical Society of America, 38,* 912 (abstract only).

Mathews, M. V. , Miller, J. B., Pierce, J. R., & Tenney, J. (1966). *Computer study of violin tones.* Murray Hill, NJ: Bell Laboratories.

McAdams, S. (1982). Spectral fusion and the creation of auditory images. In M. Clynes (Ed.), *Music, mind and brain* (pp. 279–298). New York: Plenum Press.

McAdams, S., & Bigand, E. (Eds.) (1993). *Thinking in sound: The cognitive psychology of human audition.* Oxford: Clarendon Press.

McAdams, S., & Bregman, A. (1979). Hearing musical streams. *Computer Music Journal, 3*(4), 26–43.

McAdams, S., & Saariaho, K. (1985). Qualities and functions of musical timbre. In *Proceedings of the 1985 International Music Conference* (pp. 367–374). San Francisco: Computer Music Association.

McAulay, R., & Quatieri, T. (1986). Speech analysis-synthesis based on a sinusoidal representation. *IEEE Transactions on Speech, Acoustics and Signal Processing, ASSP-34,* 744–754.

McConkey, J. (1984). Report from the synthesizer explosion. *Computer Music Journal, 8*(2), 59–60.

McIntyre, M. E., Schumacher, R. T., & Woodhouse, J. (1983). On the oscillations of musical instruments. *Journal of the Acoustical Society of America, 74,* 1325–1345.

Melara, R. D., & Marks, L. E. (1990). Interaction among auditory dimensions: timbre, pitch and loudness. *Perception & Psychophysics, 48,* 169–178.

Meyer, B., & Buchmann, G. (1931). *Die Klangspektren der Musikinstrumente.* Berlin.

Miller, D. C. (1926). *The science of musical sounds.* New York: MacMillan.

Miskiewicz, A. (1992). Timbre solfege: a course in technical listening for sound engineers. *Journal of the Audio Engineering Society, 40,* 621–625.

Modèles physiques, création musicale et ordinateur. (1994). Actes du Colloque international de Grenoble. Paris: Editions de la Maison des Sciences de l'Homme.

Moore, B. C. J. (1982). *An introduction to the psychology of hearing.* London, Academic Press.

Moore, B. C. J., & Glasberg, B. R. (1981). Auditory filter shapes derived in simultaneous and forward masking. *Journal of the Acoustical Society of America, 69,* 1003–1014.

Moore, F. R. (1990). *Elements of computer music.* Englewood Cliffs, NJ: Prentice Hall.

Moorer, J. A. (1977). Signal processing aspects of computer music: a survey. *Proceedings of the IEEE, 6S,* I108–I137.

Moorer, J. A. (1978). The use of the phase vocoder in computer music applications. *Journal of the Audio Engineering Society, 26,* 42–45.

Moorer, J. A., & Grey, J. (1977a). Lexicon of analyzed tones: Part I: A violin tone. *Computer Music Journal, 1*(2), 3945.

Moorer, J. A., & Grey, J. (1977b). Part II: Clarinet and oboe tones. *Computer Music Journal, 1*(3), 1229.

Moorer, J. A., & Grey, J. (1978). Part III: The trumpet. *Computer Music Journal, 2*(2), 23–31.

Morrill, D. (1977). Trumpet algorithms for music composition. *Computer Music Journal, 1*(1), 46–52.

Morrison, J. D., & Adrien, J. M. (1993). MOSAIC, a framework for modal synthesis: a modeling and simulation system for sound and image synthesis—the general formalism. *Computer Music Journal, 17*(1), 45–56.

Moulines, E., & Laroche, J. (1995). Non-parametric techniques for pitch-scale and time-scale modification. *Speech Communication, 16*(2), 175–215.

Music and Technology. (1971). UNESCO and *Revue Musicale.* Paris: Ed. Richard Masse.

Neisser, U. (1976). *Cognition and reality.* San Francisco: Freeman.

Olson, H. F. (1967). *Music, physics and engineering.* New York: Dover.

Panter, P. F. (1965). *Modulation, noise and spectral analysis.* New York: McGraw Hill.

Patterson, R. D. (1976). Auditory filter shapes derived with noise stimuli. *Journal of the Acoustical Society of America, 59,* 640–654.

Pierce, J. R. (1983). *The science of musical sound* (with sound examples on disk). San Francisco: Freeman/Scientific American.

Plomp, R. (1964). The ear as a frequency analyzer. *Journal of the Acoustical Society of America, 36,* 1628–1636.

Plomp, R. (1966). Timbre as a multidimensional attribute of complex tones. In R. Plomp & F. G. Smoorenburg (Eds.), *Frequency analysis and periodicity detection in hearing.* Leiden: Suithoff.

Plomp, R. (1976). *Aspects of tone sensation.* New York: Academic Press.

Plomp, R., & Steeneken, J. M. (1969). Effect of phase on the timbre of complex tones. *Journal of the Acoustical Society of America, 46,* 409–421.

Plomp, R., & Steeneken, J. M. (1971). Pitch versus timbre. *Proceedings of the 7th International Congress of Acoustics, Budapest, 3,* 377–380.

Radvansky, G. A., Fleming, K. J., & Simmons, J. A. (1995). Timbre reliance in nonmusicians' and musicians' memory for melodies. *Music Perception, 13*(2), 127–140.

Richardson, E. G. (1954). The transient tones of wind instruments. *Journal of the Acoustical Society of America, 26,* 960–962.

Risset, J. C. (1965). Computer study of trumpet tones. *Journal of the Acoustical Society of America, 33,* 912.

Risset, J. C. (1966). *Computer study of trumpet tones.* Murray Hill, NJ: Bell Laboratories.

Risset, J. C. (1969). *An introductory catalog of computer-synthesized sounds.* Murray Hill, NJ: Bell Laboratories. Reissued as part of *The historical CD of digital sound synthesis,* Mainz, Germany: Wergo, 1995.

Risset, J. C. (1971). Paradoxes de hauteur. *Proceedings of the 7th International Congress of Acoustics, Budapest, 3,* 613–616.

Risset, J. C. (1978a). Musical acoustics. In E. C. Carterette & M. P. Friedman, *Handbook of perception: Vol. IV. Hearing* (pp. 521–564). New York: Academic Press.

Risset, J. C. (1978 b). *Paradoxes de hauteur* (with a cassette of sound examples). Paris: IRCAM Report No. 10.

Risset, J. C. (1978c). *Hauteur et timbre* (with a cassette of sound examples). Paris: IRCAM Rep. No. 11.

Risset, J. C. (1986). Pitch and rhythm paradoxes. *Journal of the Acoustical Society of America, 80,* 961–962.

Risset, J. C. (1989). Paradoxical sounds. In M. V. Mathews & J. R. Pierce (Eds.), *Current directions in computer music research* (with a compact disk of sound examples) (pp. 149–158). Cambridge, MA: MIT Press.

Risset, J. C. (1991). Timbre analysis by synthesis: representations, imitations and variants for musical composition. In De Poli, G., Picciali, A., & Roads, C., eds. *The representation of musical signals* (pp. 7–43). Cambridge, MA: MIT Press

Risset, J. C. (1994). Quelques aspects du timbre dans la musique contemporaine. In A. Zenatti (Ed.), *Psychologie de la musique* (pp. 87–114). Paris: Presses Universitaires de France.

Risset, J. C., & Mathews, M. V. (1969). Analysis of musical instrument tones. *Physics Today, 22*(2), 23–30.

Roads, C. (with Strawn, J., Abbott, C., Gordon, J., & Greenspun, P.). (1996). *The computer music tutorial.* Cambridge, MA: MIT Press.

Roads, C. (1978). Automated granular synthesis of sound. *Computer Music Journal, 2*(2), 61–62.

Roads, C. (Ed.). (1985). *Composers and the computer.* Cambridge, MA: MIT Press.

Roads, C. (Ed.). (1989). *The music machine.* Cambridge, MA: MIT Press.

Rodet, X, Potard, Y, & Barrière, J. B. (1984). The Chant project: From the synthesis of the sung voice to synthesis in general. *Computer Music Journal, 8*(3), 15–31

Rodet, X. (1994). Stability/instability of periodic solutions and chaos in physical models of musical instruments. In *Proceedings of the 1992 International Computer Music Conference* (pp. 386–393). San Francisco: International Computer Music Association.

Rodet, X. (1979, July). Time-domain formant-wave-function synthesis. *Proceedings of the NATO-ASI Meeting.* Bonas.

Rodet, X., & Bennett, G. (1980). Synthèse de la voix chantée par ordinateur. *Conférences des Journees d'Etudes du Festival du Son* (pp. 73–91). Paris: Ed. Chiron.

Rodet, X., & Depalle, P., & Poirot, G. (1987). Analyse et synthèse de la voix parlée et chantée par modélisation de l'enveloppe spectrale et de l'excitation. *Actes des 16èmes Journées d'Etude sur la Parole* (pp. 41–44). Paris: Société Française d'Acoustique.

Rodet, X., Depalle, P., & Garcia, G. (1995). New possibilities in sound analysis-synthesis. In *Proceedings of the International Symposium on Musical Acoustics* (pp. 421–432). Paris: Société Française d'Acoustique.

Roederer, J. G. (1974). *Introduction to the physics and psychophysics of music.* London: The English Universities Press.

Rosenblatt, M. (Ed.). (1963). *Time series analysis.* New York: Wiley.

Rossing, T. D.(1990). *The science of sound.* Reading, MA: Addison–Wesley.

Rossing, T. D. (Ed.). (1984). *The acoustics of bells.* New York: van Nostrand Reinhold.

Rossing, T. D., Hampton, D. C., Richardson, B. E., Sathoff, J., & Lehr, A. (1988). Vibrational modes of Chinese two-tone bells. *Journal of the Acoustical Society of America, 83,* 369–373.

Saldanha, E. L., & Corso, J. F. (1964). Timbre cues and the identification of musical instruments. *Journal of the Acoustical Society of America, 36,* 2021–2026.

Sandell, G. (1995). Roles for spectral centroid and other factors in determining "blended" instrument pairings in orchestration. *Music Perception, 13*(2), 209–246.

Sandell, G., & Martens, W. (1995). Perceptual evaluation of principal components-based synthesis of musical timbres. *Journal of the Audio Engineering Society, 43,* 1013–1028.

Sasaki, L. H., & Smith, K. C. (1980). A simple data reduction scheme for additive synthesis. *Computer Music Journal, 4*(1), 22–24.

Schaeffer, P. (1966). *Traité des objets musicaux* (with three records of sound examples). Paris: Ed. du Seuil.

Schafer, R. W., & Rabiner, L. R. (1975). Digital representations of speech signals. *Proceedings of the IEEE, 63,* 662–677.

Scharf, B. (). Critical bands. In J. V. Tobias (Ed.), *Foundations of modern auditory theory* (Vol. 1, pp. 157–202). New York: Academic Press.

Schoenberg, A. (1911). *Harmonielehre.* Leipzig and Vienna: Universal Edition.

Schottstaedt, W. (1977). The simulation of natural instrument tones using frequency modulation with a complex modulating wave. *Computer Music Journal, 1*(4), 46–50.

Schroeder, M. R. (1966). Complementarity of sound buildup and decay. *Journal of the Acoustical Society of America, 40,* 549–551.

Schroeder, M. R. (1975). Models of hearing. *Proceedings of the IEEE, 63,* 1332–1350.

Schroeder, M. R., Atal, B. S., & Hall, J. L. (1979). Optimising digital speech coders by exploiting masking properties of the human ear. *Journal of the Acoustical Society of America, 66,* 1647–1652.

Serra, M.-H., Rubine, D., & Dannenberg, R. (1990). Analysis and synthesis of tones by spectral interpolation. *Journal of the Audio Engineering Society, 38,* 111–128.

Serra, X., & Smith, J. (1990). Spectral modeling synthesis: a sound analysis/synthesis system based on a deterministic plus stochastic decomposition. *Computer Music Journal, 14*(4), 12–24.

Shepard, R. N. (1964). Circularity of relative pitch. *Journal of the Acoustical Society of America, 36,* 2346–2353.

Slaney, M. (1996). Correlograms and their inversion - the importance of periodicities in auditory perception. *Acustica, 82,* S91.

Slaney, M., Covell, M., & Lassitier, B. (1996). Automatic audio morphing. *Proceedings of 1996 ICASSP,* Atlanta.

Slaney, M., & Lyon, R. F. (1993). On the importance of time: A temporal representation of sound. In Cooke, M., Beet, S., & Crawford, J. (Eds.), *Visual representations of speech signals.* Sussex, England: J. Wiley.

Slaney, M., Lyon, R., & Naar, D. (1994). Auditory model inversion for sound separation. *Proceedings of 1994 ICASSP, Adelaide, Australia, 2,* 77–80.

Slawson, A. W. (1968). Vowel quality and musical timbre as functions of spectrum envelope and fundamental frequency. *Journal of the Acoustical Society of America, 43,* 97–101.

Slawson, W. (1985). *Sound color.* Berkeley: University of California Press.

Sloboda, J. (Ed.). (1988). *Generative processes in music performance.* Oxford: Oxford Science Publications.

Smalley, D. (1986). Spectromorphology and structuring processes. In S. Emmerson (Ed.), *The language of electroacoustic music* (pp. 61–93). New York: Harwood Academic Publishers.

Smalley, D. (1993). Defining transformations. *Interface, Netherlands, 22* , 279–300.

Smith, J. (1996). Physical modeling synthesis update. *Computer Music Journal, 20*(2), 44–56.

Smith, J. (1992). Physical modeling using digital waveguides. *Computer Music Journal, 16*(4), 74–91.

Smith, J., & Van Duyne, S. (1995). Recent results in piano synthesis via physical modeling. *Proceedings of the International Symposium on Musical Acoustics* (pp. 503–509). Paris: Société Française d'Acoustique.

Stapleton, J., & Bass, S. (1988). Synthesis of musical tones based on the Karhunen-Loève transform. *IEEE Transactions on Speech, Acoustics and Signal Processing, ASSP-36,* 305–319.

Strawn, J. (1987). Analysis and synthesis of musical transitions using the discrete short-time Fourier transform. *Journal of the Audio Engineering Society, 35,* 3–14.

Strong, W., & Clark, M., Jr., (1967a). Synthesis of wind-instrument tones. *Journal of the Acoustical Society of America, 41,* 39–52.

Strong, W., & Clark, M. Jr., (1967b). Perturbations of synthetic orchestral wind instrument tones. *Journal of the Acoustical Society of America, 41,* 277–285.

Stumpf, C. (1926). *Die sprachlaute.* Berlin and New York: Springer-Verlag.

Sundberg, J. (1977). The acoustics of the singing voice. *Scientific American, 236,* 82–91.

Sundberg, J. (Ed.) (1983). *Studies of music performance.* Stockholm: Royal Academy of Music, Publication 39.

Sundberg, J. (1987). *The science of the singing voice.* DeKalb, IL: Northern Illinois University Press.

Sundberg, J., & Frydén, L. (1989). Rules for automated performance of ensemble music. *Contemporary Music Review, 3,* 89–109.

Tan, B. T. G., Gan, S. L., Lim, S. M., & Tang, S. H. (1994). Real-time implementation of double frequency modulation (DFM) synthesis. *Journal of the Audio Engineering Society, 42,* 918–926.

Tellman, E., Haken, L., & Holloway, B. (1995). Timbre morphing of sounds with unequal numbers of features. *Journal of the Audio Engineering Society, 43,* 678–688.

Tenney, J. C. (1965). The physical correlates of timbre. *Gravesaner Blätter, 26,* 103–109.

Terhardt, E. (1974). Pitch, consonance, and harmony. *Journal of the Acoustical Society of America, 55,* 1061–1069.

Terhardt, E. (1978). Psychoacoustic evaluation of musical sounds. *Perception & Psychophysics, 23,* 483–492.

Toiviainen, M., Kaipainen, M., & Louhivuoiri, J. (1995). Musical timbre: similarity ratings correlate with computational feature space distances. *Journal of New Music Research (Netherlands), 24,* 282–298.

Valimaki, V., & Takala, T. (1996). Virtual musical instruments - natural sound using physical models. *Organised Sound (England), 1,* 75–86.

Van Duyne, S., & Smith, J. (1993). Physical-modeling with a 2-D digital waveguide mesh. In *Proceedings of the 1993 International Computer Music Conference* (pp. 30–37). San Francisco: International Computer Music Association.

van Noorden L. (1975). *Temporal coherence in the perception of tone sequences.* Eindhoven, Holland: Instituut voor Perceptie Onderzoek.

Varèse, E. (1983). *Ecrits* (L. Hirbour-Paquette, Ed.). Paris: C. Bourgois.

Verge, M. P., Caussé, R., & Hirschberg, A. (1995). A physical model of recorder-like instruments. *Proceedings of the 1995 International Computer Music Conference* (pp. 37–44). San Francisco: International Computer Music Association.

Wang, K., & Shamma, S. A. (1994). Self-normalization and noise-robustness in early auditory representations. *IEEE Transactions on Speech & Audio Processing, 2,* 421–435.

Wang, K., & Shamma, S. A. (1995). Spectral shape analysis in the central auditory system. *IEEE Transactions on Speech & Audio Processing, 3,* 382–395.

Warren, R. M. (1982). *Auditory perception.* Elmsford, NY: Pergamon Press.

Warren, W. H. Jr., & Verbrugge, R. R. (1984). Auditory perception of breaking and bouncing events: a case in ecological perception. *Journal of Experimental Psychology: Human Perception and Performance, 10,* 704–712.

Wedin, L., & Goude, G. (1972). Dimension analysis and the perception of instrumental timbre. *Scandinavian Journal of Psychology, 13*(3), 228–240.

Weinreich, G. (1977). Coupled piano strings. *Journal of the Acoustical Society of America, 62,* 1474–1484.

Weinreich, G. (1979). The coupled motions of piano strings. *Scientific American, 240*(1), 118–127.

Wente, E. C. (1935). Characteristics of sound transmission in rooms. *Journal of the Acoustical Society of America, 7,* 123.

Wessel D. L. (1978). *Low dimensional control of musical timbre.* Paris: IRCAM Report, No. 12 (with a cassette of sound examples).

Wessel, D. L. (1973). Psychoacoustics and music: A report from Michigan State University. *Bulletin of the Computer Arts Society, 30.*

Wessel, D. L. (1979). Timbre space as a musical control structure. *Computer Music Journal, 3*(2), 45–52. Reprinted in Roads, C., & Strawn, J. (Eds.), (1985). *Foundations of computer music.* Cambridge, MA: MIT Press.

Wessel, D. L., & Risset, J. C. (1979). *Les illusions auditives.* Universalia: Encyclopedia Universalis, 167–171.

Wessel, D. L. (1991). Instruments that learn, refined controllers, and source model loudspeakers. *Computer Music Journal, 15*(4), 82–86.

Wessel, D. L., Bristow, D., & Settel, Z. (1987). Control of phrasing and articulation in synthesis. *Proceedings of the 1987 International Computer Music Conference* (pp. 108–116). San Francisco: International Computer Music Association.

Winckel, F. (1967). *Music, sound and sensation.* New York: Dover.

Xenakis, I. (1971). *Formalized music.* Bloomington, IN: Indiana University Press.

Yang, X., Wang, K., & Shamma, S. (1992). Auditory representation of acoustic signals. *I.E.E.E. Trans. on Information Theory, 38,* 824–839.

Young, R. W. (1952). Modes, nodes and antinodes. *Journal of the Acoustical Society of America, 24,* 267–273.

Zahorian, S., & Rothenberg, M. (1981). Principal-component analysis for low-redundancy encoding of speech spectra. *Journal of the Acoustical Society of America, 69,* 823–845.

Zwicker, E. (1961). Subdivision of the audible frequency range into critical bands. *Journal of the Acoustical Society of America, 33,* 248.

Zwicker, E., & Scharf, B. (1965). A model of loudness summation. *Psychological Review, 72,* 3–26.