Received 23 May 1966 6.9

Synthesis of Wind-Instrument Tones

WILLIAM STRONG

Physics Department, Brigham Young University, Provo, Utah 84601

MELVILLE CLARK

Melville Clark Associates, Cochituate, Massachusetts 01763

Clarinet, oboe, bassoon, tuba, flute, trumpet, trombone, French horn, and English horn tones have been synthesized with partials controlled by one spectral envelope (fixed for each instrument regardless of note frequency) and three temporal envelopes. Musically literate auditors identified natural tones with 85% accuracy and our synthesized tones with 66% accuracy; a number of the confusions were intrafamily. With intrafamily confusions tolerated in the scoring, the auditors identified natural tones with 94% accuracy and our synthetic ones with 77% accuracy.

INTRODUCTION

SYNTHESIS and analysis are complementary techniques for finding an aurally complete, nonredundant, and easily implemented representation of the tones produced by musical instruments. The aural identification of a musical instrument from its tones is independent, or nearly independent, of many variables of performance: the particular example used of a given instrument, the intensity with which the tone is produced, the note played, the acoustical environment, and so on. To the perceptual invariant of identification, there must correspond one or more physical invariants that characterize the sounds produced by an instrument that serve to identify and distinguish it aurally from other instruments. The aural significance of the physical invariants found in analysis can be assessed by the accuracy of identification of synthetic tones with systematic perturbations of the invariants. Because of the enormous number and range of the variables required, synthesis, to be fruitful, must be guided by analysis. Before examining the effects of perturbations, it is necessary to get reasonably close auditory approximations to the tones produced by musical instruments, and this is the subject of this paper for wind instruments. The matter of perturbations will be treated in a subsequent paper.

Ability to characterize the tones of familiar instruments in a simple and aurally satisfactory manner enables the invention of more controllable and flexible means for producing the tones of these instruments and may enable the creation of new types of musical instruments having greater artistic capabilities than those presently available. Such an ability will probably bring some systematization and rationality to the creation and physical characterization of new sounds artistically useful in music—or will, at least, open avenues to such a systematization.

The use of the tones of conventional musical instruments, rather than arbitrarily generated signals, is advantageous, since such tones are easily generated, can be readily reproduced at will, and provide a standard, reasonably well-quantized and well-established set of timbres known to be musically useful, familiar to, identifiable by, and capable of being uniquely labeled by, a large body of musically literate auditors. Thus, we are provided with a means for testing the accuracy of any particular synthesis; the articulation score of musically literate subjects in identifying the instruments thought to produce the tones, or in distinguishing the artificial tones from those produced by natural means, provides a good measure of the accuracy of synthesis or the sensitivity of a parameter to perturbation.

There has been some analysis of the steady states, rather less of transients of instrumental tones, and a little synthesis. Fletcher[1,2] and his associates at Brigham

[1] H. Fletcher, E. D. Blackham, and R. Stratton, "Quality of Piano Tones," J. Acoust. Soc. Am. 34, 749–761 (1962).
[2] H. Fletcher, E. D. Backham, and D. A. Christensen, "Quality of Organ Tones," J. Acoust. Soc. Am. 35, 314–325 (1963).

Young University have synthesized organ and piano tones. As regards piano tones, the choral effects arising from the presence of several strings at each note had to be reflected in the synthesis.[1] Since we are concerned only with the tones produced by single structures, no attempt is made here at examining the question of choral tones—which is itself regarded as a separate, massive subject. Fletcher *et al.* found the partials of organ tones to be harmonic for solo tones (excluding mixtures).[2] Risset, at Bell Laboratories, and his coworkers have synthesized trumpet tones with due regard to economy of specification and objectivity, and have found no evidence for inharmonic partials.[3] Their system was such that they limited their synthesis to partials below 4000 cps, a fact that leads to a rather muffled, nonbrilliant character for their sounds. It appears that only the signals of high quality were presented to auditors. Freedman, at the University of Illinois, has described an analysis–synthesis scheme and has synthesized one tone for each of the following instruments: clarinet, trumpet, saxophone, bassoon, and violin.[4] Freedman started from the wave equation and heuristically generalized his resulting representation, essentially to account for the properties of the sound source that are largely unknown. He then used time-limited Fourier transforms[4,5] to compute various parameters, at the expense of very great amounts of computer time. He did find inharmonic partials for the saxophone.

The Fourier analyses[6] of Luce and Clark on the temporal dependences of the partials of 14 nonpercussive orchestral instruments is the basis of the present work. These authors found that, within a given tone, the durations of the attack transients for different partials will differ. Also, there may be some "Zitterbewegung" in the tone of an instrument, in that the fundamental frequency may fluctuate in a fast and quasirandom manner. Waveform modulation may be quite pronounced during the steady state, especially in flute tones. The less intense partials may display short-term amplitude variations during the steady state, while the more intense partials are of nearly constant amplitude from one cycle to the next. The spectral envelope determines the relative amplitudes of the harmonics with considerable accuracy for the steady state.[6]

Work other than that in analysis also served to guide our synthesis. The attack transient appears to be of preponderant importance for identifying the tones of nonpercussive musical instruments, and the decay transient appears to be of little importance to this identification.[7–11] Again, in speech, consonants, which are of much shorter duration than vowels, contribute more significantly to the intelligibility.[12] The harmonic content of a tone becomes richer with increasing intensity; yet, the timbre is only weakly intensity-dependent.[13] Complex instrumental tones can be identified from only a few partials.[14]

I. MATHEMATICAL REPRESENTATION

Ideally, the mathematical representation of the pressure wave produced by a musical instrument should be complete enough so that pressure waves created from the representation and produced by the musical instrument are indistinguishable via the auditory process. Distinguishability by measuring methods other than the auditory process is regarded as irrelevant for our present purposes. Second, the representation should contain only elements having auditory significance in characterizing the tone. Though not necessary, it is highly desirable that the representation be readily generated in the laboratory to enhance its utility—with the principle of parsimony as a collorary. The selection of such a representation requires knowledge and insight far beyond those presently available.

However, certain facts served to narrow our choice of representations. It was observed that the signals created by an important class of musical instruments have a middle section (pseudosteady state) during which the signal is approximately cyclically repetitive, although the waveforms are, in general, very complex. Instruments producing such signals are called nonpercussive. Although a frequency is quite precisely defined for these signals, the use of a Fourier series of trigonometric functions with constant coefficients is barred by the known significance of the attack transient, during which the amplitude of the wave and usually its shape are changing markedly. A Fourier-integral representation, though complete, is obviously aurally redundant as regards characteristics that distinguish

[3] J. C. Risset, "Computer Study of Trumpet Tones," J. Acoust. Soc. Am. **38**, 912(A) (1965).

[4] M. D. Freedman, "A Technique for the Analysis of Musical Instrument Tones," Tech. Rept. No. 6 (Elec. Eng. Res. Lab., Univ. Ill., Urbana, 1965).

[5] M. Clark, "Some Problems in Musical Acoustics," J. Audio Eng. Soc. **7**, 2–4 (1959).

[6] D. Luce and M. Clark (to be published). [Meanwhile see: D. A. Luce, "Physical Correlates of Nonpercussive Musical Instruments Tones," Ph.D. Thesis, MIT (Feb. 1963)].

[7] M. Clark, D. Luce, R. Abrams, H. Schlossberg, and J. Rome, "Preliminary Experiments on the Aural Significance of Parts of Tones of Orchestral Instruments and on Choral Tones," J. Audio Eng. Soc. **11**, No. 1, 45–54 (1963).

[8] K. W. Berger, "Some Factors in the Recognition of Timbre," J. Acoust. Soc. Am. **36**, 1888–1891 (1964).

[9] E. L. Saldanha and J. F. Corso, "Timbre Cues and the Identification of Musical Instruments," J. Acoust. Soc. Am. **36**, 2021–2026 (1964).

[10] H. V. Eagleson and O. W. Eagelson, "Identification of Musical Instruments When Heard Directly and over a Public Address System," J. Acoust. Soc. Am. **19**, 338–342 (1947).

[11] W. H. George, "A Sound Reversal Technique Applied to the Study of Tone Quality," Acustica **4**, 224–225 (1954).

[12] J. C. R. Licklider and G. A. Miller, *Handbook of Experimental Psychology*, S. S. Stevens, Ed. (John Wiley & Sons, Inc., New York, 1951).

[13] M. Clark and P. Milner, "Dependence of Timbre on the Tonal Loudness Produced by Musical Instruments," J. Audio Eng. Soc. **12**, 28–31 (1964).

[14] M. Clark, private communication. See also D. Johnson, "The Importance of Natural Harmonics in Identifying Musical Instruments," thesis, MIT (1960).

the signal produced by one instrument from those produced by others of the same frequency, intensity, and duration. For example, the Fourier-integral spectra contain information concerning the duration of the tone.

The sensitivity of the auditory process to both amplitude and frequency depends on the duration of the signal.[15] A superposition of some kind of cyclically repetitive basis functions with time-dependent coefficients suggests itself. The trigonometric functions seem to be the simplest from a calculational point of view; and they have the desirable properties of completeness, orthogonality, and simple normalization. It is noted that the attack transient in most cases lasts for quite a few cycles of the wave. Accordingly, we might think of using a modified Fourier series as a representation in which the coefficients change slowly with time as compared with the basis functions themselves. The coefficients might then be computed for each cycle of the wave, as has in fact been done by Luce and Clark.[6]

Synthesis from a scheme in which the Fourier coefficients are so calculated leads to discontinuities between successive cycles that are particularly pronounced when the amplitude changes from cycle to cycle are large. The amplitude changes are particularly great during the aurally important attack transient, and the resulting noise is objectionable. While the discontinuity difficulty could have been removed by several methods, the cost in machine time when this work was being performed (1962) would have been prohibitive. There are other objections to such procedures: They involve a specification of so much information that no useful purpose would have been served. We specifically seek to characterize many different tones of a musical instrument by as few parameters and principles as possible; otherwise, one has neither science nor utility. The principles and parameters must be tested against a number of notes of different frequencies for we have found that a specification that works in one range of an instrument does not always work so well in another régime. For example, it is found that the ratios of the amplitudes of the various partials during the pseudosteady state change with note frequency but that the spectral envelope of these partials is much more nearly constant.[6] Therefore, spectral-envelope control of the amplitudes of the partials was used. Again, the statistics associated with the production of a tone on a particular instrument—the intricacies of the radiation pattern both in direction and frequency—require analysis and complementary synthesis of many notes of each instrument.

Parsimony leads to the hypothesis, which must be experimentally checked, that an auditor is not really aware of all the detail involved in the complex-looking tones of musical instruments and, therefore, that some of the detail may be omitted without impairing the auditory completeness of the representation. We assume that the auditory process perceives only certain gross features of the signal, and it is one of our purposes to determine the least amount of information needed to characterize satisfactorily the signals aurally.

The temporal courses of each partial might be represented by a linear superposition of functions: the parameters specifying the superposition and the functions could then be used to characterize the temporal behavior of the partial in question. However, the analyses of Luce and Clark revealed such a variety of behavior that such a procedure appeared unnecessarily restrictive and cumbersome—and, at least, premature when their work was performed (1961).

Luce and Clark have found that the amplitudes of the partials, especially the more intense ones, of many instruments are nearly constant from cycle to cycle during the steady state and that the amplitudes of the partials of a particular instrument may be represented by a spectral envelope that is independent of the fundamental frequency.[6] Specifically they found that, for a particular instrument, if the amplitudes of the partials of a given frequency of all notes during the steady state are normalized to a common value, then the amplitudes of all other partials of all notes will follow a common function of frequency—i.e., will fall on one amplitude-versus-frequency curve. This curve is called the spectral envelope. Further, they found that this curve is nearly independent of the dynamic marking at which an instrument is sounded. In the case of the clarinet, two spectral envelopes must be used: one for even partials and one for odd.

The representation finally used for the sound-pressure wave was

$$f(t)=f_1(t)+f_2(t)+f_3(t),$$

where

$$f_1(t)=N(t)\sum_{j=1}^{30}\delta_j a_j \sin(j\omega t+\phi_1),$$

$$f_2(t)=C_i(t)\sum_{m=1}^{30}\delta_m a_m \sin(m\omega t+\phi_m),$$

$$f_3(t)=C_k(t)\sum_{n=1}^{30}\delta_n a_n \sin(n\omega t+\phi_n),$$

where i, $k=1$, 2, 3, or 4; ϕ_j, ϕ_m, and ϕ_n are arbitrarily chosen; a_j, a_m, and a_n are determined from the spectral envelope; $N(t)$ is the natural temporal envelope of amplitudes, and $C_i(t)$ and $C_k(t)$ are temporal envelopes constructed as described below; δ_j is nonzero for those partials in Group 1 (in which case it is 1); and ω is the fundamental angular frequency of the tone. Stated in words: The 30 partials used can be separated into three groups, in each of which the temporal dependence of the amplitude can be independently specified. The fre-

[15] L. Chih-an and L. A. Chistovich, "Frequency Difference Limens as a Function of Tonal Duration," Soviet Phys.—Acoust. 6, 75–80 (1960).

quency range between 0 and 8000 cps can be subdivided into three or fewer subregions, and the partials falling within any one subregion are placed in the corresponding group. Alternatively, the amplitude range of 50 dB can be divided into three or fewer subregions, and the partials within any one are placed in the corresponding group. Partials above 8000 cps are more than 40 dB down from the most intense ones for the wind instruments and are more or less masked, in addition to being discriminated against by the auditory process. The most intense partials usually fall into the group having the natural temporal envelope.

The natural envelope $N(t)$ was determined by the computer from the successive maxima occurring in each fundamental period. Values of the envelope between successive maxima were determined by linear interpolation.

The temporal envelopes for partials in Groups 2 and 3 were constructed from one or two of the following functions:

Function 1:

$$C_1(t)=[N(t)/N(\max)]^{n/4}N(\max),$$

where $N(\max)$ is the maximum value of $N(t)$, $n=1, \cdots, 20$;

Function 2:

$$C_2(t)=\frac{(t/T)^n N(t)}{1-(t/T)^n},\quad 0\leqslant (t/T)^n\leqslant \tfrac{1}{2},$$

$$C_2(t)=N(t),\qquad \tfrac{1}{2}\leqslant (t/T)^n,$$

where T is related to the rise time of the function, $n=1, \cdots, 10$;

Function 3:

$$C_3(t)=[A_1+(A_2-A_1)(t/T_1)]X(t),\qquad 0\leqslant t\leqslant T_1,$$

$$C_3(t)=\left[A_2+(A_3-A_2)\left(\frac{t-T_1}{T_2-T_1}\right)\right]X(t),\qquad T_1\leqslant t\leqslant T_2,$$

$$C_3(t)=\left[A_3+(A_4-A_3)\left(\frac{t-T_2}{T_3-T_2}\right)\right]X(t),\qquad T_2\leqslant t\leqslant T_3,$$

$$C_3(t)=\left\{A_4+\left(\frac{N(T_4)}{N(\max)}-A_4\right)\times\left(\frac{t-T_3}{T_4-T_3}\right)\right\}X(t),\quad T_3\leqslant t\leqslant T_4,$$

$$C_3(t)=N(t)X(t)/N(\max),\qquad T_4\leqslant t,$$

where the parameters A_1, A_2, A_3, and A_4 are dimensionless and lie between 0 and 1, and T_1, T_2, T_3, and T_4 are temporal parameters;

$$X(t)=R(\max)\frac{1}{\tau}\int_{t-\tau}^{t}dt'\left[\frac{K+R(t)}{K+\frac{1}{2}R(\max)}\right],$$

K is a constant, $R(t)$ is a random number between 0 and $R(\max)$, and τ is a small smoothing time.

Function 4:

$$C_4(t)=a(t/T_1)Y,\qquad 0\leqslant t\leqslant T_1,$$

$$C_4(t)=\left[a+(b-a)\left(\frac{t-T_1}{T_2-T_1}\right)\right]Y,\qquad T_1\leqslant t\leqslant T_2,$$

$$C_4(t)=\left[b+(c-b)\left(\frac{t-T_2}{T_3-T_2}\right)\right]Y,\qquad T_2\leqslant t\leqslant T_3,$$

$$C_4(t)=\left[c+(d-c)\left(\frac{t-T_3}{T_4-T_3}\right)\right]Y,\qquad T_3\leqslant t\leqslant T_4,$$

$$C_4(t)=\left[d+(c-d)\left(\frac{t-T_4}{T_4-T_3}\right)\right]Y,\qquad T_4\leqslant t\leqslant T_5,$$

$$C_4(t)=\left[e+\left(\frac{fN(T_6)}{N(\max)}-e\right)\left(\frac{t-T_5}{T_6-T_5}\right)\right]Y,\quad T_5\leqslant t\leqslant T_6,$$

$$C_4(t)=N(t),\qquad t\leqslant T_6,$$

where $Y=N(\max)/f$, T_1 through T_6 are suitably selected temporal parameters, and a through f are suitably selected, dimensionless parameters.

The natural temporal envelope makes it possible to suit the length of the attack transient to the instrument simulated. $C_1(t)$ permits us to modulate the waveform throughout the tone, as is necessary for the flute. $C_2(t)$ permits different rates of rise to be selected for various partials during the attack transient. $C_3(t)$ allows us to incorporate short-term variations during the steady state, as is needed for some of the weaker partials and to cause high-frequency partials to increase abruptly (as is displayed by an analysis of string tones). Finally, $C_4(t)$ permits the introduction of a blip (a sudden modulation in amplitude and waveform) during the attack transient, as brasses often require.

The use of one temporal envelope was found to be insufficient, even though that envelope was provided by a natural tone. However, the tones with 30 partials do sound more natural than those produced by the Fourier-series method with only 11 partials.

There is some indication in the literature that the auditory process is somewhat sensitive to the static phases of a complex wave.[16] However, our auditors were unable to distinguish between a synthetic trumpet tone of 277 cps with phases chosen to give three different

[16] R. C. Mathes and R. L. Miller, "Phase Effects in Monaural Perception," J. Acoust. Soc. Am. 19, 780–797 (1947).

types of waves: (1) repetitive pulses, (2) the derivative of repetitive pulses, (3) a shape having approximately uniform distribution of energy throughout the fundamental period. Thus, static phases were arbitrarily chosen for the synthesis.

In view of the paucity of evidence for inharmonic partials during the pseudosteady state, no attempt was made to include such partials in the present synthesis. In the case in which Luce and Clark found inharmonic partials for the oboe, it was determined by informal experimentation that when an inharmonic partial is sufficiently strong to be heard, the naturalness of the oboe tone is impaired. Inharmonic partials during the steady state that are closely spaced around one or more of the harmonics (except those distributions resulting from or related to amplitude or frequency modulations) are regarded as the subject of choral tones[5,17]—a large topic in itself, not discussed in the present work. In other informal experiments to augment the lifelike characteristics of our tones, it was found that, in the particular manner used, initial frequency modulations and random amplitude modulations, when sufficiently intense to be heard, did not enhance the naturalness of the tones created. Consequently, our syntheses were concentrated on other aspects of tonal simulation. Nevertheless, we believe that Zitterbewegung, when properly created, will prove important, particularly to the string tones.

If the reader so chooses, he may regard the attack transient, including the blips for the brasses, as the result of the superposition of a large number of evanescent, closely spaced, inharmonic partials. For reasons of convenience, we do not choose such a representation.

II. CHARACTERIZATION OF THE SYNTHETIC TONES

In the present research, each synthesized tone is characterized by its spectral envelope, its natural and constructed temporal envelopes, and the division of the partials into groups. These quantities are listed in Table I and displayed in Figs. 1–30 for each of the nine instruments examined. In all cases, the normalization is arbitrary. Note that the first 100 msec of the amplitude plots are expanded to show greater detail in view of the great psychoacoustic significance of this segment of our 700-msec tones. Only tones having a fundamental frequency of less than 1000 cps are studied; the most often used frequency range is thus covered.

The smoothed spectral envelopes shown in Figs. 1–10 are approximations to those found by Luce and Clark.[6] The accuracy of the approximations to their results can be judged from the data in Fig. 1, where both their original results and the present smoothed curve are given. Luce and Clark determined the spectral envelope at every 25-cps interval.

Several points may be noted concerning the *brass* instruments.

[17] M. Clark, "Proposed Keyboard Musical Instrument," J. Acoust. Soc. Am. 31, 403–419 (1959).

- A blip appears during the attack transient of the lower tones on any brass instrument. This blip gradually vanishes with increasing frequency of the note played on any given instrument. Additionally, the blip tends to be somewhat larger for the low-pitched instruments than for the higher-pitched ones. The French horn has several blips during the attack transient for its lower tones. A blip was found to contribute to, and to be necessary for, the "bite" of the tones of brass instruments.
- The durations of the attack transients for the trombone are larger than those for the trumpet, and those for the tuba are longer yet.
- While the spectral envelopes for the trumpet, trombone, and tuba scale with the intrinsic frequency of the instrument, observe that this envelope decreases much more sharply for the French horn.
- The most intense partials were classified together into Group 1; the next most intense partials were put into Group 2; and the weakest ones, which display quasirandom amplitudes with time, were put into Group 3.
- All partials for the lowest note of the tuba were put into Group 1; the partials of other notes were distributed over the three groups as described for the other brasses.

The following points may be noted concerning the *woodwinds:*

- The analysis of Luce and Clark for the flute had to be extended to 0.75 sec and well beyond that (100 cps) for all other instruments in order to determine the details of the intense waveform and amplitude modulation.
- The note at 266 cps has more amplitude modulation than is normally found with the flute.
- The partials in Group 2 for the flute are amplitude-modulated more heavily than those of Group 1.
- The slow rise of the partials during the attack transient is typical of the flute and contrasts strongly with the behavior of the partials for the double reeds and the brasses.
- The clarinet displays strong odd partials and weak even partials for those below 3000 cps, so two spectral envelopes are calculated. Furthermore, it was found desirable to use separate curves for notes below 414 cps and for those above 440 cps. The rate of decrease of the two spectral envelopes above 3500 cps is about the same.
- For the clarinet, the partials in Group 2 initially increase more slowly than those in Group 1.
- A predominant feature of the spectral envelope for the oboe, especially, and to a lesser degree for the

(*Text continued on p. 49*)

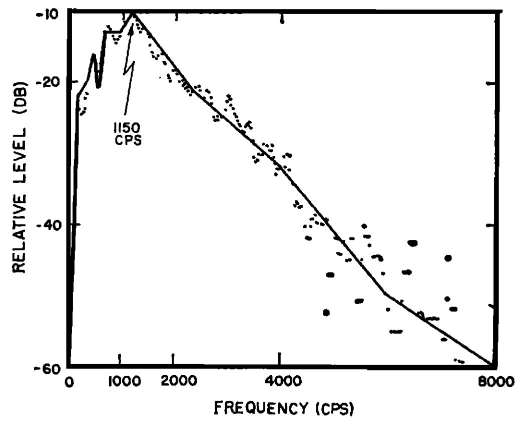

FIG. 1. Smoothed spectral envelope for trumpet. Smoothed points show actual spectral envelope used. Dotted points are from Luce and Clark.[6]

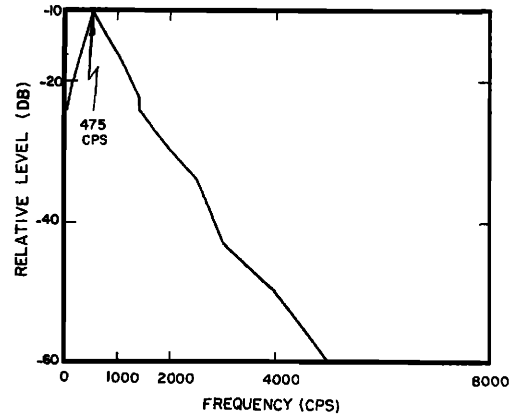

FIG. 2. Smoothed spectral envelope for trombone.

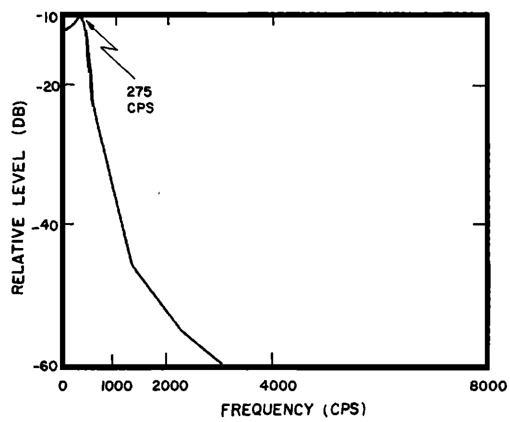

FIG. 3. Smoothed spectral envelope for tuba.

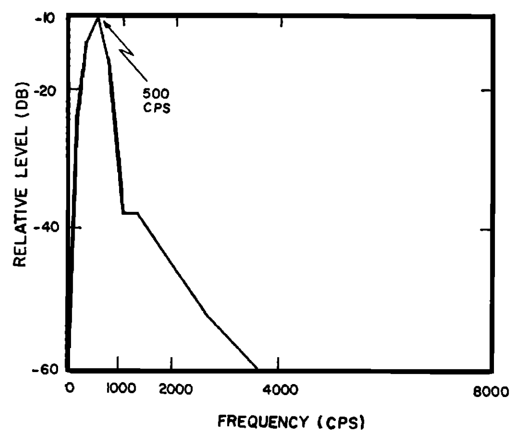

FIG. 4. Smoothed spectral envelope for French horn.

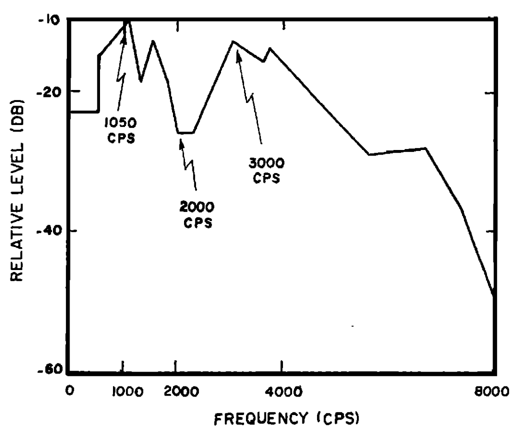

FIG. 5. Smoothed spectral envelope for oboe.

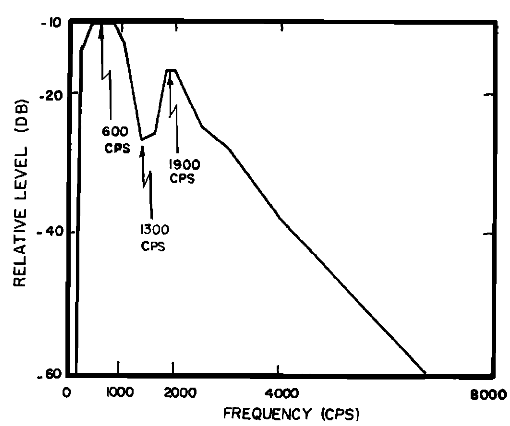

FIG. 6. Smoothed spectral envelope for English horn.

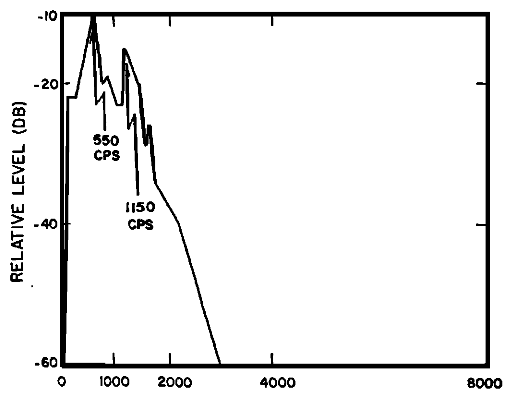

FIG. 7. Smoothed spectral envelope for bassoon.

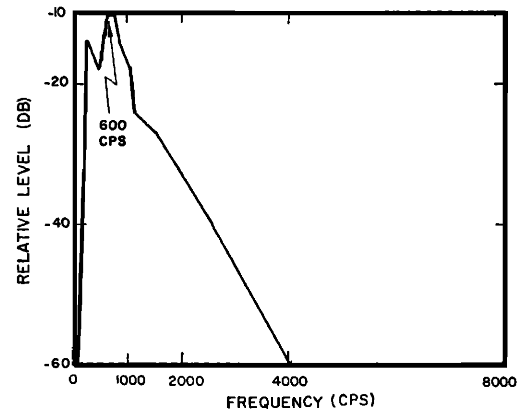

FIG. 8. Smoothed spectral envelope for flute.

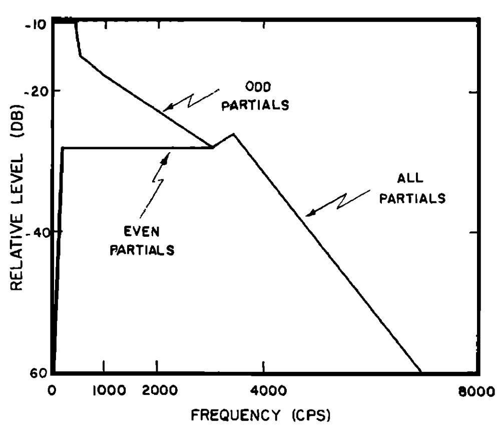

FIG. 9. Smoothed spectral envelopes for even and odd partials for clarinet. These envelopes are used for notes from 146 to 414 cps.

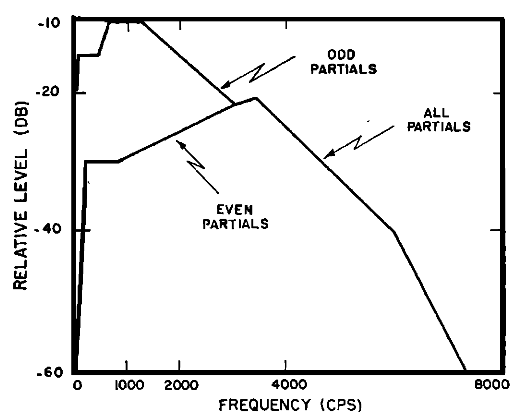

FIG. 10. Smoothed spectral envelopes for even and odd partials for clarinet. These envelopes are used for notes from 440 to 834 cps.

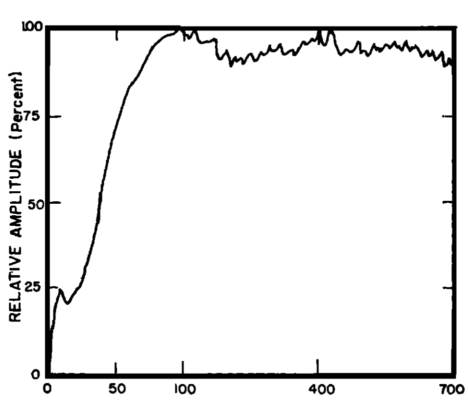

FIG. 11. Natural temporal envelope for a trumpet tone of 212 cps.

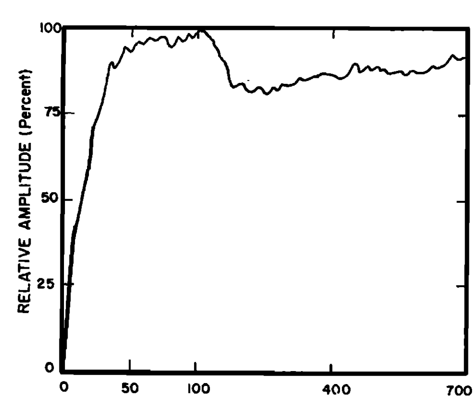

FIG. 12. Natural temporal envelope for a trumpet tone of 881 cps.

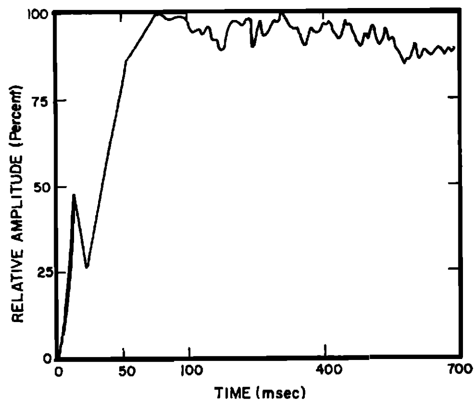

FIG. 13. Natural temporal envelope for a trombone tone of 128 cps.

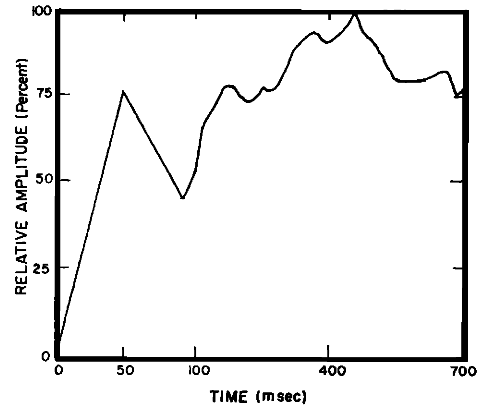

FIG. 14. Natural temporal envelope for a tuba tone of 43 cps.

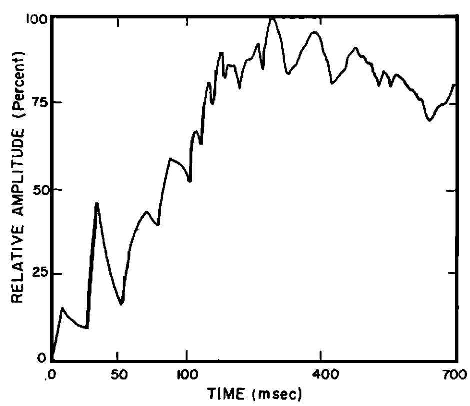

FIG. 15. Natural temporal envelope for a French horn tone of 115 cps.

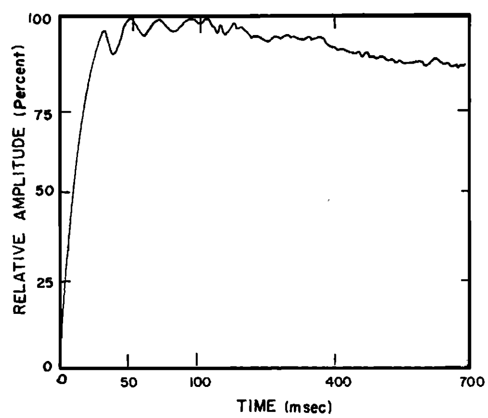

FIG. 16. Natural temporal envelope for a French horn tone of 447 cps.

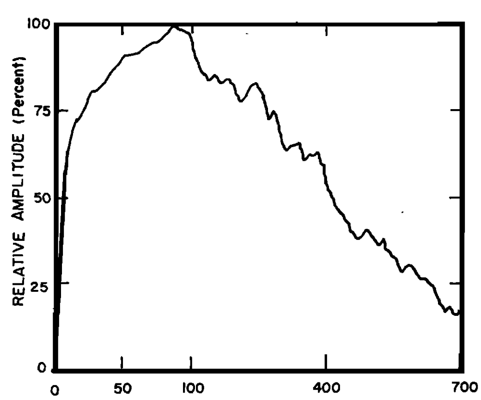

FIG. 17. Natural temporal envelope for an oboe tone of 466 cps.

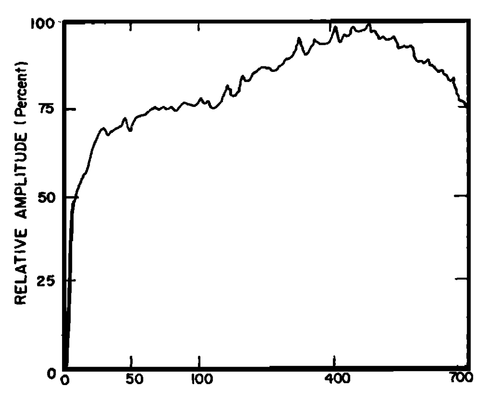

FIG. 18. Natural temporal envelope for an English horn tone of 315 cps.

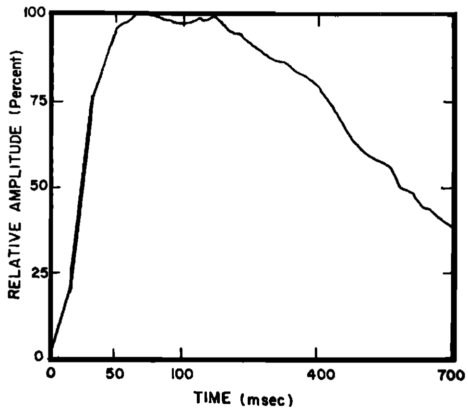

FIG. 19. Natural temporal envelope for a bassoon tone of 62 cps.

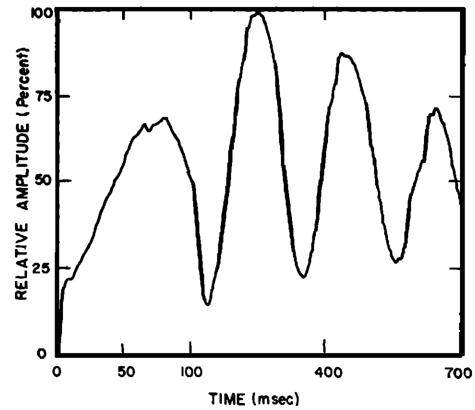

FIG. 20. Natural temporal envelope for a flute tone of 266 cps.

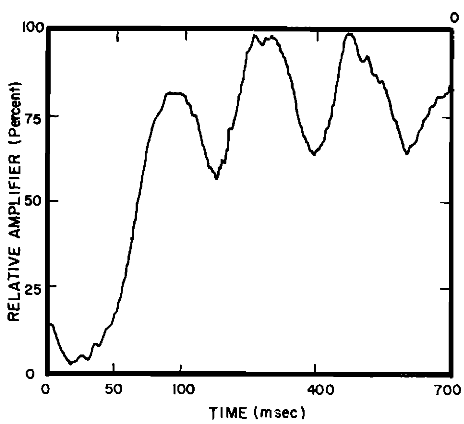

FIG. 21. Natural temporal envelope for a flute tone of 542 cps.

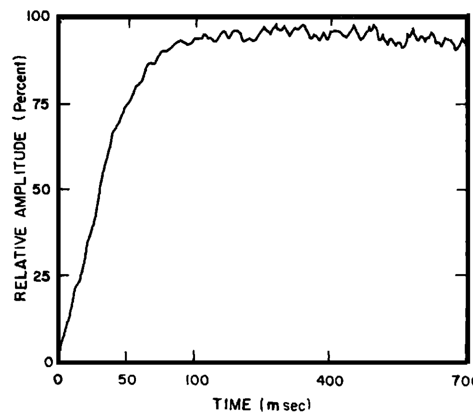

FIG. 22. Natural temporal envelope for a clarinet tone of 232 cps.

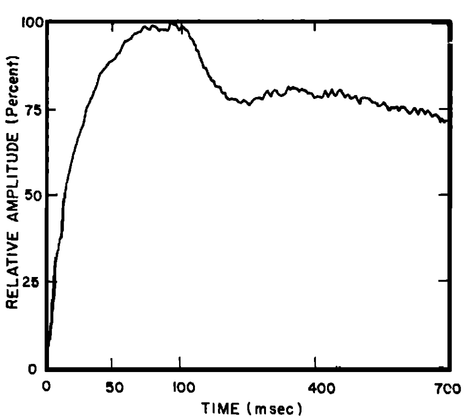

FIG. 23. Natural temporal envelope for a clarinet tone of 553 cps.

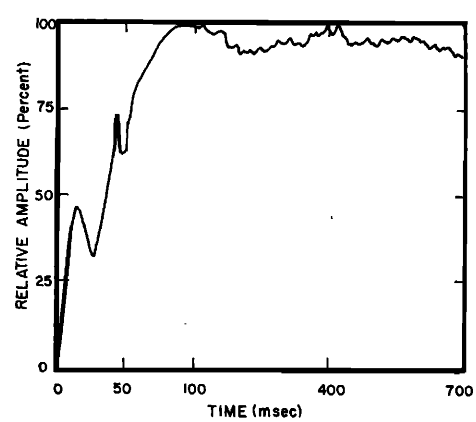

FIG. 24. $C_4(t)$ temporal envelope for a trumpet tone of 212 cps.

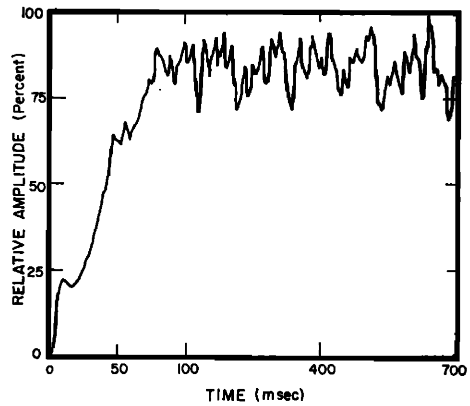

FIG. 25. $C_3(t)$ temporal envelope for a trumpet tone of 212 cps.

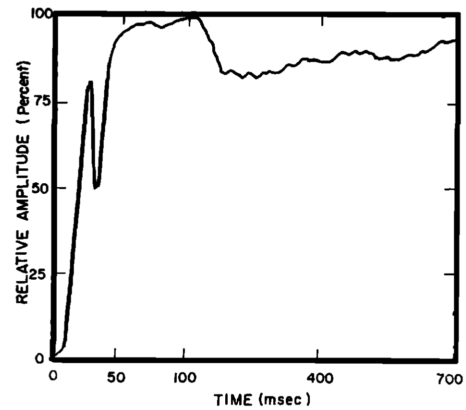

FIG. 26. $C_4(t)$ temporal envelope for a trumpet tone of 881 cps.

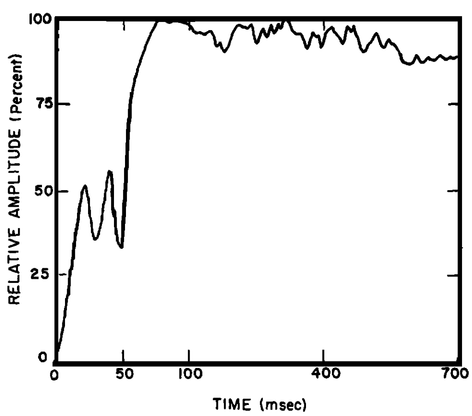

FIG. 27. $C_4(t)$ temporal envelope for a trombone tone of 128 cps.

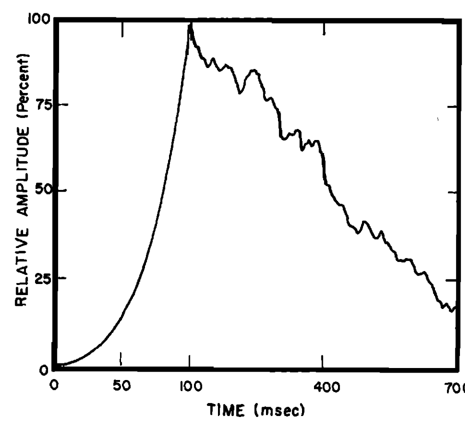

FIG. 28. $C_2(t)$ temporal envelope for an oboe tone of 466 cps.

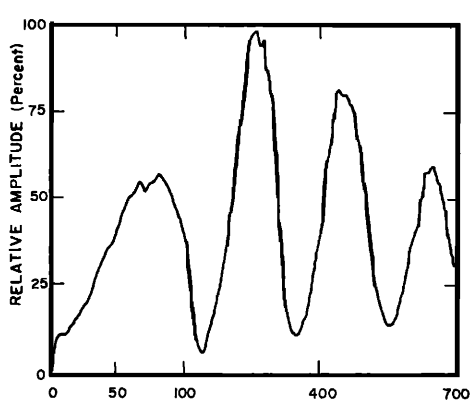

FIG. 29. $C_1(t)$ temporal envelope for a flute tone of 266 cps.

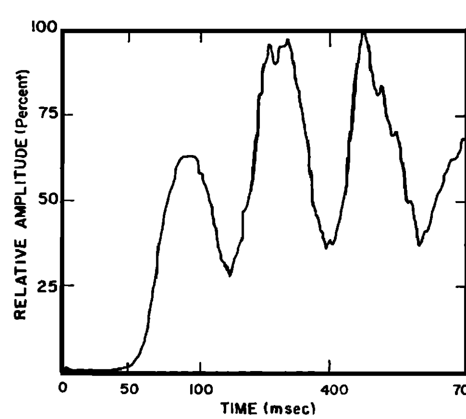

FIG. 30. $C_1(t)$ temporal envelope for a flute tone of 542 cps.

English horn and bassoon, is the existence of two peaks separated by a valley.

- For the oboe, the second group of partials increases initially more slowly than those in Group 1 and more rapidly and rather abruptly during the latter part of the attack transient.
- At high frequencies, the spectral envelope for the oboe is greater than that for other instruments.
- The partials of the English horn are much less consistent than those of any other wind instrument, and the spectral envelope for this instrument is a much less accurate representation of the amplitudes of the partials than that for any other instrument.
- The attack transients of the oboe and English horn are similar in shape to the attack transient for the bassoon; but the latter is longer in duration, and all partials rise approximately together.

III. EXPERIMENTAL PROCEDURES

Tones of 700-msec duration were synthesized digitally by an IBM 7090 computer at MIT's Computation Center. The 20 000 digital samples generated per tone at a rate of 28 000/sec excited a 12-bit, bipolar, digital-to-analog converter. Quantization noise was removed by a low-pass filter down by 3 dB at 10 kc/sec and by 54 dB at 20 kc/sec. The filter output was recorded on Ampex 611 professional recording tape by an Ampex 300 tape recorder running at 15 in./sec with NAB equalization. Several tones for each instrument were synthesized at representative frequencies for that instrument, as displayed in Table I.

The spectral envelope, the natural temporal envelope, the fundamental frequency, the variables that govern the division of partials into groups, and the information needed to construct the artificial temporal envelopes were read into the computer for each tone. The relative amplitudes of the partials were specified by the spectral envelope; the frequency or amplitude of a partial was used to classify it into its group. The partials in each group were combined with arbitrary phases, and their relative amplitudes were modulated with the temporal envelope peculiar to that group. Economy of computer time barred the generation of decay transients; these are of little aural importance to the characterization of nonpercussive tones.[7]

For each of the 59 synthetic tones, a natural one of the same frequency and amplitude was provided with the same duration by amputating the decay transient and some of the steady state. The 118 tones were recorded in a known, random order, and were numerially labeled vocally. The tones were reproduced at an approximately normal level for each instrument by a KLH-9 full-range electrostatic loudspeaker in the MIT listening studio, designed to give diffusion for psychoacoustic tests.

The natural tones were recorded in an anechoic chamber for reasons reported elsewhere,[18] using an Altec Lansing 21D microphone, an Ampex 300 tape recorder, and Irish 220 tape. Scales, with rests between each note and its successor, were played over the full range of each instrument at dynamic markings of *pp*, *mf*, and *ff*. (The instrumental control of sound-pressure level proved to be very artificial.[19]) The players were selected as the best from each of three local orchestras in the judgment of the conductors and fellow players.

As is well known, one who works for extended periods with audio signals adapts extensively to these signals and can no longer judge the quality of these signals. For this reason, to get objective measures of the quality of our syntheses and to provide a basis for comparison with the work of others (as well as with other work of our own), psychoacoustic tests were performed with eight musically literate auditors. These auditors were required to identify the instrument producing the tone as one of the following: trumpet, trombone, tuba, French horn, oboe, English horn, bassoon, flute, or clarinet. The auditors were given a short rest period after each 30–40 min of testing, but the test was completed in one session. To qualify as "musically literate," each auditor was required to identify accurately each instrument from its amputated, natural tones, apart from intrafamily confusions.[18] (The subjects had, in fact, played instruments themselves, had a long history of exposure to music, and were familiar with orchestral instruments. Several had been used in other psychoacoustical experiments.) Each tone was repeated as many times as any auditor requested, but each subject gave, and was required to give, one and only one response to each tone. As can be seen from the confusion matrix of Table II, the subjects were well qualified.

IV. VALIDITY OF THE SYNTHETIC TONES

The responses of the auditors to the synthetic tones are displayed in Tables II and III. Table II is the gross confusion matrix listing the tones we were trying to simulate against the instruments named in the response to those tones. It may be seen that the identifications of the instruments from synthetic tones were not so accurate as those from natural tones and that there were usually more confusions with instruments not in the same family with synthetic tones than was the case with natural tones. (Chance scores in all cases would be 11%.) Indeed, from the detailed results (not presented here) at each particular frequency for each particular instrument, the identifications, from the natural tones (apart from intrafamily confusions) were always as accurate as, or more accurate than, those

[18] M. Clark, P. Robertson, and D. Luce, "A Preliminary Experiment on the Perceptual Basis for Musical Instrument Families," J. Audio Eng. Soc. **12**, 194–203 (1964).
[19] M. Clark and D. Luce, "Intensities of Orchestral Instrument Scales Played at Prescribed Dynamic Markings," J. Audio Eng. **13**, 151–157 (1965).

Table I. Characterization of synthesized tones.

Instrument simulated	Frequencies of synthesized tones (cps)	Spectral envelopes	Temporal envelopes: Group 1 partials Character	Group 1 partials Definition	Group 2 partials Character	Group 2 partials Definition	Group 3 partials Character	Group 3 partials Definition
Trumpet	212 277, 350, 442, 551, 712	Fig. 1	Fig. 11	−10−−16 dB	Fig. 24	−16−−26 dB	Fig. 25	−26−−60 dB
	881	Fig. 1	Fig. 12	0–1000 cps	Fig. 26	1000–4000 cps	$C_3(t)$	4000–8000 cps
Trombone	98 128 163, 210, 266, 336	Fig. 2	Fig. 13	0–800 cps	Fig. 27	800–1300 cps	$C_3(t)$	1300–8000 cps
Tuba	43 66, 86, 110, 137, 172, 220	Fig. 3	Fig. 14	All partials				
French horn	115 141, 175, 221, 284	Fig. 4	Fig. 15	0–800 cps	$C_3(t)$	800–8000 cps		
	447	Fig. 4	Fig. 16	0–500 cps	$C_3(t)$	500–8000 cps		
Oboe	232, 292, 370 466 593, 747, 931	Fig. 5.	Fig. 17	−10−−20 dB	Fig. 28	−20−−28 dB	$C_3(t)$	−28−−60 dB
English horn	198, 248 315 420, 532, 723	Fig. 6	Fig. 18	−10−−18 dB	$C_2(t)$	−18−−26 dB	$C_3(t)$	−26−−60 dB
Bassoon	62 72, 93, 123, 154, 203	Fig. 7	Fig. 19	0–900 cps	$C_3(t)$	900–3000 cps		
Flute	266 361, 452	Fig. 8	Fig. 20	0–900 cps	Fig. 29	900–1800 cps	$C_3(t)$	1800–4000 cps
	542 677, 854	Fig. 8	Fig. 21	0–900 cps	Fig. 30	900–1800 cps	$C_3(t)$	1800–4000 cps
Clarinet	186 232 291, 348, 441	Fig. 9	Fig. 22	−10−−18 dB	$C_2(t)$	−18−−26 dB	$C_3(t)$	−26−−60 dB
	553 740, 996	Fig. 10	Fig. 23	−10−−18 dB	$C_2(t)$	−18−−26 dB	$C_3(t)$	−26−−60 dB

Table II. Probability, in percent, of naming instrument, listed for each instrument sounded (whether synthetic or natural).

Instrument sounded	Type of tones	Instrument named by musically literate auditors: Trumpet	Trombone	Tuba	French horn	Oboe	English horn	Bassoon	Flute	Clarinet
Trumpet	Natural	96	4							
	Synthetic	72	4			9	9			5
Trombone	Natural	2	92		6					
	Synthetic	6	58		10		13	10		
Tuba	Natural			89	11					
	Synthetic		4	61	20			14		2
French horn	Natural		35	2	63					
	Synthetic		13	6	54		2	19	6	
Oboe	Natural	4				75	12			9
	Synthetic	2	2			73	12		2	9
English horn	Natural	2			2	6	75	6	4	4
	Synthetic	23	11		6	4	42	2	2	8
Bassoon	Natural		6	8			6	79		
	Synthetic		6	8			21	65		
Flute	Natural								100	
	Synthetic	15	2			2	4		77	
Clarinet	Natural					2				98
	Synthetic	3				5	3		2	88

TABLE III. Frequencies of notes, probabilities of confusion, and confusion instruments. For any instrument simulated, the frequency of the note is listed in cps for each instrument confused with the simulated one; the probability, at each frequency generated, of naming the designated confusion instrument is listed in parentheses.

Instrument simulated	Instrument confused with simulated instrument								
	Trumpet	Trombone	Tuba	French horn	Oboe	English horn	Bassoon	Flute	Clarinet
Trumpet		212(13) 350(12)			277(12) 442(12) 551(37)	277(25) 350(38)			881(37)
Trombone	336(37)			128(12) 210(25) 266(12) 336(13)		210(38) 266(38)	98(37) 163(12) 266(13)		
Tuba		86(12) 172(13)		86(38) 172(37) 220(63)			43(12) 66(12) 86(12) 110(50) 172(12)		220(12)
French horn		141(38) 175(25) 284(13)	115(12) 175(13) 284(12)			141(12)	115(63) 141(25) 175(12) 221(12)	447(37)	
Oboe	747(12)	232(12)				232(50) 292(12) 593(12) 747(12)		593(12)	747(12) 931(50)
English horn	315(25) 420(37) 532(63) 723(12)	198(25) 315(13) 420(12) 532(13)		248(12) 420(25)	532(12) 723(12)		248(13)	723(13)	723(50)
Bassoon		72(13) 123(13) 203(12)	62(25) 72(12) 123(12)			154(62) 203(63)			
Flute	266(13) 361(25) 542(50)	542(12)			266(12)	266(25)			
Clarinet	996(25)				441(25) 740(13)	186(12) 348(12)		740(12)	

from the synthetic ones. All tones synthesized according to a particular algorithm are presented, not merely those that happened to come out well. It is surprising that a method that works well for many of the notes of an instrument may fail somewhat for other notes.

Comparison of the results in Table II reveals that synthetic brass tones tended to be confused with those of the double reeds, whereas the natural ones displayed no such confusion. In no case were synthetic brass tones identified as accurately as the natural ones. The trumpet and trombone tones showed nearly the same intrafamily confusions for both natural and synthetic tones. The synthetic tuba tones displayed more confusion with those of the French horn than did the natural ones. The synthetic French horn tones showed less confusion with trombone tones than did the natural ones, but the synthetic tones were more often confused with those of the double reeds.

Synthetic and natural oboe tones were equally satisfactory for identification. The synthetic English horn tones were very poor as compared with the natural ones in characterizing this instrument, and the synthetic ones led to much confusion with those of the brasses. Synthetic bassoon tones were as good as the natural ones for identification, except for more intrafamily confusion with those of the English horn.

For a few of the synthetic flute tones, there was a marked tendency to identify them as those of the trumpet, probably owing to too fast an attack. Several other synthetic flute tones were as good as the corresponding natural ones for identification purposes.

Roundoff errors, and occasional failure of our subjects to respond in each case, account for the tallies' not summing to 100% in Table II.

Table III lists the frequency of each note of any instrument with which any synthetic tone was confused and gives a measure of the errors of our syntheses as determined by the auditory process at each particular frequency. For those cases in which there was any confusion the probability (in percent) with which any synthetic tone at the particular frequency listed was confused with a given instrument is shown in parentheses. For example, of all synthetic tones created that purport to simulate those of the French horn at 175 cps, 25% were identified as those of the trombone, 13% as those of the tuba, and 12% as those of the bassoon. At such a low frequency, the confusions with tones of the trombone (or tuba) may not be regarded as serious, since the characteristics aurally distinguishing the horn from the trombone appear most strongly only above about 284 cps. However, the confusions with the bassoon would be regarded as serious if they were more probable.

From Table III, we observe that, at particular frequencies, confusions with instruments other than those simulated had a probability of occurrence of 50% or more in the case of the tuba note at 220 cps (horn) and at 110 cps (bassoon); the French horn note at 115 cps (bassoon); the oboe note at 232 cps (English horn) and at 723 cps (clarinet); the bassoon at 154 cps and 203 cps (English horn in both cases); and flute at 542 cps (trumpet). The confusions outside a family are to be regarded as much more serious than those within the same family. We remark that the timbres of a number of instruments belonging to different families but having approximately identical instrinsic frequencies tend to converge at the extremely high end of their scales.

With the list of fundamental frequencies at which notes were generated in Table I, we may summarize the confusions listed in Table III for synthetic tones as follows:

- *Trumpet* tones are confused with tones of the oboe and English horn in the midrange and with those of the clarinet at the highest note.

- *Trombone* tones are confused with those of the bassoon and English horn throughout most of the range of the trombone. The trombone tones (including the natural ones) tend to be confused with those of the trumpet and French horn—a tendency more pronounced for the synthetic tones at the highest note generated.

- *Tuba* tones are confused with those of the bassoon, trombone, and most especially with those of the French horn.

- *French horn* tones are confused with those of the trombone and more strongly with those of the bassoon. The highest horn tone is confused with flute tones. (Note that natural, low horn tones are strongly confused with trombone tones.)

- Low *oboe* tones are confused with English horn tones; and high oboe tones, with clarinet tones.

- High *English horn* tones are confused with those of the other double reeds, clarinet, flute, French horn, and especially the trumpet and trombone.

- *Bassoon* tones are confused with those of the trombone and tuba in the upper and lower parts of the range and with tones of the English horn in the upper range.

- *Flute* tones, if confused at all, are identified as those of the trumpet.

- In the lower range, the *clarinet* tones are confused as those of the English horn and in other ranges with tones of the trumpet, oboe, flute, and English horn.

We may summarize our results as follows: Synthetic oboe tones are as good as the natural ones for identification purposes. Synthetic bassoon and clarinet tones are nearly as good as the natural ones. The synthetic flute tones are somewhat less valid, but are still consistently identified as those of the flute. All of the synthetic brasses are identified in over 50% of all cases. If intrafamily confusions are overlooked, then the synthetic brass tones are about as good as synthetic flute tones. Synthetic English horn tones are rather unsatisfactory.

It is thought that the subject would have difficulty in distinguishing simulated and natural tones, for several of the better cases, in a paired comparison.

The significance of spectral envelopes, temporal envelopes, and perturbations therein will be reported later.

ACKNOWLEDGMENTS

This work was supported in part by a grant from the National Association of Music Merchants. The assistance of Dr. Michael P. Barnett, Dr. Irving Kaplan, Dr. David Luce, and Ercolino Ferretti is very much appreciated. The computations were performed at the MIT Computation Center and the Cooperative Computing Laboratory. This work would have been impossible without the aid of many subjects and players who volunteered their time and services.