XIV. SIGNAL PROCESSING*

Academic and Research Staff

Prof. A. G. Bose
Prof. J. D. Bruce
Prof. A. V. Oppenheim
Prof. C. L. Searle
Prof. H. J. Zimmermann
Dr. M. V. Cerrillo

Graduate Students

J. B. Bourne
M. F. Davis
D. A. Feldman
J. M. Kates
R. M. Mersereau
J. R. Samson, Jr.
R. M. Stern, Jr.

A. MUSICAL TIMBRE RECOGNITION BASED ON A MODEL OF THE AUDITORY SYSTEM

As reported previously,[1] we have been investigating the relative perceptual importance of the spectral and temporal characteristics of musical notes.[2] To carry out these studies, we constructed a model of the monaural spectral analysis operations of the peripheral auditory system, and analyzed single musical notes passed through the model. Initial results confirmed the importance of the steady-state spectrum and suggested a closer study of the attack transients than had previously been made with either psychophysical or analytic techniques.

The most significant temporal characteristic found in individual note samples was a distinct ordering of the onset or starting times of the partials of the note (Fig. XIV-1). Because the onset time of each partial is measured relative to the onset time of the fundamental, the data for several notes can be averaged by using the formant curve technique developed by Strong and Clark for the amplitudes of the individual partials.[3] The results of applying this technique are shown in Figs. XIV-2, XIV-3, and XIV-4 for clarinet, French horn, and trumpet, respectively. The present approach of averaging with respect to frequency produced results with less variance than had been reported previously,[4] which were obtained by averaging with respect to partial number. Clearly, the onset times, if indeed they are characteristic of an instrument, cannot assume a pattern consistent with both partial number and partial frequency; therefore, we now favor averaging with respect to frequency.

The 10-90% rise times of the partials and the average derivatives over the rise interval were also computed. Unlike the amplitude and onset data, rise times and derivatives are absolute quantities and cannot be averaged by using the formant technique. Instead, simple means and variance were computed for comparison with similar quantities obtained from the onset data. As for the onset data, averages expressed as a function of partial number showed greater variance than those expressed as a function of frequency,

*This work was supported by the Joint Services Electronics Programs (U.S. Army, U.S. Navy, and U.S. Air Force) under Contract DAAB07-71-C-0300.

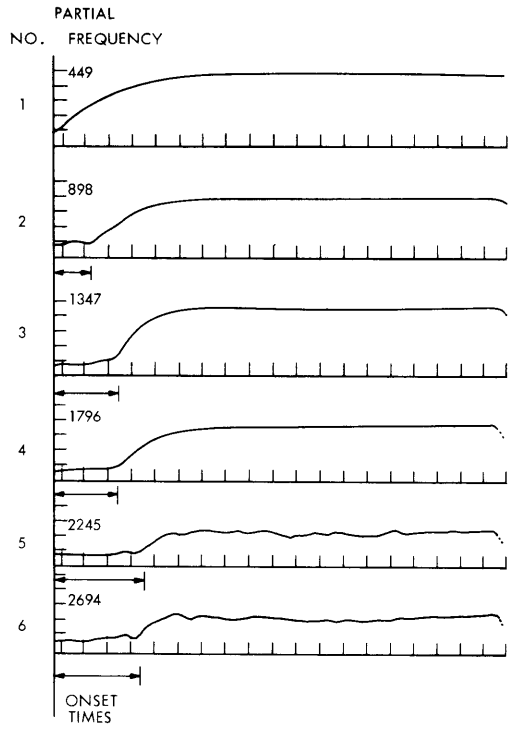

Fig. XIV-1. Envelopes of the partials of a typical clarinet note. Scale: vertical, 10 dB/div; horizontal, 10 ms/div.

and therefore were rejected. In general, all of the derivative and rise-time data showed greater variance than either the spectra or the onset-time data.

A Bayesian template-matching recognition algorithm was written to assess the perceptual importance of these characteristics, considered both individually and in groups. This algorithm is not being suggested as a model for perception, but only as a method

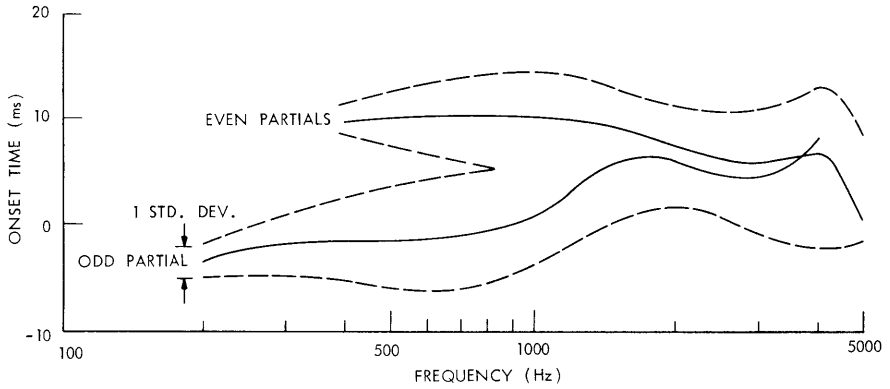

Fig. XIV-2. Clarinet onset data.

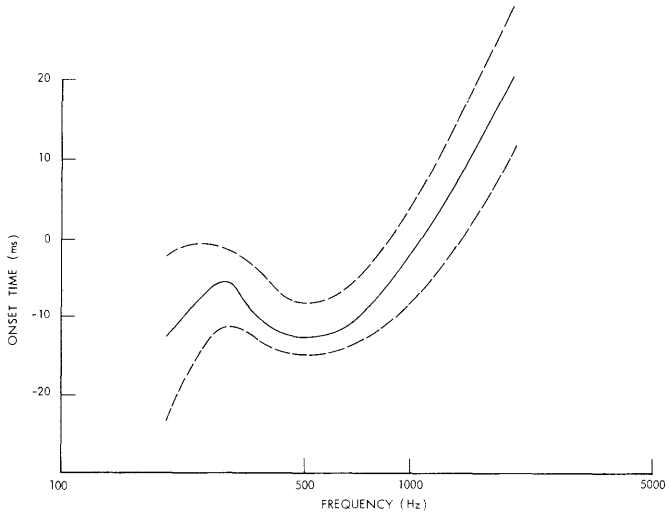

Fig. XIV-3. French Horn onset data.

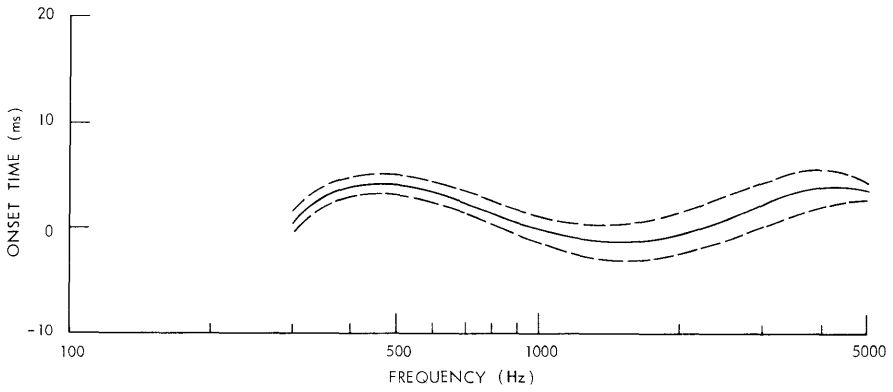

Fig. XIV-4. Trumpet onset data.

of determining the relative contributions of spectrum, onset, rise time, and derivative patterns, based on the variance of those patterns. The program as applied to the present data calculated the probability of an unknown note sample being one of three possibilities. Fifteen samples, eight of which were not used in computing the templates for the decision algorithm, were applied to the program. For each note, decisions were calculated by using each of several combinations of the four characteristics. The combination in which the program converged to the highest probability was considered the best. Typical values were from 0.5 to 0.7, the chance probability being 0.33. The program converged

to the correct decision in 14 of 15 cases. The probabilities for the 14 correct answers were rank-ordered; these ranks were then averaged with respect to the characteristics used to arrive at the decision.

Table XIV-1. Rank ordering of recognition results.

Characteristics				Score
AMP	ONS	RTM	DER	40
AMP	ONS			49
AMP	ONS	RTM		53
AMP	ONS		DER	55
AMP		RTM		82
AMP			DER	84
AMP		RTM	DER	85
AMP				89
	ONS	RTM	DER	104
	ONS			112
			DER	150
		RTM		156

Note: Worst possible score = 168. Best possible score = 14.

The results, as shown in Table XIV-1, clearly indicate the basic importance of the steady-state amplitudes; indeed, using only amplitude (AMP) data in this test produced better results than using only attack information. The use of onset (ONS) data, however, produced significantly better results than the use of rise-time (RTM) or derivative (DER) data, either separately or together. As might be expected, the best identifications resulted from using all of the data, although the derivative and rise-time data, when used together, contributed far less to the identification than did the onset-time data.

J. B. Bourne

References

1. J. B. Bourne, "Musical Timbre Recognition Based on a Model of the Auditory System," Quarterly Progress Report No. 104, Research Laboratory of Electronics, M. I. T., January 15, 1972, pp. 362-365.

2. J. B. Bourne, "Musical Timbre Recognition Based on a Model of the Auditory System," S. M. and E. E. Thesis, Department of Electrical Engineering, M. I. T., June 1972.

3. W. J. Strong and M. Clark, "Synthesis of Wind Instrument Tones," J. Acoust. Soc. Am. 41, 39-52 (1967).

4. J. B. Bourne, Quarterly Progress Report No. 104, op. cit., see Fig. XXIV-8, p. 363.

B. PERCEPTION OF SIMULTANEOUSLY PRESENTED MUSICAL TIMBRES

An investigation of human perception of musical instrument timbres is now under way, in an attempt to determine the way in which one can hear and "track" individual instruments separately and selectively, even though many of them may be played at once. This report summarizes some preliminary findings of the study.

A series of synthetic trumpet and clarinet tones was prepared on the Speech Communication Group's PDP-9 computer. The tones were generated by modeling the musical notes with a seven-harmonic periodic function whose amplitudes as a function of time were piecewise-linear approximations to the envelope curves produced by the analysis method described by Bourne.[1] The duration of the synthesized notes was approximately 200 ms, with an artificial 10-ms linear decay transient. This skeletal representation does not take into account the effects of small fluctuations in amplitude and frequency, inharmonic partials, or noise associated with the instrument. All of these factors may play a part in making a tone sound realistic and, because of their omission, some of the synthesized tones (particularly those of the trumpet) tended to sound artificial. It was always possible, however, to identify the instrument that a single synthesized note was imitating.

The computer-generated tones were added and presented simultaneously in pairs, one from each "instrument," at the musical intervals of the perfect fourth, major third, minor third, and minor second, each being repeated 10 times. Informal tests showed that musically trained listeners could consistently identify the instrument on which each note was played. This would imply that it is possible to perceive two different timbres presented simultaneously, without external cues such as melodic context, provided that the sounded notes are separate in frequency.

Tone combinations were also prepared in which trumpet and clarinet sounded the same note at the same time. In all of these cases the resultant musical complex sounded like neither trumpet nor clarinet, nor like a predictable mixture of their perceptual qualities. Time-delaying one note of a unison pair with respect to the other enabled listeners to perceive the timbre of the earlier tone more clearly, but the timbre of the following tone could not be identified with delays at least as long as 100 ms. Hence it does not appear possible to perceive separately the timbres of two different instruments synthesized in the manner that we have described if exactly the same note is played on

both instruments. In fact, the first note that is sounded will, in some sense, mask the timbral perception of succeeding tones at the same frequency for a period of time.

We are now examining these phenomena more rigorously.

R. M. Stern, Jr.

References

1. J. B. Bourne, "Musical Timbre Recognition Based on a Model of the Auditory System," S.M. and E.E. Thesis, Department of Electrical Engineering, M.I.T., June 1972.