Multidimensional perceptual scaling of musical timbres*

John M. Grey

Center for Computer Research in Music and Acoustics, Department of Music, Stanford University, Stanford, California 94305
(Received 12 August 1976; revised 23 November 1976)

Two experiments were performed to evaluate the perceptual relationships between 16 music instrument tones. The stimuli were computer synthesized based upon an analysis of actual instrument tones, and they were perceptually equalized for loudness, pitch, and duration. Experiment 1 evaluated the tones with respect to perceptual similarities, and the results were treated with multidimensional scaling techniques and hierarchic clustering analysis. A three-dimensional scaling solution, well matching the clustering analysis, was found to be interpretable in terms of (1) the *spectral energy distribution*; (2) the presence of *synchronicity* in the transients of the higher harmonics, along with the closely related amount of *spectral fluctuation* within the the tone through time; and (3) the presence of *low-amplitude, high-frequency energy* in the initial attack segment; an alternate interpretation of the latter two dimensions viewed the cylindrical distribution of clusters of stimulus points about the spectral energy distribution, grouping on the basis of musical instrument *family* (with two exceptions). Experiment 2 was a learning task of a set of labels for the 16 tones. Confusions were examined in light of the similarity structure for the tones from experiment 1, and one of the family-grouping exceptions was found to be reflected in the difficulty of learning the labels.

PACS numbers: 43.66.Jh, 43.75.Wx, 43.66.Lj

INTRODUCTION

In discussing the attributes of complex tones, Licklider (1951) concluded that "until carful scientific work has been done on the subject, it can hardly be possible to say more about timbre than that it is a 'multidimensional' dimension." In the last decade, investigators of timbre have taken marked steps in the direction of finally being able to deal with the multidimensionality of their subject. These recent improvements in timbre research are largely the result of technological advances in the use of digital computers which have given the investigator powerful new means for the analysis and synthesis of complex, time-variant musical instrument tones (Luce, 1963; Risset, 1966; Freedman, 1967, 1968; Beauchamp, 1969; Mathews, 1969; Chowning, 1973; Moorer, 1975), and for the analysis and presentation of complex, multidimensional data structures of the type that may be collected from studying the perception of timbre (Plomp, 1970; Wessel 1973, 1974; Miller and Carterette, 1975; Grey, 1975; Grey and Moorer, 1976).

A major aim of research in timbre perception is the development of a theory for the salient dimensions or features of classes of sounds. Given that timbre is clearly a *multidimensional* attribute of sound, the computer-based techniques of perceptual data analysis which are in the category of multidimensional scaling seem to be especially well-suited for examining the complex aspects of timbre perception (Shepard, 1962a, 1962b; Kruskal, 1964a, 1964b; Carroll and Chang, 1970). In multidimensional scaling, the investigator gathers perceptual data consisting of the subjective similarities between all pairs in a set of stimuli. The perceptual similarity judgment is treated as a measurement of subjective *distance*, from which a best-fitting geometric *map* of the stimuli, plotted as points in a space, is constructed. The multidimensional scaling algorithm maps the subjective distance relationships into a geometric space which has the number of dimensions specified by the investigator.

Although the scaling algorithm reports a goodness-of-fit statistic between the constructed spatial distances and the subjectively judged distances of the stimuli, a most important evaluation of any particular geometric mapping of similarities is its usefulness in interpreting the bases for the perceptual judgments. The investigator attempts to interpret the geometric configuration of stimulus points with regard to the factors which may explain the ordering of points along the various axes, or dimensions, of the *space*. An alternate approach has been to explain the factors involved in the clustering of points in the space.

Multidimensional scaling techniques are attractive in the investigation of extremely complex stimuli (such as music instrument tones) since the researcher need not physically construct his stimuli to conform to the test of a specific *a priori* hypothesis. Rather, he may start with the perceptual judgments of similarity among a diverse set of (naturalistic) stimuli, and then explore the various factors which contributed to the subjective distance relationships. These factors may be physical parameters of the stimuli, which then would lead to a psychophysical model; yet, multidimensional scaling techniques may also uncover any other factors involved in judgment strategies.

In the current investigation we have collected two sorts of psychological distance measurements between 16 music instrument notes. One measurement was a judgment of the timbral similarities for all pairs of the 16 instrument notes. Here, the psychological distance of two tones was related to the inverse of their similarity. This data was processed with multidimensional scaling. The other measurement taken was of the accuracy of listeners in associating specific names with the notes, in a learning task with feedback. In this case the psychological distance of two tones was related to the number of confusions that occurred between them

in learning their respective names. A comparison was of these two measurements.

The instrument notes were previously equalized for perceived pitch, loudness, and duration, in order to eliminate confounding dimensions from the judgments on timbre (Grey, 1975). We feel that this step, which has often been missing in previous studies, was necessary because timbre is classically defined to exclude the dimensions of pitch, loudness, and duration. Furthermore, one may expect that detectable differences in pitch, loudness, or duration would show up as additional dimensions in the stimulus space—presenting a judgment context which would minimize the attention paid to timbral differences alone. [Miller and Carterette, (1975) present an example of the effect of pitch on similarity structures for tones.]

A second important feature of the stimuli in these studies was that they were computer generated by an analysis-based synthesis algorithm (Grey and Moorer, 1976). Specifically, the tones were synthesized from information derived from the waveforms of actual recorded and digitized musical instrument tones. By using synthetic stimuli rather than stimuli on tape, the frequencies, amplitudes, and durations could be altered as specified, thereby allowing for the independent perceptual equalization of pitch, loudness and duration stated above.

An equally important benefit derived from the use of synthetic stimuli was that the exact physical properties of the tones were known. Formulation of the psychophysical relationships involved in timbre perception depends upon a knowledge of the physical nature of the stimuli as much as it does upon the perceptual scaling. The most direct and reliable information about the physical properties of stimuli is a knowledge of the parameters used for their synthesis. This type of information is much stronger than any *post facto* analysis of stimuli. The use of synthetic stimuli also allowed for a control over the complexity of the physical makeup of the tones. Synthetic tones were used which were almost indistinguishable from the originally taped tones, but which had greatly simplified physical-control parameters, compared to the direct analysis of the originals (see Grey and Moorer, 1976). It would be expected that the simplification of the physical representations of the stimulus tones correspondingly would simplify the task of psychophysical interpretations of the perceptual data.

I. EXPERIMENT 1. MULTIDIMENSIONAL SCALING OF TIMBRAL SIMILARITIES

A. Stimuli

The stimuli were derived from 16 instrumental notes played near the pitch of E♭ above middle-C, approximately 311 Hz, whose durations ranged between 280 and 400 msec. The tones were performed and recorded in an IAC acoustical isolation chamber, which is a very dry environment, but passes some small amount of reverberation. The recording was done on a Revox tape recorder, half-track monaural at 7.5 ips, using Scotch 206 $\frac{1}{4}$-in. low-noise tape. They were then digitized by an Analogic 14-bit analogue-to-digital converter at a rate of 25 600 samples per second, and the high-order 12 bits were stored in digital form for computer analysis or playback. The 16 instrumental tones consist of two oboes (different instruments and players), English horn, bassoon, E♭ clarinet, bass clarinet, flute, two alto saxophones (one instrument, played at p and mf), soprano saxophone, trumpet, French horn, muted trombone, and three celli (one instrument, played normally, muted sul tasto, and sul ponticello).

The tones were digitally analyzed with a heterodyne filter technique (Moorer, 1975; Grey and Moorer, 1976). This technique produced a set of time-varying amplitude and frequency functions for each partial in the instrumental tone. Digital additive synthesis was used to produce a synthetic tone, where a set of partials, each controlled through time in amplitude and frequency, were added together. The functions obtained from the analysis were considerably simplified before being used to control the synthesized partials. Where the original functions were quite complex, consisting of 300 to 500 linear segments each, the data-reduced functions consisted of only 4 to 8 line segments. The data reduction was performed empirically, where there was an attempt to model activity in the attack of the tone, including initial, low-amplitude inharmonicity. The resulting signals were difficult to discriminate from signals based on the more complex, originally analyzed functions (Grey and Moorer, 1976).

The stimuli were equalized in an on-line experiment for loudness, pitch, and perceived duration (Grey, 1975). They were then recorded on tape after being played back from the computer through a 12-bit digital-to-analogue converter at a rate of 25 000 samples per second. The recording was done on a Sony 854-4 tape recorder, quarter-track monaural at 7.5 ips, on Maxell U50 $\frac{1}{4}$-in. low-noise tape. Playback was accomplished by means of a Dynaco PAT-4 preamplifier and Dynaco Stereo 120 amplifier connected to an Altec 804-E loudspeaker, and was done in a moderately reverberant room approximately 12×12 ft in dimensions. Playback level was a comfortable, moderate level for listeners.

B. Listeners and procedure

Twenty listeners at Stanford University were employed for this experiment, 15 of whom took the experiment a second time, making a total of 35 data sets collected. Listeners were musically sophisticated, some actively involved in advanced instrumental performance and others in conducting or musical composition.

Data sets were collected in an hour session. Each trial consisted of a warning knock, the two tones for comparison, and a decision interval. The warning knock preceded the first tone by 2.5 sec, 1.5 sec separated the two tones, and 6 sec were given for the listener to make a judgment. There were a total of 270 trials, 30 of which were practice trials. The remaining 240 trials consisted of the $n*(n-1)$ possible pairs of 16 tones, given in both directions. Trials were presented in a random order.

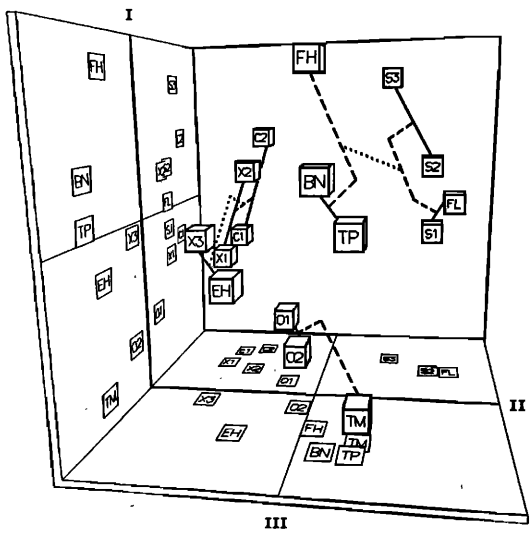

FIG. 1. Three-dimensional spatial solution for 35 similarity matrices generated by multidimensional scaling program INDSCAL (Carroll and Chang, 1970). Hierarchical clustering analysis (Johnson, 1967) is represented by connecting lines, in clustering strengths order: solid, dashed, dotted. Two-dimensional projections of the configuration appear on the wall and floor. Abbreviations for stimulus points: O1, O2 = oboes; C1, C2 = clarinets; X1, X2, X3 = saxophones; EH = English horn; FH = French horn; S1, S2, S3 = strings, TP = trumpet; TM = trombone; FL = flute; BN = bassoon.

Listeners were told to rate the *similarity* of the two tones, relative to that of all other pairs of tones heard. They were instructed that the first 30 pairs were practice, and that they could change their rating strategies during that time. The similarity rating was made on a scale of 1 to 30, and this scale was presented to listeners as having three general ranges: (1) 1–10—*very dissimilar*, (2) 11–20—*average level of similarity*, and (3) 21–30 = *very similar*, relative to all pairs.

C. Results and discussion

The similarity judgments for each listener were stored as a 16 by 16 matrix of data, recording responses with respect to the exact order of presentation for any pair of stimuli. The overall Pearson product-moment correlation between the 35 pairs of half-matrices was 0.905, indicating that there were differences in judgments with respect to the order of presentation. An examination was made for the existence of consistent order-related response differences across all listeners for any pair of stimuli. A half-matrix of order-related response differences was generated for each listener by subtracting the upper from the lower half of the response matrix. A Student t-test was made for each cell of the half-matrix across the 35 listeners for a consistent direction of differences. None of the 120 cell means for the 35 half-matrices were significantly different from zero, and they were all within 0.8 standard deviations of zero. Therefore the original response matrices were transformed into half-matrices for each listener by averaging the two responses to a pair of stimuli presented in different orders.

The 35 averaged half-matrices were treated with a multidimensional scaling algorithm that takes individual differences into account, INDSCAL (Carroll and Chang, 1970). Spatial representations were obtained in two, three, and four dimensions. Goodness-of-fit measures, defined as the correlation between the scalar products of the actual and predicted distances, for the solutions were: 0.78 for four dimensions, 0.75 for three dimensions, and 0.68 for two dimensions.

In order to compare the similarity structures between the first and second runs of the 15 listeners who gave two data sets each, the two respective sets of 15 half-matrices were solved for in two and three dimensions using INDSCAL. The spatial solution for the first run was rotated to best fit that of the second run, and Pearson product-moment correlations were found for spatial coordinates. The correlations for both the two-dimensional and three-dimensional cases were 0.98, suggesting that there were no fundamental changes in response strategy for the similarity judgments between runs.

In addition to subjecting the similarity matrices to multidimensional scaling analysis, they were also treated with a hierarchical clustering algorithm (HICLUS by Johnson, 1967). A group matrix was formed by averaging the rank orders of the ratings in the 35 individual matrices (since HICLUS works on rank orders of responses). The clustering algorithm produced an analysis of the similarity data which was independent of the spatial-dimensional reduction generated by multidimensional scaling. The *compactness*, or diameter, method of clustering was found to give the most interpretable results. The analysis grouped the most similar stimuli into clusters and then grouped such clusters into high-order clusters, continuing this way until the whole set of stimuli were in one cluster. Cluster strength reflected degree of similarity, so that the lowest level clusters were the strongest.

The INDSCAL and HICLUS analyses were used in conjunction with one another to interpret the data (see Shepard, 1972). The three-dimensional INDSCAL solution was found to be the most useful for interpreting the similarity structure of the stimuli. The two-dimensional spatial solution presented several discrepancies with the clustering analysis, and as a spatial solution was difficult to interpret. The three-dimensional solution overcame the problems of clustering and seemed more interpretable. However, there was no benefit found for interpreting the data by increasing the number of dimensions to four.

The three-dimensional INDSCAL solution is shown in the perspective plot of Fig. 1. Dimension I is the vertical axis, II is the horizontal axis, and III is the depth axis. The abbreviations adopted for the 16 tones are O1 and O2 = the two oboes; EH = the English horn; BN = the bassoon; C1 = the E♭ clarinet and C2 = the bass clarinet; X1 and X2 = the two saxophone tones (*mf* and *p* respectively); X3 = the soprano sax; FL = the flute; TP = the trumpet; FH = the French horn; TM = the muted trombone; S1, S2, and S3 = the cello tones (labelled strings: *sul ponticello*, normal bowing, and muted *sul tasto*, respectively).

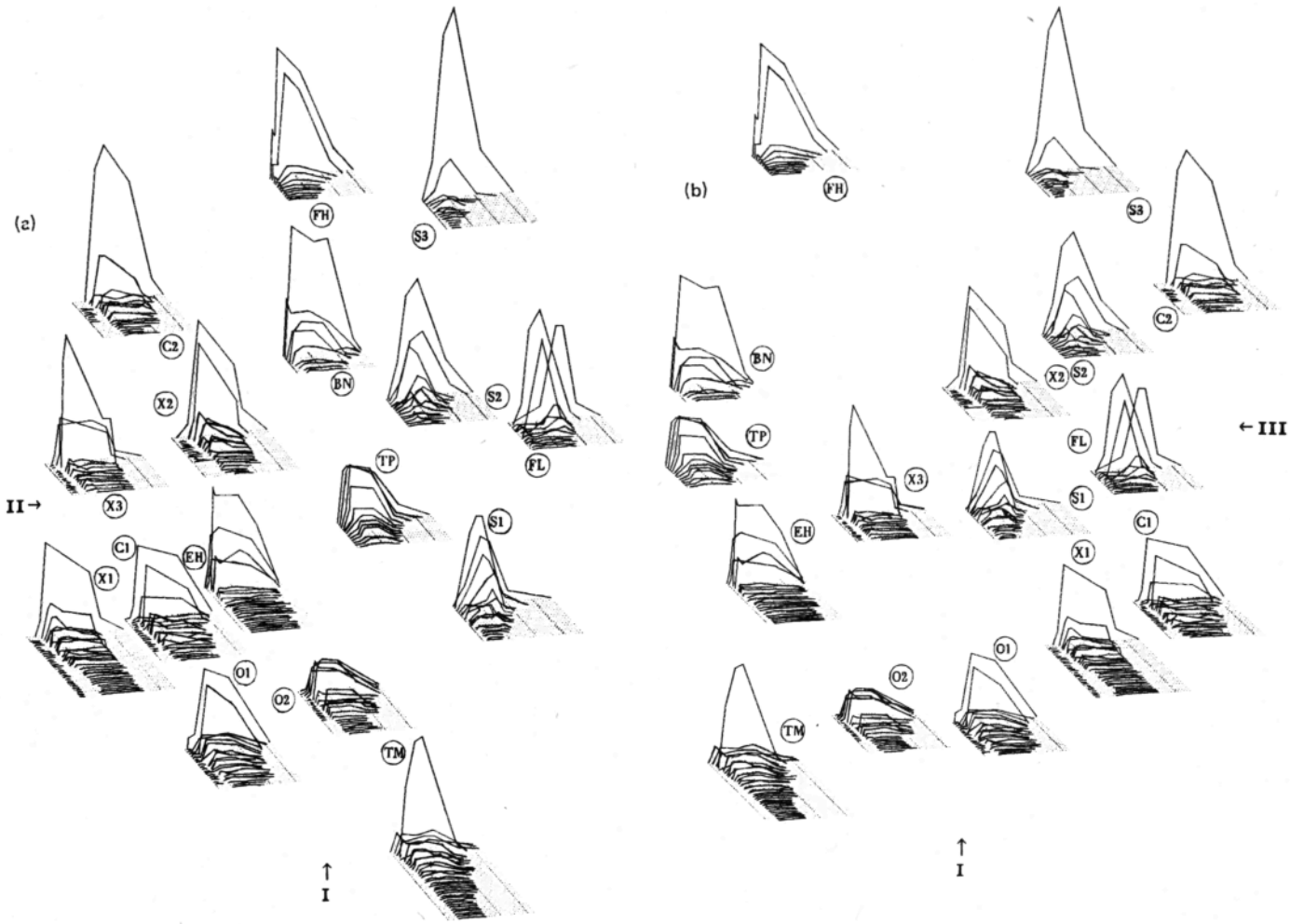

FIG. 2. (a) Two-dimensional projection of axis I versus axis II of Fig. 1. Stimulus points are neighbored by their physical representations in terms of (amplitude) × (time) × (frequency) perspective plots (amplitude = ordinate; time = abscissa; frequency = depth axis, with fundamental harmonics furthermost behind). (b) Two-dimensional projection of axis I versus axis III, as in (a).

The distances of the stimuli are given by their relative sizes and the two-dimensional projections of the configuration on the wall and the floor. The hierarchical clustering analysis is also represented in this spatial solution by lines connecting the components of the clusters. The primary-level clustering, having the greatest clustering strengths, are represented by unbroken straight lines connecting the elements. The next level, weaker clustering, is represented by dashed lines, and the highest level, weakest clustering, is shown by dotted lines in the space.

To aid in making a psychophysical interpretation of this spatial solution, two-dimensional projections of axes I versus II and of axes I versus III have been prepared, in which stimulus points are represented with two different types of physical analyses of the tones. Figures 2(a) and 2(b) present (amplitude) × (time) × (frequency) plots of the stimuli, and Figs. 3(a) and 3(b) present spectrographic displays to the tones (see Grey and Moorer, 1976).

A physical interpretation can be attempted of the three-dimensional space in terms of the spectral energy distribution and certain temporal features of the tones. Axis I can be interpreted with respect to the *spectral energy distribution*. On the one extreme, FH and S3 have the properties of narrow spectral bandwidth and a concentration of low-frequency energy. On the other extreme, TM has a very wide spectral bandwidth and less of a concentration of energy in the lowest harmonics; O1 and O2 have significant upper formant regions about the 10th harmonic.

Axis II would appear to relate to the form of the onset–offset patterns of tones, especially with respect to the presence of *synchronicity* in the collective attacks and decays of upper harmonics. Related to this is the overall amount of *spectral fluctuation* through time for a tone. The woodwinds are on one extreme and tend to have upper harmonics enter, reach their maxima, and often exit in close alignment. Their spectra tend to have little fluctuation in shape while all harmonics are present. On the other extreme, the strings, brass, FL, and BN do not have harmonics which display such synchronicity, but rather tend to have tapering patterns of entrances and/or exists. Correspondingly higher amounts of spectral fluctuations through time are found in these tones. It should also be pointed out that axis II may also be considered to express a musical instrument family partitioning. With two exceptions, the

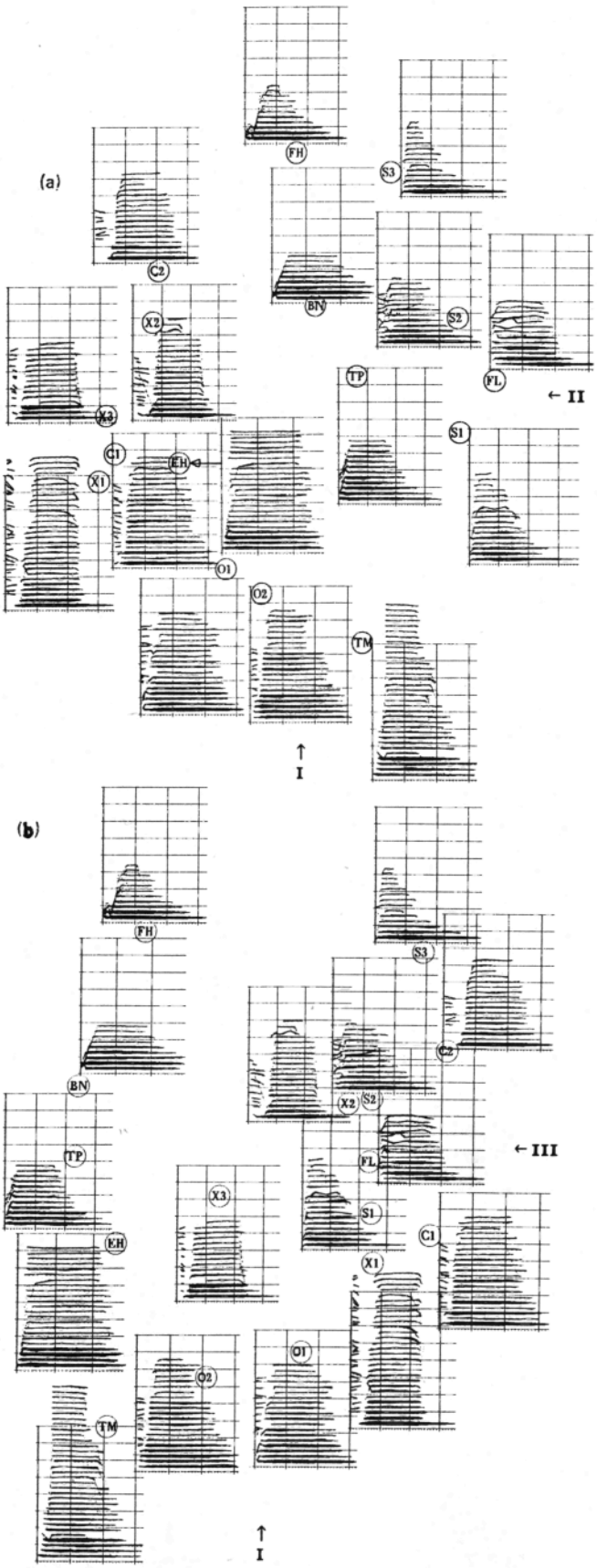

FIG. 3. (a) Two-dimensional projection of axis I versus axis II of Fig. 1. Stimulus points are neighbored by their physical representations in terms of spectrographic plots (frequency = ordinate; time = abscissa; energy proportional to width of vertical bars about the frequency of energy). (b) Two-dimensional projection of axis I versus axis III, as (a).

woodwinds appear on the far left, the brass in the middle, and the strings on the far right. The exceptions to this pattern are the clustering of BN with the brass and FL with the strings.

Axis III can also be interpreted in terms of temporal patterns. On this axis, the tones on one extreme, the strings, FL, clarinets, X1, X2, and O1 display *precedent high-frequency, low-amplitude energy*, most often *inharmonic energy*, during the attack segment. The tones at the other extreme, which include the brass, BN, and EH, either have low-frequency inharmonicity or, at least, no high-frequency precedent energy in the attack. The two tones on the latter side of the scale, X3 and O2, while quite near the center of the axis, show in their analyses the existence of precedent high-frequency energy. However, in an evaluation of the importance of that initial segment of the attack (Grey and Moorer, 1976) it was found that the discriminability between the existence or nonexistence of that precedent high-frequency energy was 0.57 and 0.58 for the two tones, respectively. This is lower than the average discriminability of the presence of that segment for the tones on the other extreme, which was 0.71 accuracy.

An alternate way of interpreting axes II and III in combination is that they represent the dispersion of family-related clusters, cylindricially about axis I. The projection of this clustering arrangement is seen on the floor of Fig. 1. Two dimensions are taken to represent a nonordered clustering arrangement for three clusters. At this point, there has been no formal modeling of the physical attributes of the stimuli used in interpreting the axes above. Such models will become useful to evaluate the adequacy of the above psychophysical interpretations.

D. Conclusions

In the present multidimensional scaling of timbral similarities between 16 time-variant musical instrument tones, a three-dimensional spatial solution best represented the perceptual relationships. A preliminary psychophysical interpretation of the bases for the similarity judgments found that one dimension related to the *spectral energy distribution*, while the other two related to various temporal patterns of the tones, namely, the presence of *low-amplitude, high-frequency energy* in the initial attack segment and the presence of *synchronicity* in the transients of the higher harmonics along with the closely related level of *spectral fluctuation* in the tone through time. An alternate interpretation of the spatial solution views the cylindrical arrangement of three clusters of stimuli around the spectral energy-distribution axis, grouped on the basis of instrument families.

This study marks the first such scaling of naturalistic, time-variant tones which necessitated a three-dimensional solution for interpretability. Previous studies by other investigators (Wedin and Goude, 1972 as analyzed by Wessel, 1974; and Wessel, 1974) and a pilot study performed at our laboratory used stimuli which differed from the tones employed for this study in three ways: (1) they were *longer* in duration, averaging 1 sec or

more rather than 350 msec; (2) they were *nonprocessed* tones, that is, simply recordings of original tones rather than, as in the present study, computer-synthesized based on an analysis of the real tones; and (3) they were *not equalized* in the dimensions of pitch, loudness or duration.

In previous studies, the spectral energy distribution axis was always found. However, the solutions were only in two dimensions, and the other dimension was difficult to pin down with respect to a specific attribute of tone. Instead, this second dimension was interpreted as having to do with musical instrument *families*, as it showed clustering of brass, woodwind, and string instruments. A physical interpretation could be made only insofar as it related to the complete and unknown set of temporal features defining family membership, rather than to more specific acoustical dimensions. It is hoped that this study has introduced an interpretation of some of the temporal bases for instrument-family clustering, as two time-related aspects of tone appeared to parallel the familylike clustering found in the spatial solution.

The exceptions to the family clustering found in this study may prove to be interesting, namely BN, which clusters with the brass, TP and FH, and FL, which clusters with the strings. Also, from the hierarchical clustering, TM grouped with the oboes, although in the three-dimensional solution it may be seen to join with the brass instruments. Suggested is the possibility that certain physical factors may override the tendency for instruments to cluster by family. The factors suggested here are the articulatory components of tone which occur in the attack. FL is overblown, so it has two similarities to the strings: (1) low-amplitude, high-frequency precedent (inharmonic) energy and (2) nonsynchronicity of onsets along with a high degree of spectral fluctuation through time. BN likewise shares the physical properties of the brass tones with which it clusters, namely, (1) low-amplitude, low-frequency precedent inharmonicity and (2) tapering of onsets of the higher harmonics, leading also to spectral fluctation.

It would appear then that, just as performing instrumentalists can imitate different instruments with varying degrees of success, so might instruments actually cluster in a perceptual similarity mapping. Articulatory features seem to play a central role in this nonfamilial clustering, but it is also important to keep in mind that the selected instrumental ranges and durations have an effect as well. Quite clearly, the use of isolated tonal stimuli is important to make note of, and it will be very important to make similar scalings for timbres that are in musical phrases, where a more complete picture of the instruments involved will emerge.

Finally, it seems necessary to reiterate this investigator's concern for the importance of using computer-synthesized tonal stimuli in experiments on timbre perception. They have the multiple advantages of being capable of equalization in the confounding dimensions of pitch, loudness, and duration, and of being exactly specified with respect to their physical features. Clearly, the potentiality of performing data reductions upon these physical factors is of further importance in obtaining easily and perhaps even uniquely interpretable results.

EXPERIMENT 2: CONFUSIONS IN THE LEARNING OF INSTRUMENT LABELS

A. Stimuli

The stimuli used in this study were identical to those used in Experiment 1, described above, consisting of 16 computer-generated timbres based on the analysis of instrumental notes.

B. Listeners and procedure

There were 22 listeners in this experiment, twenty of whom participated in Experiment 1 as well. All had musical training and were familiar with the instruments of the orchestra. Many were professional orchestral players, and some had considerable experience in conducting. Data collection was spread over a number of short sessions lasting about 10 min apiece, usually occurring before some other experiment. The number of listeners participating in successive sessions decreased almost linearly from 22 to 15 by the fourth session; there were only 12 listeners taking part in a fifth session and 6 in the final session.

Each session consisted of 80 trials. On each trial listeners heard a single tone and then they would be required to identify it choosing freely between the 16 labels provided. They received feedback as to the correct answer. All listeners served in an initial session, where they saw the name of each tone as it was played. In the subsequent sessions, the first 16 trials were considered practice trials, leaving a total of 64 experimental trials. Each of the 16 stimulus tones was presented four times in a random order.

The labels were the same as the abbreviations used in referring to the multidimensional scaling study: O1 and O2 = the two oboes; EH = the English horn; BN = the bassoon; C1 = the E♭ clarinet and C2 = the bass clarinet; X1 and X2 = the two saxophone tones (*mf* and *p*, respectively); X3 = the soprano sax; FL = the flute; TP = the trumpet; FH = the French horn; TM = the muted trombone; S1, S2, and S3 = the cello tones (labelled strings: *sul ponticello*, normal bowing, and muted *sul tasto*, respectively).

Each trial was preceded by a knock, which was followed 1.5 sec later by the tone to be identified. There was a response interval of 4.5 sec before the next warning knock. Listeners were required to identify the tone with one of the labels above, and then to reveal the correct answer by uncovering the next entry on an answer sheet.

C. Results and discussion

A confusion matrix derived from the data is printed as Table I. It consists of the total number of identifica-

TABLE I. Confusion matrix derived from identification/learning study. Stimulus tone labels are rows; responses by listeners are in columns. (22 listeners participating in an average of 4 sessions).

	Response															
Stimulus	O1	O2	EH	BN	C1	C2	X1	X2	X3	FL	TP	FH	TB	S2	S1	S3
O1	173	82	35	4	8	5	10	6	3	-	8	-	6	6	5	2
O2	115	218	24	3	1	-	2	-	-	1	-	-	-	-	-	-
EH	40	38	248	12	-	-	5	3	3	-	1	-	8	2	4	1
BN	1	4	8	305	-	-	-	-	-	-	14	26	9	-	-	-
C1	1	-	-	-	294	60	8	6	-	-	-	-	-	-	-	1
C2	-	-	2	-	77	258	10	12	6	-	1	-	-	-	-	2
X1	1	-	2	3	1	2	229	86	39	-	1	-	3	-	-	-
X2	1	-	2	3	6	8	67	231	39	1	-	1	-	-	-	-
X3	6	9	29	4	3	2	30	42	236	-	1	-	3	-	-	-
FL	-	-	-	-	-	-	-	-	-	358	-	-	-	5	8	1
TP	1	-	5	5	-	-	-	-	-	-	342	4	7	1	-	-
FH	-	-	2	1	-	-	-	-	-	5	7	356	-	-	-	-
TB	3	4	1	-	-	-	1	-	-	1	9	-	346	-	1	-
S2	-	-	-	-	1	-	-	-	-	3	-	-	-	267	74	24
S1	6	2	3	6	1	-	-	-	-	7	2	-	1	57	263	9
S3	-	-	-	1	2	-	-	-	-	1	-	2	-	26	15	320

tions which were made in all trials for all listeners. About 3% of the responses could not be scored, because of insufficient information for identification, and were discarded. The accuracy of judgments did improve with practice over the sessions, rising from an initial 60% correct to 84% by session five.

A large source of the confusions found involved identifying the numeric index of a tone in a subset of similarly labeled notes: S1–S2–S3, O1–O2, C1–C2, and X1–X2–X3. In addition to reflecting an expected increase in perceptual confusions among tones within instrumental classes, as opposed to tones between classes, this distribution of confusions may have been enhanced by a number of procedurally induced difficulties involving the labeling system for the tones. In future experiments of this type, it will be preferable to give stimuli mutually exclusive, perhaps even neutral or nonsuggestive labels.

There were some interesting confusions between instruments of different classes. Two of these confusions are asymmetrical, so that one tone was often erroneously identified as the other, but not the converse. These two cases also have a relationship to the structure of perceptual similarities collected in Experiment 1.

The first case was with X3, identified as EH with about the same frequency of confusion that it had in being identified as either of the other two saxophone tones, that is, a little over eight percent. The group scaling solution shown in Fig. 1 places X3 in the same neighborhood as the three confused tones, about intermediate between them. The second case was with BN, identified as FH about seven percent of the time, and as TP in about four percent of the time. Notice that in the scaling solution, the close relationship of these three tones is also manifest. The only other similarly strong case of confusion was between the oboes and EH, a confusion which occurred in all directions of stimulus-identification, and is also somewhat reflected in the similarity solution.

D. Conclusions

The potentially informative aspect of the asymmetrical confusions is the way in which they may relate to timbral dominance or neutrality for identification. Consider, for example, BN, which was confused with the two brass tones mentioned above, and also rated as being most similar to them in Experiment 1, thereby taking it out of its musical family environment of the woodwinds. This would seem to indicate that certain attributes of this particular tone made it hard to identify as a bassoon. It was taken from the high extreme of the instrumental range, and this certainly must have contributed to its identification with brass tones. It is often noted in musical performances that the high range of the bassoon seems quite similar to a trumpet line in the same range; hence it is probably not the case that the findings of this study are strictly a result of the particular tones utilized. Clearly, *in-context* studies will greatly expand the data base relating to the phenomena of timbral dominance in identification.

ACKNOWLEDGMENTS

We would like to thank John Chowning, Loren Rush, Andy Moorer, Max Mathews, and Dave Wessel for their encouragement and advice in this investigation, and Elizabeth Dreisbach for running the experimental sessions. Also, we are indebted to the Stanford Artificial Intelligence Laboratory, whose computer facilities were used to perform this research.

*This work was supported in part by NEA grant C 50-31-282 and by NSF contracts BNS 75 17715 and DCR 75-00694.

Beauchamp, J. W. (1969). "A computer system for time-variant harmonic analysis and synthesis of musical tones," in *Music by Computers*, edited by H. von Foerster and J. W. Beauchamp (Wiley, New York).

Carroll, J. D., and Chang, J. J. (1970). "Analysis of individual differences in multidimensional scaling via an *N-way* generalization of 'Eckart-Young' decomposition," Psycho-

metrika 35, 283–319.
Chowning, J. M. (1973). "The synthesis of complex audio spectra by means of frequency modulation," J. Audio Eng. Soc. 21, 526–534.
Freedman, M. D. (1967). "Analysis of musical instrument tones," J. Acoust. Soc. Am. 11, 793–806.
Freedman, M. D. (1968). "A method for analysing musical tones," J. Audio Eng. Soc. 16, 119–125.
Grey, J. M. (1975). *An explanation of musical timbre*, Ph. D. dissertation (Dept. of Music Report No. STAN-M-2, Stanford University) (unpublished).
Grey, J. M. and Moorer, J. A. (1976). "A perceptual evaluation of synthetic music instrument tones," (unpublished).
Johnson, S. C. (1967). "Hierarchic clustering schemes," Psychometrika 32, 241–254.
Kruskal, J. B. (1964a). "Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis," Psychometrika 29, 1–27.
Kruskal, J. B. (1964b). "Nonmetric multidimensional scaling: a numerical method," Psychometrika 29, 115–129.
Licklider, J. C. R. (1951). "Basic correlates of the auditory stimulus," in *Handbook of experimental psychology*, edited by S. S. Stevens (Wiley, New York).
Luce, D. A. (1963). *Physical correlates of nonpercussive musical instrument tones*, Ph. D. dissertation (Massachusetts Institute of Technology) (unpublished).
Mathews, M. V. (1969). *The Technology of Computer Music* (M. I. T. Press, Cambridge, Mass).
Miller, J. R., and Carterette, E. C. (1975). "Perceptual space for musical structures," J. Acoust. Soc. Am. 58, 711–720.
Moorer, J. A. (1975). *On the segmentation and analysis of continuous musical sound by digital computer*, Ph. D. dissertation (Stanford University) (unpublished).
Plomp R. (1970). "Timbre as a multidimensional attribute of complex tones," *Frequency Analysis and Periodicity Detection in Hearing*, edited by R. Plomp and G. F. Smoorenburg (A. W. Sijthoff, Leiden).
Risset, J. C. (1966). "Computer study of trumpet tones," Bell Labs., Murray Hill, New Jersey.
Shepard, R. N. (1962a). "The analysis of proximities: multidimensional scaling with an unknown distance function. I," Psychometrika 27, 125–140.
Shepard, R. N. (1962b). "The analysis of proximities: multidimensional scaling with an unknown distance function. II." Psychometrika 27, 219–246.
Shepard, R. N. (1972). "Psychological representation of speech sounds," *Human Communication*, edited by E. E. Davis and P. B. Denes (McGraw–Hill, New York).
Wedin, L., and Goude, G. (1972). "Dimension analysis of the perception of instrumental timbre," Scand. J. Psych. 13, 228–240.
Wessel, D. L. (1973). Report to Psychometric Society Meeting, San Diego (unpublished).
Wessel, D. L. (1974). Report to C. M. E. University of Calif., San Diego (unpublished).