Perceptual and Acoustical Features of Natural and Synthetic Orchestral Instrument Tones

Article *in* Music Perception: an interdisciplinary journal · March 1999

DOI: 10.2307/40285796

3 authors, including:

John M. Hajda
University of California, Santa Barbara
10 PUBLICATIONS **230** CITATIONS

Perceptual and Acoustical Features of Natural and Synthetic Orchestral Instrument Tones
Author(s): Roger A. Kendall, Edward C. Carterette and John M. Hajda
Source: *Music Perception: An Interdisciplinary Journal*, Vol. 16, No. 3 (Spring, 1999), pp. 327-363
Published by: University of California Press
Stable URL: http://www.jstor.org/stable/40285796
Accessed: 06-08-2015 20:46 UTC

Music Perception
Spring 1999, Vol. 16, No. 3, 327–364

Perceptual and Acoustical Features of Natural and Synthetic Orchestral Instrument Tones

ROGER A. KENDALL, EDWARD C. CARTERETTE, & JOHN M. HAJDA
University of California, Los Angeles

Four experiments were conducted to explore the timbres of natural, continuant orchestral instruments with emulation based on sampling, frequency modulation (FM) synthesis, and a hybrid consisting of sampling and synthesis techniques combined. Identification of instruments using verbal labels was significantly better for the natural and sampling-based signals than for either FM synthesis or the hybrid technique, a result also found for aural categorization. Perceptual scaling of timbral similarities demonstrated great consistency across a series of independent variables, including musical training, monophonic and stereo presentation, and long versus short signal durations. The first dimension of the classical multidimensional scaling (CMDS) solutions mapped onto long-time-average spectral centroid. The second dimension mapped onto a measure of spectral variability. Little evidence was found for the mapping of attack time or signal duration onto either dimension. A third dimension separated most natural instruments from their emulated counterparts. Experiments using verbal attribute ratings confirmed the correlation of spectral centroid, the first dimension of the perceptual space, and ratings of *nasality*; the second dimension correlated with spectral variability and modestly correlated with ratings of *rich*, *brilliant*, and *tremulous*. Mismatches of spectral distribution and variability resulted in poor emulations of the natural instruments. Results suggest that further study of centroid and time-variant psychophysical properties is warranted.

OUR purpose here is to understand the relationships among natural, acoustic instrument timbres and their synthetic counterparts. The major research questions that condition our study are (1) What is the relationship of natural instrument timbres to timbres reproduced by various means of digital synthesis? (2) How do convergent experimental methods explicate different facets of perceived timbral differences? (3) What acoustical properties map to any differences among these timbres? There is a

Address correspondence to Roger A. Kendall, Music Perception and Acoustics Laboratory (MPAL), Program in Systematic Musicology, Schoenberg Hall, University of California, Los Angeles, CA 90095-1657. (e-mail: kendall@ucla.edu)

long history of the use of naturalistic sound in electronic music, from *musique concrète* to the present near-ubiquitous use of orchestral instrument emulation in the commercial sphere. In addition, timbres from synthesizers have been used in several timbre studies (e.g., Iverson & Krumhansl, 1993; Krumhansl, 1989; McAdams, Winsberg, Donnadieu, De Soete, & Krimphoff, 1995; Wessel, Bristow, & Settel, 1987). Therefore it is timely to consider in detail the relationships, both perceptually and acoustically, among natural and synthetic stimuli.

In order to explicate the relationships among the natural and synthetic instruments, we used a series of convergent methods. What we learn about a question depends upon what we ask and how we operationalize the search for the answer. We can expect that certain timbral attributes may be stable across methods, and others may vary. This approach can give us clues about the stable features. In addition, as we shall show, our approach provides a degree of replication that leads to high reliability and consistency of the results.

For this study, we selected four methods, one of which is central in each of the four main experiments that follow. We then proceed to link perceptual and acoustical frames of reference by means of correlative and mathematical and physical analyses of the stimuli and their relationships to the response data arising from several experiments. Inasmuch as there are a number of independent variables beyond timbre, including stereo versus mono, musical training, and tonal duration, we were able to lay to rest a number of questions about context.

Experiment 1 employs *name categorization* (i.e., *identification*, see later). This requires the assignment of a label to a timbre and so demands that verbal and musical schemata be linked. Experiment 2 involves *aural categorization*,[1] that is to say, the grouping of timbres according to a model. This method does not require a subject to know instrument names, in contrast to identification. In the multifaceted Experiment 3, we explore the similarities among timbres of the different natural and synthetic generators. Similarity scaling requires a subject to make direct comparative judgments of all possible pairs of timbres; this does not require a subject to make decisions of grouping or to associate a name with an instrument timbre. Experiment 4 extends our work with verbal attributes of timbral dyads (Kendall & Carterette, 1993a, 1993b) to the natural single notes used as the canonical group in the other experiments. We connect the verbal attributes to the perceptual spaces and acoustical analyses of Experiment 3, and we use the attributes in interpreting the data of the other experiments.

1. Our approach to categorization is best related to "prototype theory" (see Ashby and Maddox, 1998, pp. 252–293).

Motivations for the Study

Recent implementations of digital technology have led to a remarkable array of synthesizers, most often in the form of "keyboards," in which the techniques for generating sound are many and varied. In particular, new timbral palettes and possibilities were explored in the realm of electronic art music as exemplified by the efforts of Vladimir Ussachevsky, Lejaren Hiller, and Morton Subotnick. At the same time, commercial efforts were focused on the simulation of natural instruments motivated by economic pressures to minimize labor and cost. An early emulation of natural instruments was the Mellotron, a device that used tape loops of individual instrument tones actuated by a keyboard (the instrument was under discussion for banning by The Musicians' Union).

A clear distinction between many early models of synthesizers and natural instruments was the degree of variability, both in the signals produced and in the degrees of freedom of performance. Indeed, the history of evolution of electronic musical instruments has been connected to a relatively uncontrolled search for the enhancement of variability, often in the absence of a musically based empirical model or theory. We enumerate several issues connected with the mappings among electronic instruments and their natural counterparts:

1. Loudspeakers are not soundboards, resonance chambers, or oscillating air columns; musical instruments should be conceived of as a generator coupled to the environment in a way that is quite different from that of a loudspeaker. This issue is connected to understanding the relationship between actual and recorded performances.
2. Piano keyboards are rather like switching devices; mapping a string/fingerboard onto a keyboard leads to performance mismatching because the degrees of freedom are different. Evidence for difficulties such as this can be found in the existence of such controls as wheels, sliders, toggles, and pressure and velocity mechanisms.
3. An important issue is the capacity of human information processing in performance on musical instruments. There is a complex multidimensional interaction among performer, musical instrument, the environment, and the music itself that is intended to be communicated. The musician is part of a dynamic feedback system. The successful musician possesses many schemas, for example, motor, sound, pattern, and meaning, all of which are implicated in the feedback loop. It is far-fetched to believe that a keyboard artist can have sufficient cognitive capacity to

switch among multiple sets of these schemas, to say nothing of doing it on continuous demand. Playing a keyboard as a 'clarinet' and immediately switching to 'violin' for example would at least stretch cognitive limits.

Musical validity in timbre research is a difficult problem because so many nonorthogonal variables are involved. As a beginning, our approach leaves the character of standard musical instruments and synthesizers intact and comparatively examines the timbres of single notes. We hypothesize that differences between natural, orchestral instrument timbres and their emulation can be explained in terms of measures of spectral distribution and variability. We also hypothesize a consistency among timbre relations that holds across subject tasks. Our experimental design is a set of essentially parallel, not consecutive, experiments; one experiment does not lead to the next.

The present study was motivated by our recent research into simultaneously sounding orchestral wind instruments (Kendall & Carterette, 1989, 1991, 1993a–1993c). We found an interesting space of these dyads, such that they spanned a two-dimensional perceptual space that was circumplicial. Treating points of the paired instruments as vectors allowed us to apply standard vector algebra to create hypothetical spaces of single instruments and to combine them in various ways. It is clear to us from our earlier work that the evolution of the palette of orchestral winds is not accidental; it has been shaped to yield a perceptual space that is bounded by the extremes of *nasality* (brightness) on the one hand and *richness* and *brilliance* on the other. The principal dimension of musical timbre for continuant signals appears to be *nasality*, which we believe is similar to what others call *brightness* (Lichte, 1941; Plomp & Steeneken, 1969) or *sharpness* (von Bismarck, 1974a, 1974b). Jakobson, Fant, and Halle (1951) also studied the attribute *acute* in terms of phonetics, which is probably strongly correlated with *nasality*.

The principal physical attribute of timbre that maps onto these perceptual attributes is spectral *centroid*, defined as the midpoint of the spectral energy distribution (see below). Correlations between this measure and the *nasality* or *brightness* timbre dimensions are in the range of about .90 in a number of studies, for example, Grey and Gordon (1978), Krimphoff et al. (1994), and Donnadieu et al. (1996). These high correlations hold also for combinations of instruments, including relations with blend (Kendall & Carterette, 1989, 1993c; Sandell, 1995), such that pairings of instruments with disparate centroids blend less well than pairings of instruments with similar centroids.

We noticed, however, differences between our analyses and those of others in the literature and were driven to reexamine single instrument timbres

using insights from our work. Much work has been done using only a few synthesis methods, which have become nearly canonical, such as line-segment resynthesis (e.g., Grey, 1975) and frequency modulation (FM) synthesis (e.g., Krumhansl, 1989). In addition, the advent of new synthesis techniques, especially sampling and hybrid techniques, demands a fresh study comparing natural instrument tones with others. We make it clear from the outset that our goal is not to determine whether one timbre is good or bad or one timbre is better than another, but rather to understand what makes timbres distinct from each other when they share many common features. We will, however, use the natural set for comparisons with emulated counterparts. In effect, the natural signals will become the canonical set.

General Instrumentation

Kendall and Carterette (1989, 1993b, 1993c) have reported psychoacoustical relationships between measures of spectral distribution and the principal dimension of perceptual spaces of instrument dyads. We suggested that attack time correlates with a primary perceptual dimension for orchestral instrument timbres because of the inclusion of both impulse and continuant instrument envelopes in the same stimulus set (Kendall, Carterette, & Hajda, 1994, 1995). The multidimensional scaling (MDS) solutions for studies with such hybrid instrument populations clearly show a near categorical separation of impulse and continuant instruments along the first dimensions (e.g., Iverson & Krumhansl, 1993; McAdams et al., 1995, p. 185).[2]

In exploratory research, we conducted a perceptual scaling that included all of the natural timbres used in the present experiments along with a set of impulse instrument tones: tubular bell, xylophone, marimba, classical guitar, pizzicato violin, and piano (these impulse signals were the subject of experiments reported in Hajda, 1995). The resulting similarity matrix was subjected to classical MDS (see Schiffman, Reynolds, & Young, 1981) and hierarchical clustering analysis. The two-dimensional perceptual space could be cleanly split by a vertical line perpendicular to the first dimension, thus separating continuant from impulse timbres. This grouping of instrument classes was also confirmed even more strongly by cluster analysis. In addition, Hajda (1995) found difficulty in mapping long-time-average centroid to impulse signals, in contrast to continuant signals. We therefore concentrated in these experiments on a homogeneous set of natural, continuant tones.

2. For a critical, comparative study of the relationships among MDS studies that focuses on methodological issues, see Hajda, Kendall, Carterette, and Harshberger (1997).

We chose orchestral instruments capable of easily sounding $B\flat_4$ (ca. 466 Hz)[3]: flute, oboe, clarinet, soprano saxophone, alto saxophone, tenor saxophone, French horn, English horn, violin, bassoon, and trumpet. Three classes of keyboards/sound modules: (1) FM, (2) hybrid synthesis/sampling, and (3) processed sampling were represented by appropriate factory presets or standard ROM card voicings on the Yamaha DX7, Roland D-50, and E-mu Emax II and Proteus/2, respectively. All modules were driven from a single master keyboard. Whenever possible, the factory presets or ROM card voicings were used (Appendixes A and B). Note that not all of the 11 instruments used in the natural set were available on a given emulator used in this study. The voicings and setup were configured by a professional Hollywood synthesist. Signals were recorded in a studio using direct electrical connection to a Sony digital audio interface (PCM 601 esD) with 14-bit linear quantization at 44.056 k samples per second.

Preparation of Stimuli

In order to approximate the recording conditions for the natural instruments, the digital tape was played on stage in a configuration similar to that used for live performances of electronic music. Recordings were made in a moderately reverberant concert hall (ca. 1.6-s reverberation time). The transducers were two Infinity 5001 3-way speakers spaced 132 cm apart and situated 167.6 cm from the microphone; the angle of speaker center to microphone was 50 degrees. The microphone, which was centered, was an AKG C-S42 coincident stereo condenser microphone operated in M-S configuration by using the AKG S42 matrix controller. The M and S microphone pick-up patterns were set for an orthogonal figure eight. Natural instruments were recorded with the same microphone configuration. Microphone recordings were made using a 16-bit Sony PCM F1 digital audio interface in stereo. Recording levels and the geometry for instrument/microphone positions were set by a professional recording engineer to maximize quality and perform a first-order equalization of loudness (later refined by a psychophysical procedure). Recording sessions were monitored for pitch accuracy by a musicologist (one of the authors, R. A. K.), and feedback was provided during multiple takes until a uniformity was achieved.

3. Our experiments (Kendall & Carterette, 1991, 1993a–1993c) have used $B\flat_4$, which is a moderate tessitura for the majority of instruments we employed. We do not use $E\flat_4$ (ca 311 Hz), which has been a standard for the studies of Grey (1975), Krumhansl (1989), McAdams et al. (1995), and others. Except for Grey (1975), those researchers used purely synthetic tones not requiring an instrumentalist. $E\flat_4$ is a characteristically poor note for the majority of soprano winds. We would note, however, that the bassoon is in a very uncomfortably high register in our study. We had an excellent professional bassoonist, however. We eliminated other tenor/bass instruments from our study.

Eleven natural continuant instruments were recorded: B♭ soprano clarinet, bassoon, English horn, flute, French horn, oboe, B♭ trumpet, violin, E♭ alto saxophone, B♭ soprano saxophone, and B♭ tenor saxophone. All performers were music majors studying their instrument at UCLA, except for the bassoonist, who was a member of the UCLA music faculty and the Los Angeles Philharmonic.

Many different scale patterns, melodies, and articulations were recorded. For the experiments of the present study, we use only the $B\flat_4$ single notes of 500-ms and 3000-ms performed duration. Actual durations in the reverberant environment were measured from the digitized signals and had a mean length of 4.35 s for the 3000-ms signals, which corresponds reasonably well to the estimated hall reverberation time. Similarly, the 500-ms signals had a mean length of 1.75 s.

Signals were resampled from digital tape to the hard disk of an IBM class PC computer at 25,000 samples per second per channel. A computer program handled all facets of the experiments, including randomization and data collection. All experimental stimuli were equalized for loudness. Four musicians adjusted each stimulus relative to a standard, and the average adjustments were used to set playback levels using a digitally controlled attenuator of custom design. The computer equipment provided 16-bit linear recording and reproduction with appropriate anti-aliasing filters.

Here we clarify an issue regarding the instrument variable. Several instruments included in the larger, natural set did not have counterparts across all emulators (soprano saxophone, alto saxophone, bassoon, English horn). Seven of the instruments (clarinet, flute, French horn, oboe, tenor saxophone, trumpet, and violin) were in common among all generators. In our analyses, we generally compare the data generated from the larger set of instruments with the data generated from the subset of seven in-common instruments to analyze whether a context effect exists.

Experiment 1: Identification

Identification in timbre research explores the degree to which subjects can attach a label to a sound. Actually, many of the studies called "identification" do not fit strictly the formal definition that the number of stimuli and response items be equal. Although one might use the phrase "name categorization" for the situation where the number of stimuli is greater than the number of names, as in the present experiment, we will adhere to common practice and use the term "identification" for brevity. Historically, such tasks have involved three basic procedures (Hajda et al., 1997, p. 262). The first type is free identification, in which subjects simply write down the instrument they hear (e.g., Wedin & Goude, 1972). In the second type, subjects provide a forced-choice response by using a list of instru-

ments that includes distracters (Saldanha & Corso, 1964). The third type is forced-choice, but the list does not include distracters (e.g. Clark, Robertson, & Luce, 1964), a method that adheres to the classical definition of identification.

It is possible to hypothesize different cognitive operations relative to subject task among identification, sound categorization (Experiment 2, below) and aural matching (for timbre, see Kendall, 1986). Name categorization requires a subject to have explicit knowledge about the relationship between sound and word, a higher cognitive operation requiring some musical expertise. A subject must have learned the meaning of the word "clarinet," for example. In contrast, with aural categorization and matching the task is nonverbal.

Previous experiments in timbre identification demonstrate the influence of methodological variables on the accuracy of responses obtained. In Clark, Luce, Abrams, Schlossberg, and Rome (1963), with only 2 distracters in the response list, instruments were identified with 70% accuracy (other details of their experiment were not presented). In contrast, the experiments of Saldanha and Corso (1964), which used 10 orchestral instruments at three pitches and a response list with 29 distracters, yielded only 41% correct identification. For five disparate studies analyzed by Hajda et al. (1997, Table 12.1), the natural, unaltered signals were identified an average of 55.6% of the time.

Identification studies have been widely used to compare natural, unaltered timbres with degraded signal conditions, such as missing attacks, decays, and all manner of signal editing. In the present study, our independent variables are the instruments, and we look here to compare the identification accuracy of recorded natural instruments with their synthesized counterparts.

METHOD

Subjects

Subjects were college-age musicians (N = 16) who responded to an advertisement for subjects; they were paid $10 to participate in the experiment. Because the identification task required knowledge of instrument labels, subjects were required to take a pretest. We required 100% accuracy in selecting instrument labels for natural trumpet, clarinet, oboe, flute, and alto saxophone using a subset of the loudness-equalized long steady-state tones. Eight subjects were randomly assigned to respond to short continuant tones (ca. 500 ms) and eight to the long continuant tones (ca. 3000 ms).

Stimuli and Procedure

The monophonic stimuli for the entire set of 11 natural instruments and available emulations (see Appendix B, a total of 36) were sequenced by computer program and played back using a laboratory-grade amplifier diotically through Sennheiser HD-222 headphones. The voltage output level from the amplifier was fixed relative to an arbitrary standard tone

(oboe) at a comfortable listening level near 68 dB SPL. Each stimulus was auditioned once by a given subject. The subject selected an instrument name from a word list, and the next stimulus was presented automatically. This procedure was repeated until the random presentation of the entire stimulus set was complete.

RESULTS

We scored answers correct if the label selected corresponded to the intended emulation, that is, if the label "flute" was selected when E-mu flute was sounded. The experimental design was a mixed, within/between model with repeated measures. The between-groups factor was length (short vs. long). One within-subjects factor was emulator (natural, E-mu, Roland, Yamaha); the other was instrument.

Figure 1 is a bar graph of the identification score data for all 36 stimuli ($N = 16$). The more standard, familiar orchestral instruments, like clarinet, trumpet, violin, French horn, oboe, and flute, are more accurately identified, particularly for natural versions. More poorly identified instruments comprise the less common orchestral instruments, like English horn, bassoon, and saxophones. The saxophones were identified in the context of other members of the family, and so intraclass confusions arose, which would be apparent from the confusion matrix (too complex to present here).

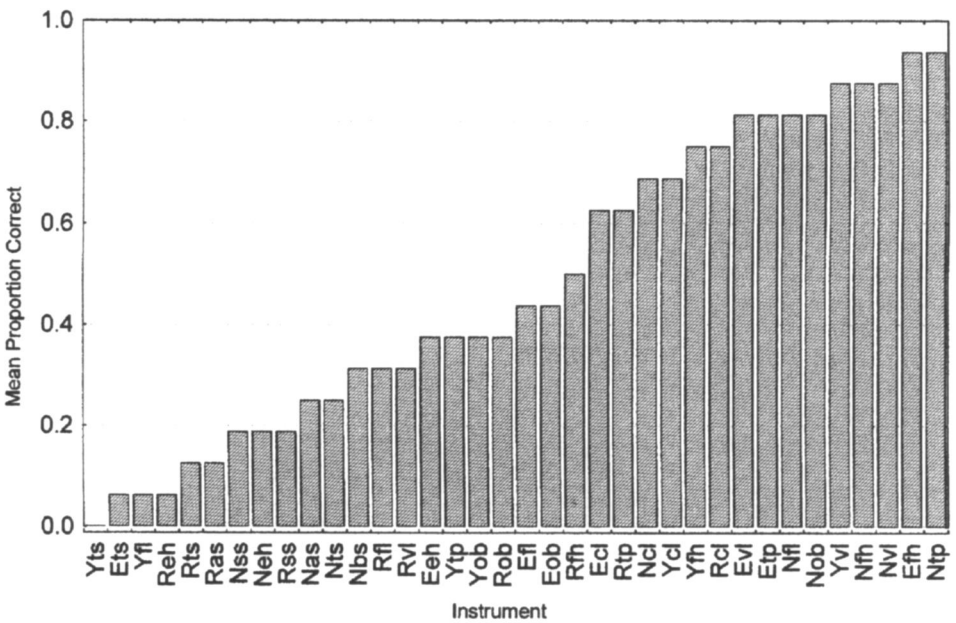

Fig. 1. Bar graph of identification score data ($N = 16$). N = natural, E = E-mu, R = Roland, Y = Yamaha, as = alto sax, bs = bassoon, cl = clarinet, eh = English horn, fh = French horn, fl = flute, ob = oboe, ss = soprano sax, tp = trumpet, ts = tenor sax, vl = violin.

We turn to inferential statistics to probe these differences in identification accuracy and present the analyses of variance (ANOVAs) for the seven in-common instruments. The instrument factor was collapsed and included the data for the main effects of emulator and length. There was no difference between ANOVA results for the larger instrument sets and the subset of the seven in-common instruments.

An ANOVA based on scored data showed that the main effect for length was not significant, $F(1, 14) = .126$, $p > .728$. This result indicates that a 500-ms performed duration in a reverberant environment is as easily identified as a 3000-ms performed tone. This is a unique finding because the identification literature has ignored performed duration as a variable.

The main effect for emulator was significant, $F(3, 42) = 19.97$, $p < .0000009$. The mean correct name categorization ranges from 75% for natural, to 58% for E-mu, to about 43% for Roland and 44.5% for Yamaha. Tukey's honestly significant difference (HSD) post-hoc analyses demonstrate that the natural means are significantly higher ($p < .05$ for all tests) than the other three emulators, that E-mu has higher identification accuracy than either Roland or Yamaha, and that Roland and Yamaha have equally low identification accuracy. It is clear that the greater the degree of naturalness (as an omnibus variable), the greater the ease of selecting the correct instrument name. Sampled instrument tones, such as the E-mu, are more easily categorized than either hybrid or FM synthesis.

The instrument main effect was significant, $F(6, 84) = 18.246$, $p < .0000009$. Tukey's HSD shows that the means for clarinet, French horn, trumpet and violin, across emulators, are not different, but are significantly higher than flute, oboe, and tenor saxophone. Flute and oboe are not different but are higher than the mean for tenor saxophone. We find, therefore, that flute and oboe are surprisingly poorly emulated; recall that correct naming of these natural instruments was prerequisite to participation in this experiment. Tenor saxophone, once again, is probably the victim of family distracters (but see Experiment 2, below), although for Yamaha and E-mu there is a clear confusion outside the members of the saxophone family.

The interaction of instrument and emulator, thus, is significant, $F(18, 252) = 2.942$, $p < .000088$. Figure 2 shows this interaction. What emerges as the source of the interaction, tested with Tukey's HSD, is a substantial difference in means between natural flute and the emulators and the poor identification of the Roland violin compared with excellent identification of the natural and the other two emulators. The Yamaha flute is mistaken for French horn by nearly half the subjects, a confusion that will reappear in other experiments.

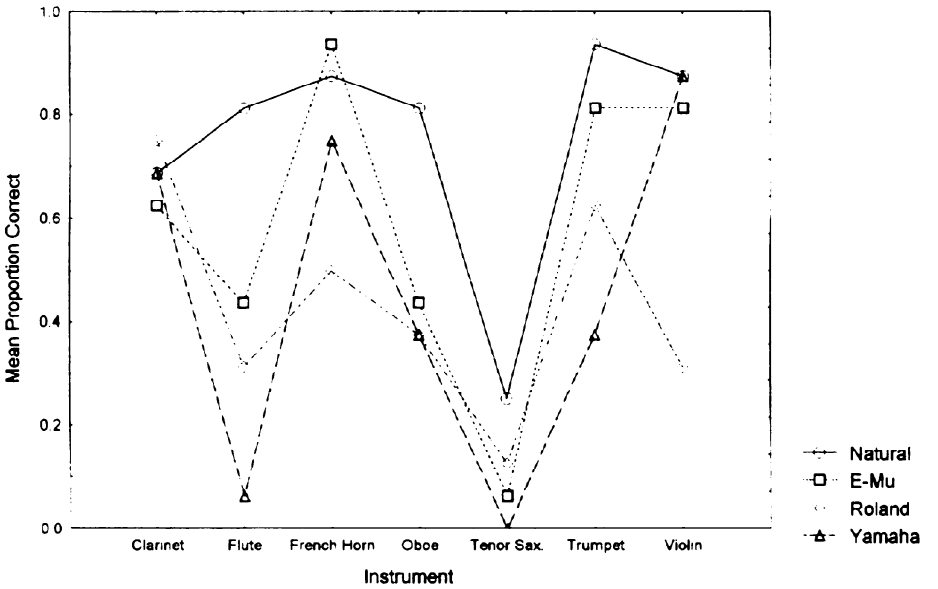

Fig. 2. Graph of the two-way interaction ($N = 16$) between instrument (x axis) and emulator (see legend) for identification score means (y axis).

Experiment 2: Aural Categorization

Over the years, we have developed a unique experimental methodology for auditory classification without the use of verbal labels. A more extensive discussion of this technique can be found elsewhere (Carterette & Kendall, 1996; Kendall & Carterette, 1990, 1992). In the categorization procedure, groups are primed by model stimuli chosen as independent variables. Choice stimuli are grouped under the primes using whatever strategy the subject desires according to the most salient characteristic. The approach is an analog to categorization experiments done with pictures and objects.

Aural categorization is different from identification (name categorization) in that knowledge of the verbal or pictorial attributes of the musical sound is not required; direct sense impressions can be used. Second, the primed groups establish a range similar to a closed list in identification studies. Third, the subject can review choices among categories and make changes in categorization, an ability that is not present in most identification procedures. We would expect mean scores in categorization to go up relative to identification. However, because different perceptual strategies are involved, it is not clear whether interactions among factors and levels

would be the same, or whether errors of categorization would be made to the same instruments or emulators as with identification.

METHOD

Subjects

All subjects were college-level trained musicians. The stimuli were a subset of those used in the identification experiment; only the seven in-common instruments were used (clarinet, flute, French horn, oboe, tenor saxophone, trumpet, and violin) because the computer display limited categories to a maximum of 10. Thirteen subjects responded to long stimuli and 14 responded to short stimuli, for a total of 27 subjects. Because no significant differences were found between long and short results, the data were combined.

Stimuli and Procedure

The priming stimuli were the seven natural instrument signals, which served as models. Twenty-eight randomly ordered choices were placed at the bottom of the screen (7 instruments x 4 emulators, including the natural identities). The subject's task was to move a colored rectangle representing the choice stimulus underneath the best fit for a model. The experiment concluded when all 28 choices were thus categorized. Subjects could review the model and choice stimuli at will and could make changes in their groupings until they were satisfied.

The experimental apparatus and stimulus conditions were the same as those used for the previous experiment. Stimuli were presented diotically by headphones (Sennheiser HD 220).

RESULTS

An answer was scored as correct if the choice instrument was grouped with its model. The factorial design of this experiment is identical to that of identification (Experiment 1). We combine data for the seven common instruments under identification and categorization to test these effects, with appropriate adjustment of alpha.[4]

The response accuracy was significantly higher for categorization (mean = .72) than for identification (mean = .55), $F(1, 39) = 19.69$, $p < .0002$. There is a significant interaction between method and instrument; however, the interaction is largely due to a mean difference between tenor saxophone in identification (.11) and categorization (.59). Within categorization, however, the pattern of significant effects mirrors that of identification. There was again a significant main effect for emulator, $F(3, 75) = 52.45$; $p < .00018$. Tukey's HSD shows, like identification, that the natural mean is higher (.97) than E-mu (.74), and both are higher than either Roland (.58) and Yamaha (.57), which are not significantly different.

The same pattern of mean differences is found for instruments under aural categorization, $F(6, 150) = .984$; $p < .00018$. Therefore, the flute remains very poorly categorized overall. From the data, we can conclude

4. The combined identification and aural categorization data sets are later reseparated to avoid the loss of discrimination owing to unequal *N*s.

that the tenor saxophone's poor showing cannot be attributed to method or its inclusion in the family of saxophones under identification, because it fairs relatively poorly in the absence of family members under aural categorization.

Figure 3 shows the significant interaction of instrument and emulator, $F(18, 450) = 6.64$, $p < .0005$. Tukey's HSD demonstrates a near-perfect mean correct categorization for natural instruments with no significant differences among them. Once again, flute and tenor saxophone mean categorization is significantly lower for the synthesizers. However, the tenor saxophone natural signals are well-categorized, a contrast with identification. Here, then, may lie the influence of the distracter list in identification; it affects the *identification* of all emulators for tenor saxophone. However, under *categorization*, emulators for tenor saxophone other than the naturals do not fare well. This finding should caution interpretation of research using word lists requiring differentiation among members of a family. The correlation of categorization and identification results is .654.

Experiment 3: Similarity Scaling

Proximity rating or similarity scaling measures the degree of similarity or dissimilarity among all pairs of a set of objects and has proved to be a

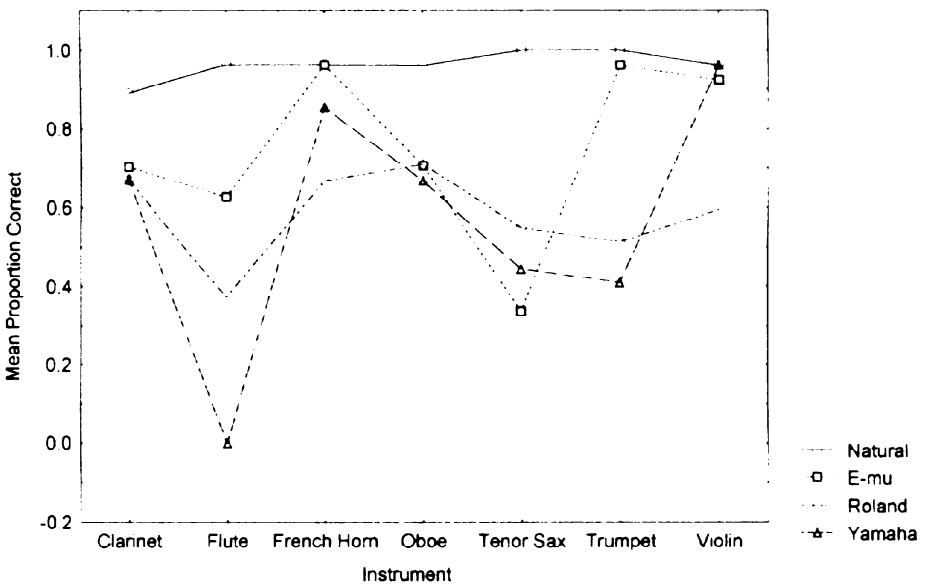

Fig. 3. Graph of the two-way interaction ($N = 27$) between instrument (x axis) and emulator (see legend) for categorization score means (y axis).

useful approach to explicating relationships among timbres. During the past 25 years, the method has been used with considerable success in a number of timbre studies, for example by Plomp (1970), Wedin and Goude (1972), Wessel (1973), Miller and Carterette (1975), Grey (1975), and more recently by Krumhansl (1989), Kendall and Carterette (1991), and McAdams et al. (1995). We have analyzed the data of the aforementioned and other workers (see Hajda et al., 1997, pp. 284–295) as well as our own studies of dyads (Kendall & Carterette, 1991) and conclude that, if the stimulus set is homogeneous with respect to envelope and generator (continuants without impulse), there is a consistency of the interpretation of the principal dimensions. We would note, however, that many of these studies have used "naturalistic" timbres; by this the researchers mean emulation, for example, FM by Krumhansl (1989) and McAdams et al. (1995), resynthesis by Grey (1975, 1977), and synthesis by Plomp (1970) and by Miller and Carterette (1975).

We turn to the method of proximity (Kruskal & Wish, 1978), which has not been used before to compare natural timbres with their naturalistic counterparts. Similarity may exist along a number of dimensions, yet the subject appears to be able to incorporate the amalgam in a single judgment of proximity. It should be interesting to compare the results of similarity scaling with results of our previous methods in order to converge on properties relevant to timbral naturalness.

METHOD

Procedure

The experimental design had two phases. The subject counts (N) that we give for both phases are subjects actually included in our analysis. A small number of outliers (fewer than 10%) were excluded from the analysis by means of an a priori, quantificational procedure.

In Phase 1, we scaled the natural instruments in four blocks to ascertain influences of musical training (music major and nonmusic major) or playback (monophonic and stereophonic). There were four groups: (1) nonmusicians and mono (N = 10), (2) nonmusicians and stereo (N = 10), (3) musicians and mono (N = 10), and (4) musicians and stereo (N = 10). The results of Phase 1 (see later) led us to drop training and playback as variables in subsequent experimentation. Therefore, in Phase 2, we investigated relations among natural and synthetic timbres independent of training and playback. Phase 2 subjects were randomly assigned to one of the following stimulus contexts: (1) E-mu alone, (2) Roland alone, (3) Yamaha alone, (4) E-mu plus natural, (5) Roland plus natural, and (6) Yamaha plus natural. The total number of subjects for these contexts was 60.

In both phases, stimuli were presented randomly and responses were recorded by computer. The stimuli were all possible pairs of 3-s sounds with 500 ms between pairs. On the computer screen, a 127-mm bar appeared with the label "similar" at one end and the label "dissimilar" at the other end. On each trial, a pair of sounds was heard and a scale with a randomly placed arrow was seen. The subject's task was to set the arrow on the line to indicate the similarity between the two sounds just heard. The computer translated the

relative distance of the arrow to a scale of 0 to 99, where 0 was highly similar (identical) and 99 was highly dissimilar.

Subjects

Musicians were those in or having graduated from a college music program or participants in the Los Angeles Philharmonic Summer Program. Musicians were paid for their participation. Nonmusicians were drawn from the subject pool in the psychology department at UCLA. These subjects were given partial academic credit for their participation. Each subject was run individually and separately on the scaling task.

RESULTS

Phase 1: Analyses of Experimental Condition Effects

Data from the four Phase 1 experimental conditions involving musicians versus nonmusicians, stereo versus mono, were subjected to classical multidimensional scaling (CMDS). In addition to these four matrices, the natural-instrument data were extracted from the three Phase 2 stimulus contexts that included natural instruments. Pearson correlation was performed across these seven natural-instrument matrices (Table 1). The average correlation among matrices (N = 10 x 7 = 70) across all conditions was .836 and the range was .13.

Individual Differences Scaling (INDSCAL, Carroll & Chang, 1970) was conducted with the seven natural-instrument matrices entered as subjects. The resulting plot had all points within the upper right-hand quadrant, nearly on top of each other. Therefore, in view of the high consistency and strong relations among matrices, the data were averaged over contexts involving musicians versus nonmusicians and stereo versus mono so that all further analyses were made on the four stimulus contexts: natural, E-mu, Roland, and Yamaha. There were thus four comparisons: The natural in-

TABLE 1
Correlations of Stimulus Conditions

	Nonmusician		Musician		Natural		
	Mono	Stereo	Mono	Stereo	Yamaha Context	Roland Context	E-mu Context
Nonmusician mono	1.00						
Nonmusician stereo	.82	1.00					
Musician mono	.82	.83	1.00				
Musician stereo	.76	.81	.81	1.00			
Yamaha	.87	.83	.86	.78	1.00		
Roland	.89	.87	.87	.79	.89	1.00	
E-mu	.85	.83	.89	.79	.85	.85	1.00

struments alone as well as the natural instruments compared with each of the other three emulators.[5]

Phase 2: Similarities

Figures 4, 5, 6, and 8 show the two-dimensional CMDS results for natural alone and natural combined with E-mu, Roland, and Yamaha respectively. We used classical MDS (Kruskal, 1964a, 1964b) because we desired to rotate these configurations onto acoustical analysis results (see later).[6] Variance accounted for by the two-dimensional solutions is 95%, 84%, 81%, and 84% for natural alone and natural combinations with E-mu, Roland, and Yamaha, respectively.

We have interpreted perceptual spaces of natural timbres in terms of *nasality*, *reediness*, *richness*, and *brilliance* (Kendall & Carterette, 1993a, 1993b). We can apply the same descriptions in interpreting the present spaces (we return to verbal attributes relative to the *present* stimuli in a separate experiment, and we connect all of this to spectral variables in the next section). Dimension 1 can be interpreted as the *nasal* (–) versus *not nasal* (+) axis; dimension 2 as the *not brilliant* (+) versus *brilliant* (–) axis. We hypothesized, on the basis of our previous work with dyads, that *reediness* increases for instruments in the upper left quadrant (*not brilliant* combined with *nasality*), and *richness* increases for instruments in the upper right quadrant (*not nasal* combined with *not brilliant*). However, Experiment 4 (see later) provided little support for this idea.

In these terms, the natural instruments (Figure 4) range in *nasality* from the violin and oboe, which are *nasal*, to the French horn and alto saxophone, which are *not nasal*. The trumpet is *brilliant* and the tenor saxophone is maximally *not brilliant*. Therefore, the soprano saxophone is somewhat *brilliant* and *nasal*. The English horn is somewhat *nasal* and moderately *not brilliant*. To a musician, these interpretations make perfect sense. For example, in a school-band score, one might find a substitution of soprano saxophone for English horn (this is called a cross-cue), which would retain the *nasality* at the expense of a little too much *brilliance*. Alto saxophone is often cross-cued as a substitute for the French horn.

Please note that the correlation of the 11 naturals across all conditions was relatively high (see preceding paragraphs). We have, for the first time, a consistent set of scalings across seven subject groups ($N = 70$).

5. There is no space here to include the data and results for the emulators alone. However, the results did not differ substantially from scalings that included the natural-instrument set.

6. For a discussion of rotation for mapping of nonperceptual data onto a perceptual space, see Kruskal and Wish (1978). For an application of this technique to timbre spaces, see Grey and Gordon (1978).

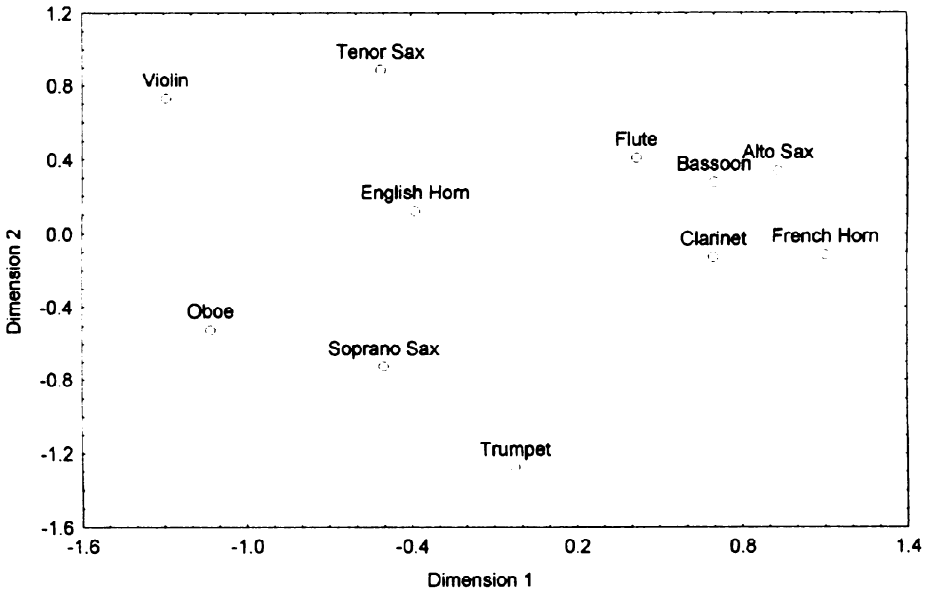

Fig. 4. Two-dimensional classical multidimensional scaling solution (R^2 = .95, N = 70) for natural instruments across training, playback, and emulator context.

Figure 5 shows the results for the E-mu plus natural condition. The emulation of natural instruments by this generator is quite variable. For example, the E-mu tenor saxophone is too *nasal* and *brilliant* to place it in proximity to its natural counterpart. The same is true of the E-mu oboe; the E-mu flute is too *brilliant* and too *nasal* as well. In general, E-mu emulations of brass instruments (trumpet and French horn) are less nasal than required, and of wind instruments, with the exception of the oboe, are more *nasal* than their natural counterparts, and often too *brilliant* as well. The E-mu violin is too *nasal* and somewhat too *brilliant*.

Figure 6 shows the results for the natural plus Roland condition. In general, Roland timbres are more *nasal* than their natural counterparts. Note that many Roland timbres cluster in the lower left quadrant; they are too *nasal* and too *brilliant*. Because the Roland timbres were poorly identified and categorized, we get a clue regarding the interpretation of a higher dimensional solution to the similarity data. Figure 7 shows the three-dimensional solution for natural plus Roland, which accounts for 12% more of the total variance. The third dimension most often separates natural from Roland timbres whereas in the two-dimensional solution (Figure 6) such natural-Roland clustering is less evident. In every case except the tenor saxophone, which is poorly identified in Experiment 1, the natural instrument is higher on Dimension 3 than its Roland counterpart. We interpret

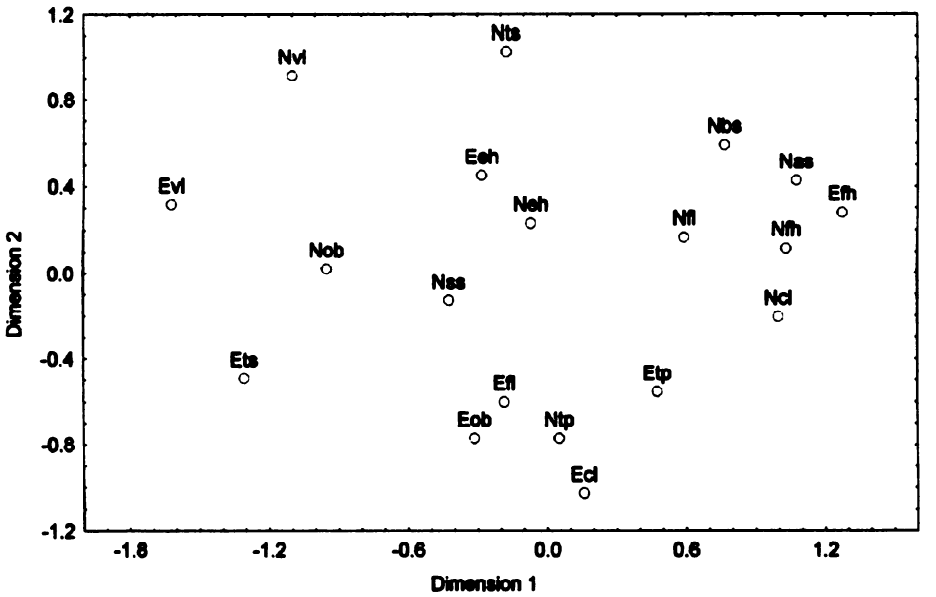

Fig. 5. Two-dimensional classical multidimensional scaling solution (R^2 = .84, N = 10) for natural and E-mu instruments. N = natural, E = E-mu, as = alto sax, bs = bassoon, cl = clarinet, eh = English horn, fh = French horn, fl = flute, ob = oboe, ss = soprano sax, tp = trumpet, ts = tenor sax, vl = violin.

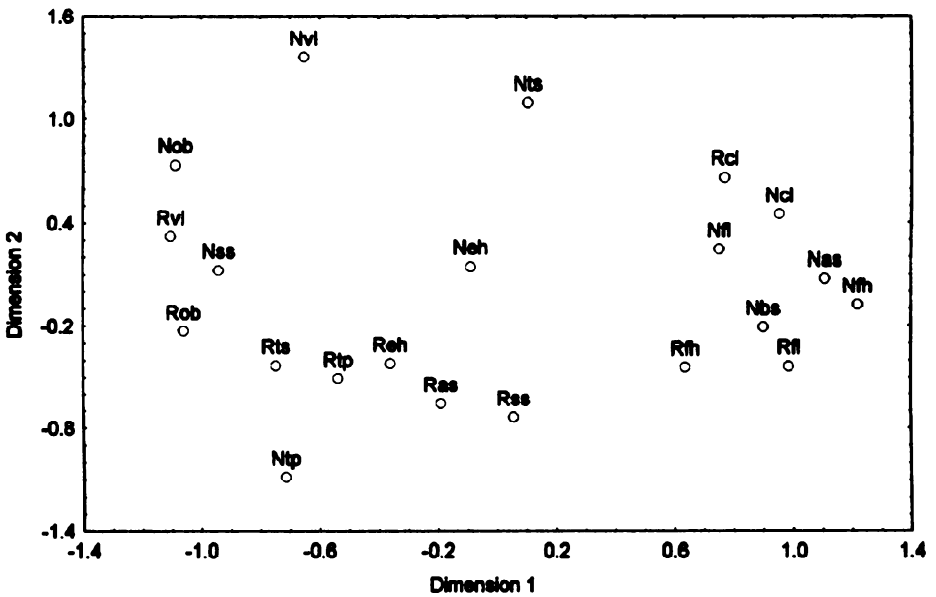

Fig. 6. Two-dimensional classical multidimensional scaling solution (R^2 = .81, N = 10) for natural and Roland instruments. N = natural, R = Roland, as = alto sax, bs = bassoon, cl = clarinet, eh = English horn, fh = French horn, fl = flute, ob = oboe, ss = soprano sax, tp = trumpet, ts = tenor sax, vl = violin.

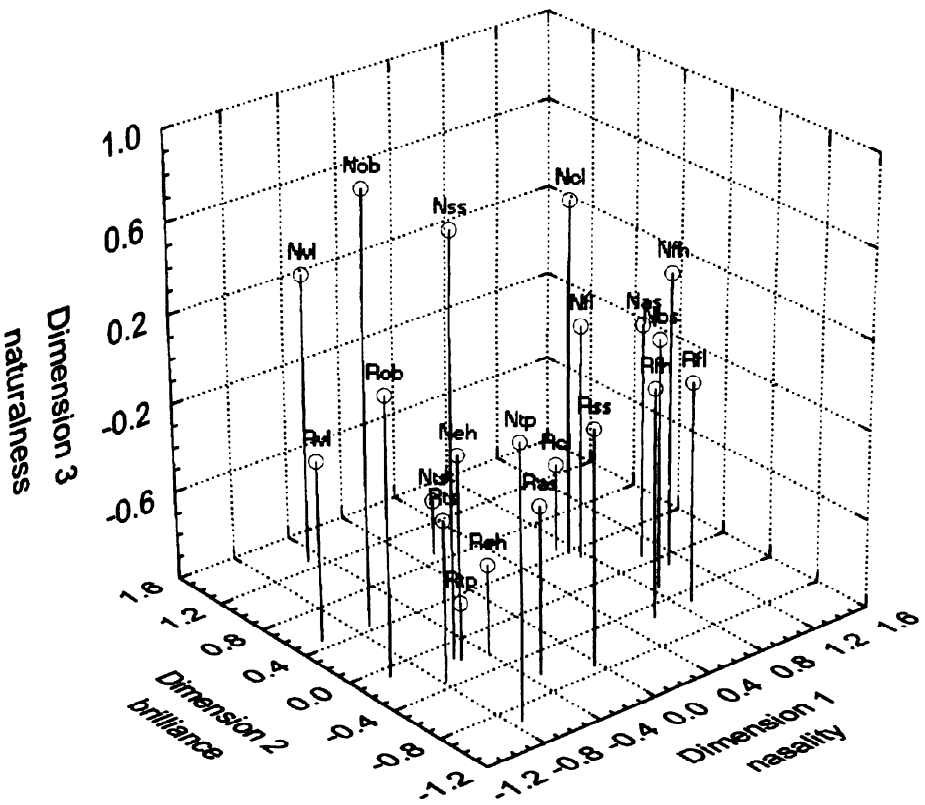

Fig. 7. Three-dimensional classical multidimensional scaling solution (R^2 = .96, N = 10) for natural and Roland instruments. N = natural, R = Roland, as = alto sax, bs = bassoon, cl = clarinet, eh = English horn, fh = French horn, fl = flute, ob = oboe, ss = soprano sax, tp = trumpet, ts = tenor sax, vl = violin.

the third dimension as *naturalness*. The divergence from naturalness is ordered among generators from E-mu (least) through Roland to Yamaha (most). In the interest of space, the other two three-dimensional graphs are not presented.

Figure 8 shows the results for natural plus Yamaha. The Yamaha flute is nearly on top of the Yamaha French horn and very close to the natural French horn. This easily accounts for the identification and categorization mismatches reported here. Yamaha trumpet, oboe, and violin are less *nasal* than their counterparts, and the Yamaha clarinet is somewhat too *nasal* and it needs to be more *brilliant*.

For the seven in-common instruments, the mean dissimilarity rating for comparisons of naturals versus synthetics was calculated (e.g., natural clarinet vs. E-mu clarinet). The correlation of these means with the corresponding identification percentage correct (combined long and short signals) was –.365. The same procedure was applied to the categorization ratings, and

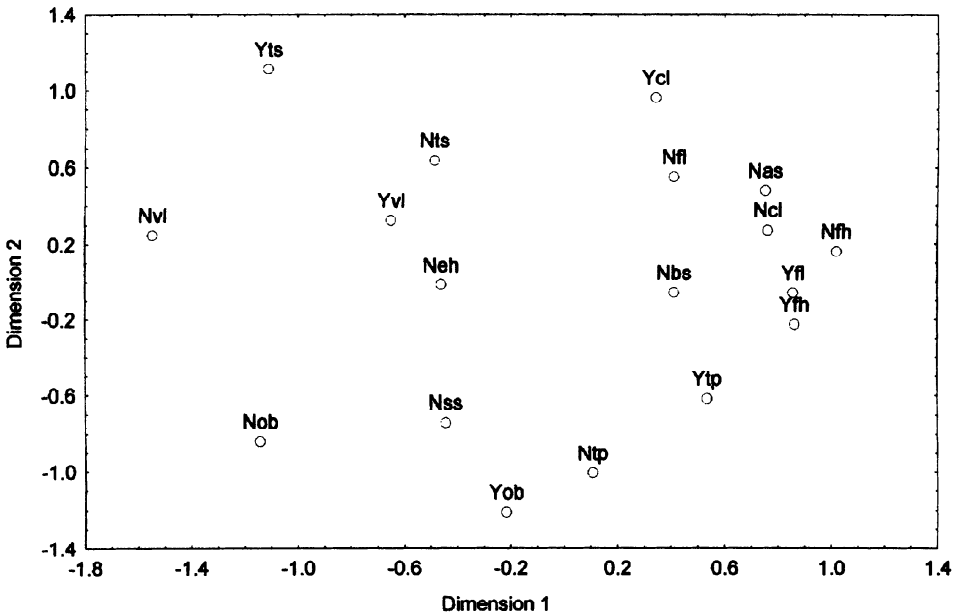

Fig. 8. Two-dimensional classical multidimensional scaling solution (R^2 = .84, N = 10) for natural and Yamaha instruments. N = natural, Y = Yamaha, as = alto sax, bs = bassoon, cl = clarinet, eh = English horn, fh = French horn, fl = flute, ob = oboe, ss = soprano sax, tp = trumpet, ts = tenor sax, vl = violin.

the correlation was –.305. Common sense suggests that if a synthetic emulation is dissimilar to its natural counterpart, then the identification accuracy should be lower. A stronger negative correlation than we found might be anticipated; however, some instruments (e.g., tenor saxophone) were poorly identified no matter what the emulation, a situation that worked against a stronger result.

In general, data show that synthesized *nasal* instruments are less similar to their natural counterparts than are *not nasal* instruments such as French horn. This corresponds to the just-noticeable difference (JND) for centroid, which is larger for lower centroids, such as the French horn, than for higher centroids, such as the oboe (Kendall & Carterette, 1996). We turn now, in fact, to interpreting these geometrical scalings in terms of acoustical variables.

Acoustical Analyses

Our purpose here is to connect the results of similarity scaling, identification, and categorization with acoustical features of the various natural

and synthetic signals. We used the monophonic signals, 16-bit, sampled at 22.05 k samples per second. It is our contention that the envelope characteristics of continuant signals are not a primary variable in their perceptual differentiation (Hajda, 1996; Kendall, 1986).

TEMPORAL VARIABLES

We calculated total signal durations using the root mean square (RMS) within 64-sample windows (2.9 ms). The start and end of the signal were defined relative to the noise floor (recall that all signals, synthetic or not, were recorded in a concert hall). The correlation of the signal duration and the 2-dimensional CMDS solutions reported above was .176 for the natural timbres alone (11 stimuli), .310 for the natural and E-mu (19 stimuli), .134 for the natural and Roland (21 stimuli), and –.098 for the natural and Yamaha (18 stimuli). We conclude that there was only the weakest relationship between total signal duration and the orientation of our perceptual spaces.

Next we approached calculating attack time using a number of criteria, including those of Krimphoff et al. (1994) and McAdams et al. (1995). Because natural signals do not follow the standard attack-decay-sustain-release (ADSR) envelope, and because of the level of the noise floor for our signals, which had been recorded in an auditorium, we were prevented from using the method of Krimphoff et al. (1994). In addition, the global maximum amplitude occurs well into the steady-state portion of our continuant signals. Therefore, rise time was taken to begin after any preattack noise, identified by the presence of undulating RMS amplitude. The rise time ended with the ensuing local maximum, or change in slope, of the global RMS function.[7]

The correlations of the natural log transforms of the rise times with the CMDS spaces were .371 for the naturals alone, .262 for the naturals plus E-mu, .261 for the naturals plus Roland, and .050 for the naturals plus Yamaha. Therefore, we failed to obtain the high correlations of Krimphoff et al. (1994) and McAdams et al. (1995), which were .94 and –.94, respectively. This was expected because our stimuli were recorded in a natural environment and did not include impulse instruments.

7. Rise times were calculated by using the following procedure: (1) RMS data files were calculated by using a 64-sample analysis window (2.9 ms). (2) The RMS functions were smoothed by using a running mean comprised of three consecutive analysis windows. (3) The start of the rise time was assumed to be within the first 200 ms of the signal. It was marked by at least seven consecutively increasing amplitude values (20.3 ms). (4) The end of the rise time was marked at either (a) the point of two decreasing amplitude windows or (b) the point of one decreasing amplitude window followed by an increasing amplitude window such that, for the xth window, amplitude(x + 2) < amplitude(x).

ANALYSIS PARAMETERS

Because we found only weak relations between global temporal features and the perceptual spaces, we chose to perform acoustical analyses on the steady-state portion of the signal, beginning 500 ms from the onset and ending 1520 ms later, for a total of 1020 ms.

Our physical variables are based on ninth-order fast Fourier transforms (FFTs) using a Hanning window with a frame size of 512 samples that gave a bandwidth resolution of 43.07 Hz. Each frame of the analysis spans 23.2 ms; there were 44 analysis frames.

Three physical measurements were calculated. First, we measured the long-time average centroid (LTAC). This is the mean spectral centroid across time. The equation for spectral centroid is

$$\frac{\sum_{n=1}^{p} f_n A_n}{f_1 \sum_{n=1}^{p} A_n} \tag{1}$$

where p is the number of analyzed partials (9), f_n is the frequency of the nth component (band), f_1 is the fundamental frequency, and A_n is the linear amplitude of the nth component (band). We calculated the mean spectral centroid across 44 analysis time frames (1020 ms). Second, we measured centroid variability as the standard deviation of the centroid values across time. Third, we calculated the mean coefficient of variation, a measure of spectral flux we used in our previous dyad studies (Kendall & Carterette, 1993b),

$$\frac{\sum_{n=1}^{p} \frac{\sigma_n}{\mu_n}}{p} \tag{2}$$

where σ_n is the standard deviation of the amplitude of the nth component, μ_n is the mean amplitude of the nth component, and p is the number of bands (partials).

LONG-TIME AVERAGE CENTROID

In order to explain the perceptual scaling relative to these acoustical variables, a program was written to rotate the spaces for maximal correlation (Pearson r) with each physical measure in turn. The iterative procedure rotates a space in 0.5-degree steps, finding the maximum positive or negative correlation.

Table 2 shows results for LTAC across seven different contexts involving both the natural and synthetic timbres. Each context involves an N of 10. All rotations are clockwise in degrees. Recall that Figures 4, 5, 6, and 8 show the two-dimensional CMDS spaces for the natural instruments alone, and for the natural together with the three emulators. It is easy to see that

TABLE 2
Long-Time Average Centroid

Context	Maximal *r*	Rotation (°)
E-mu	.990	35.0
Roland	.963	19.5
Yamaha	.955	159.5
Natural	.952	2.5
Natural plus E-mu	.941	–7.0
Natural plus Roland	.937	31.5
Natural plus Yamaha	.924	7.0

long-time average centroid maps very strongly to the timbre spaces; these best fits require little rotation generally. It should be understood that the apparently large rotation for Yamaha (159.5°) arises from the transposition of sign on the axes.

From seven different scalings with different groups of subjects, a statement is adduced of the reliability of mapping between LTAC and the principal dimension of timbre in the case of continuant signals. Because of this strong mapping (ca. $r = .90$ and above), we can conclude that synthetic instruments diverge from naturals along Dimension 1 because of centroid mismatches, a fact most dramatically in evidence for the confusion of Yamaha flute with natural and Yamaha French horn. This can be seen in Figure 9, which shows a complete linkage clustering of the centroid data across all instruments. It is worth noting that the major branch division separates cleanly between nasal and less nasal instruments at a point roughly corresponding to the origin ($x = 0$) of the perceptual spaces. The upper secondary branches separate natural French horn and alto saxophone (and emulators with similar centroids) from natural flute and clarinet. Based on Figure 9, you can judge the consecutive centroid differences, noting that under laboratory conditions, the JND for centroid was found to be maximally .153 (Kendall & Carterette, 1996) using the fundamental-weighted centroid (Eq. 1), which is a unitless measure. Higher centroids, reflected in the cluster diagram by the lower branches, tend to have greater consecutive differences, and cluster with more branching levels. The natural oboe, tenor saxophone, and violin are conspicuously clustered. Tenor sax emulations from E-mu and Yamaha, and violin emulations from E-mu, have higher centroids than their natural counterparts. E-mu French horn, natural French horn, and Yamaha flute cluster at the lowest centroids, corresponding to the confusion of Yamaha flute and these French horns in both identification and categorization.

Centroid measures are useful in musical psychophysics by providing benchmark values for comparative analysis, particularly because a strong mapping to the primary dimension of continuant instrument perceptual

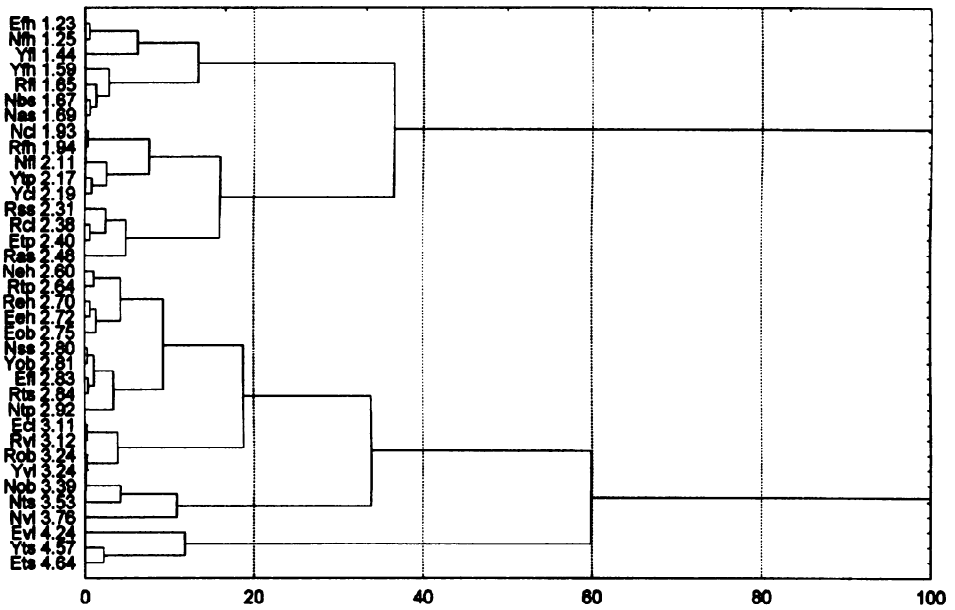

Fig. 9. Hierarchical cluster analysis (complete linkage) of spectral centroids for all instruments used in these experiments. The value of the centroid is to the right of each instrument label. N = natural, E = E-mu, R = Roland, Y = Yamaha, as = alto sax, bs = bassoon, cl = clarinet, eh = English horn, fh = French horn, fl = flute, ob = oboe, ss = soprano sax, tp = trumpet, ts = tenor sax, vl = violin.

space is invariably found under many conditions. Do not be misled, however, into concluding that, because two centroid differences are smaller than the JND for centroid, the instruments always will have indistinguishable timbres. Other physical variables may be effective in telling such instruments apart, such as time variancy.

TIME VARIANCY

Three forms of time variance include: (1) The change in spectral centroid over time as the standard deviation of successive centroids; (2) Spectral flux as measured by the mean coefficient of variation (Eq. 2) for the first nine partials; and (3) The overall change in RMS of the two-dimensional signal. We investigate the first two of these variances below.

The MDS solutions for seven experimental contexts were rotated into maximum correlation with the mean coefficient of variation (Eq. 2) for the first nine partials. Table 3 shows the results of this analysis. It is striking that the natural signals alone correlate well (r = .752) at a rotation of 73 degrees. This demonstrates near-orthogonality relative to the centroid dimension. The E-mu instruments in the presence of naturals also share this

TABLE 3
Mean Coefficient of Variation

Context	Maximal *r*	Rotation (°)
E-mu	.905	−17.5
Roland	.753	7.5
Yamaha	.074	No correlation
Natural	.752	73.0
Natural and E-mu	.726	246.5
Natural and Roland	.040	No correlation
Natural and Yamaha	.243	62.0

property, because a 246.5° rotation approaches 270°. Figure 10 shows the natural instrument two-dimensional MDS space rotated onto the mean coefficient of variation. Note that the oboe is now at the bottom of the figure, the trumpet is at the right, and the violin and tenor sax are at the left. Maximum spectral flux is at the left of the figure; minimum at the right. Long-time average centroid is now Dimension 2 in this figure (e.g., oboe vs. French horn).

We note that, for synthetic signals alone, spectral flux is correlated highly with the first dimension with little rotation, or not correlated at all, as is the case with the Yamaha. The time-variant statistics somewhat overlap

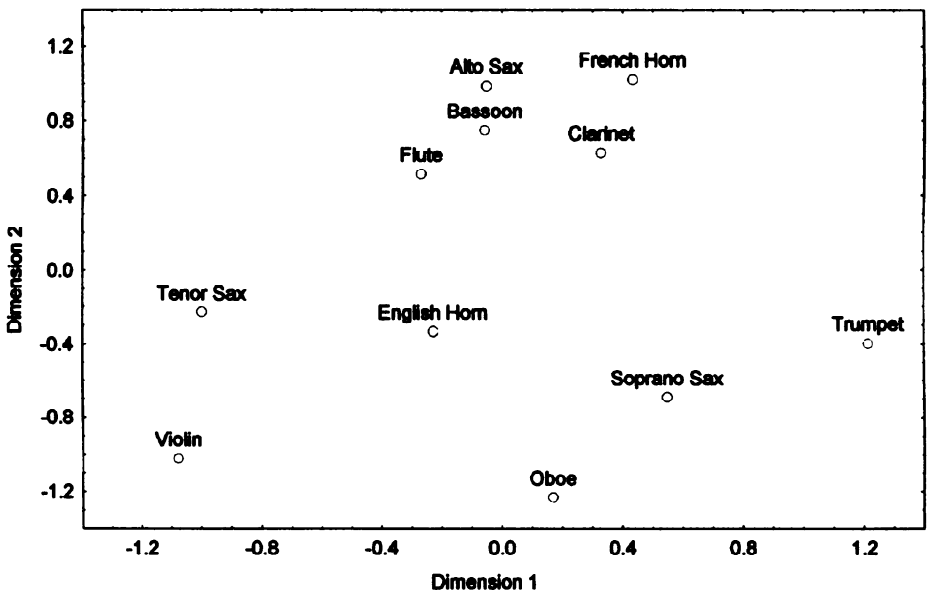

Fig. 10. Natural instrument classical multidimensional scaling (Figure 4) rotated onto the mean coefficient of variation.

for synthetic signals, but for naturals, they approach complete independence. This can be explained by a general tendency, evident in the spectra, for synthetic instruments to have a uniformly applied amplitude variation across all components; thus their mean coefficient of variation is large relative to the natural signals, in which variability arises from the interaction of the driving vibration (e.g., reed or string) with the resonance characteristics of the system. In addition, these differences in time variancy reflect a context effect such that Rolands in the presence of naturals do not produce a mean coefficient of variation correlation, but analyzed separately, the Roland mean coefficients of variation overlap the same dimension (Dimension 1) as long time average centroid (Table 2).

Table 4 shows the results of rotating the solutions for the seven contexts on a variable that combines centroid and time variance. For the naturals, maximum correlation (r = .931) is achieved with a rotation of 37.5°, approaching half the rotation for spectral flux, and this property is shared by many of the synthetic instruments. The natural, E-mu, and natural plus E-mu contexts produce higher correlations for centroid variability than do the other four contexts. This kindred property of natural and E-mu holds for the other measures of variability as well. Perhaps a fundamental property that distinguishes natural instrument tones and their sampled E-mu counterparts is the existence of two partially independent sources of variability, a property less characteristic of pure FM synthesis (Yamaha) and hybrid synthesis (Roland).

Previously we noted (Kendall & Carterette, 1993b) that interpretation of dyad timbre spaces in terms of verbal attributes implied that *reedy* tones combined relatively high time variance with relatively high spectral centroid, thus occupying the upper left quadrant of our perceptual spaces. Similarly, *rich* tones have relatively high time variance with low spectral centroid and occupy the upper right quadrant. The rotation of the natural spaces in the present experiment seems to add weight to this interpretation. We turn now to an experiment with verbal attributes that uses the stimuli of the present research.

TABLE 4
Centroid Variability (Standard Deviation)

Context	Maximal r	Rotation (°)
E-mu	.911	10.5
Roland	.758	49.0
Yamaha	.658	145.5
Natural	.931	37.5
Natural and E-mu	.843	34.0
Natural and Roland	.666	47.5
Natural and Yamaha	.720	16.5

Experiment 4: Verbal Attributes

Our final perceptual experiment explores the verbal characteristics of natural and synthetic single instrument tones. Previously (Kendall & Carterette, 1993a, 1993b) we conducted extensive research into verbal attributes of simultaneously sounding wind instrument dyads. We use some of the findings to guide our work here, which is directed toward assisting the interpretation of the perceptual spaces relative to acoustical variables. For this purpose, we confine our attention to the natural instruments that appeared in four of the major contexts.

METHOD

Subjects

Twenty-two subjects, 11 musicians and 11 nonmusicians, participated in the experiment. Musicians were paid $10 and nonmusicians received academic credit for their efforts.

Procedure

The eleven $B\flat_4$ natural-instrument signals (p. 333), 3 s in length, were used: Trumpet, bassoon, French horn, soprano saxophone, alto saxophone, tenor saxophone, flute, oboe, English horn, clarinet, and violin. Eight timbral descriptors from the principal components analysis of Kendall and Carterette (1993b) were chosen: strong, tremulous, light (Factor 1), nasal, rich (Factor 2), brilliant, ringing (Factor 3) and reedy (Factor 4). These were used in a VAME (verbal attribute magnitude estimation) procedure (Kendall & Carterette, 1993a). Subjects heard diotically by headphone (Sennheiser HD-222) a randomly selected stimulus and rated the magnitude of the attribute on a 100-point scale presented on a computer screen. A mouse was used to move a slider to the selected position; the initial position of the slider was randomly set on the scale. An example scale would be "*not nasal–nasal.*" The poles were randomly assigned. The stimulus was heard before every scale for the 88 items of the experiment.

RESULTS

The data across subjects by instruments were subjected to principal components analysis with Varimax rotation. We present here the combined data sets from musicians and nonmusicians because the data do not vary significantly on the main effects (see following for ANOVA analysis). Table 5 shows the results of this analysis. As with our previous research with timbral dyads (Kendall & Carterette, 1993b), the factors include a *power* factor that suggests strongly the idea of *potency* from traditional semantic differential work. The words *strong* and to a certain extent *brilliant* and *ringing* are antipodes to *light*. This factor accounts for 33% of the variance. Factor 2, accounting for 24% of the variance, is the *stridency* or *nasal* factor. The words *nasal* and *reedy*, and to a certain extent *brilliant*, stand in opposition to *rich* and *ringing*; there was little weighting on Factor 2 for *light*, *tremulous*, *ringing*, and *strong*. Finally, our third factor, ac-

TABLE 5
Principal Components Analysis of VAME: Varimax Rotation

Variable	Factor 1 (Power/Potency)	Factor 2 (Stridency/Nasal)	Factor 3 (Vibrato)
Strong	–.912	–.037	–.197
Tremulous	.203	.133	.903
Light	.930	.127	–.128
Nasal	.006	.981	.089
Rich	–.240	–.409	.701
Brilliant	–.556	.492	.571
Ringing	–.586	–.115	.553
Reedy	.426	.681	–.280
Proportion of variance accounted for	.329	.236	.260

counting for 26% of the variance, is the *vibrato* factor on which *tremulous* and *rich* load positively relative to *reedy*, *light*, and *strong*. This factor is somewhat different than that obtained with dyads, but, as we shall see, makes perfect sense with regards to our present set of experiments. Recall that there was a strong correlation between our acoustical analysis of time variancy, expressed in the coefficient of variation (Eq. 2), and the second dimension of the natural MDS space (Figure 4); this should map onto the *vibrato* factor.

The cross correlation of verbal attribute ratings across instruments and subjects ($N = 22$) was submitted to CMDS. The two-dimensional solution, which accounted for 97% of the variance, is shown in Figure 11. The axes of the solution have been rotated 90° counterclockwise for comparison with the timbral similarity spaces. In this space, Dimension 2 is the *power* dimension with extremes of *strong, ringing* versus light. Dimension 1 separates *nasal*, *strong*, and *reedy* versus *rich* and *tremulous*. *Brilliant* lies near the origin.

One must note, however, that words are not instrument sounds, and that sonic and verbal spaces are not isomorphic. We believe that this is best shown in the case of the first factor of the principal components analysis, *power*. And in fact we previously noted (Kendall & Carterette, 1993b) that timbral loudness may be different from loudness as traditionally conceived. However, it is also possible that this *power* factor represents the cultural associations of instruments or sounds, implying that, particularly for the music majors, the instruments are identified and then rated. Trumpet, for example, may be a masculine instrument, suggesting *strength*, whereas flute is considered feminine and therefore *light*.

We analyze the VAME data further by using repeated measures ANOVA. The mixed design has two within-subjects factors, instrument and rating

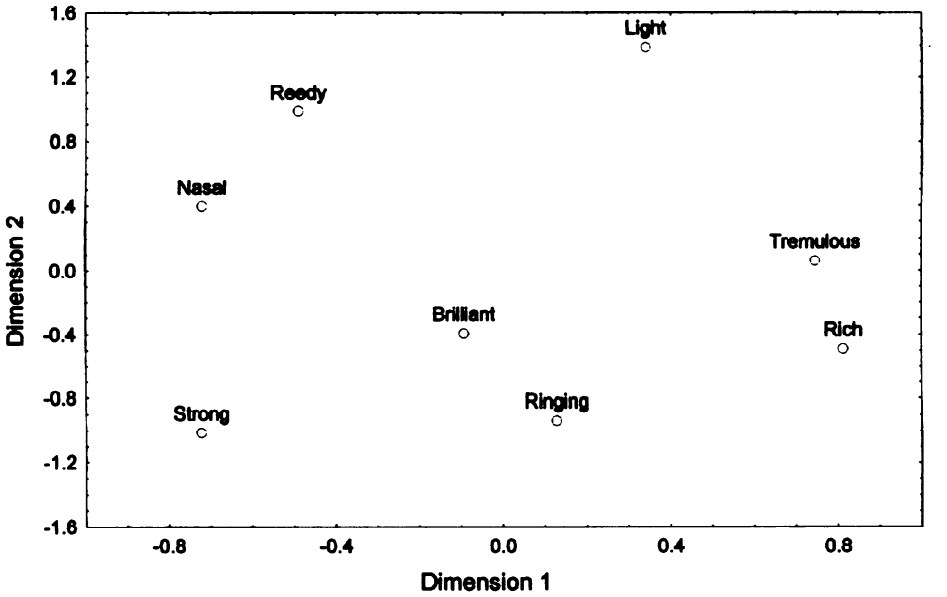

Fig. 11. Two-dimensional classical multidimensional scaling (R^2 = .97, N = 22) for verbal attributes based on correlations across subjects and natural instruments.

scale, and one between-subjects factor, group (musicians or nonmusicians). One would expect main-effect differences for instrument and rating scale, and that is found, $F(10, 200)$, $p < .0000009$ and $F(7, 140)$, $p < .000027$, respectively. We do not analyze these post-hoc because they have little meaning relative to the current study. Also, the group means (musicians vs. nonmusicians) are not significantly different.

The two-way interaction of instrument by rating scale is significant, $F(70, 1400)$, $p < .0000009$. As predicted, *nasal* instruments are oboe, with the highest mean, in the company of soprano sax and tenor sax. Instruments of moderate *nasality* include trumpet, English horn, and violin. Clarinet also has a moderate nasality, which is not predicted by our interpretation of the similarity scalings. Instruments that are not nasal include bassoon, French horn, and flute.

We cross-correlated the physical measures, discussed earlier, with the verbal rating scales. As predicted, the attribute *nasal* correlates highly with long-time average spectral centroid (r = .77). It follows that *nasal* correlates well with the first dimension of the perceptual space for natural instruments as well (r = .80).

In contrast to nasal, *reedy*, originally hypothesized to correlate with the upper left quadrant of the perceptual spaces, fails to differentiate among instruments as expected. For example, the violin is rated as *not reedy*, along with the flute, trumpet, French horn, and to a certain extent the bassoon

(on the high note of $B\flat_4$). Most of the saxophones are rated *reedy*, with the alto saxophone being less so. *Reedy* fails to correlate well with the dimensions of the similarity space for natural instruments.

Dimension 2 of the perceptual space is correlated moderately well with *rich* ($r = -.43$), *brilliant* ($r = -.52$), and *tremulous* ($r = .40$). This modestly supports our previous interpretation of perceptual spaces, but also indicates a lack of orthogonality in these measures. In fact, *brilliant* and *tremulous* correlate with Dimension 1 as well, which is hardly surprising considering that *nasality* and time variance are rotated for maximum correlation approximately 45°. The tendency among the natural instruments is for instruments of relatively high spectral centroid to also possess relatively high time variancy, such as the violin; an exception is the trumpet, which is *brilliant*. Although the range of means for *rich* is small, the values vary such that instruments above the origin of the similarity space (Figure 4) are *richer* than those below. This does not confirm our specific hypothesis, derived from previous work with dyadic spaces, that *rich* is always a combination of low spectral centroid and time variancy, associated with the upper right quadrant of the space, and instead suggests that *brilliant* and *rich* are simple antipodes.

The verbal attributes *strong* and *light*, clearly antipodes, fail to map to the perceptual similarity space; they appear highly charged with semantic, cultural meaning and, we repeat, they are the primary component in the principal components analysis of verbal attributes. They simply do not map readily to timbral similarity, as they account for less than .04% of the covariance between either dimension and the attributes. *Ringing* shares a similar fate. The three-way group by instrument by rating scale interaction is significant, $F(70, 1400)$, $p < .000313$. Few trends in the interaction are discernible, except for the fact that musicians and nonmusicians more often disagree on verbal ratings for instruments of higher spectral centroid. For example, musicians rate violin as *nasal* and *strong* in contrast to nonmusicians. Oboe is less *light* to musicians than nonmusicians, and English horn is less *reedy*. There are exceptions even to this trend, in that French horn is more *nasal* to nonmusicians than to musicians, and flute more *rich*.

Conclusions

All of these data and all of these analyses drive us to conclude that there is a remarkable stability in the behavior of the instruments and their perceptions, across seven contexts of instruments and variables such as musical training, long versus short notes, monophonic versus stereo. The spatial MDS relationships remain the same and are consistent with our previous work with dyads (Kendall & Carterette, 1991).

The departure of emulated instrument coordinates from natural instrument coordinates was readily explained in terms of physical measures. Spectral centroid mapped with correlations near or exceeding .90 to the principle dimension of similarity, *nasality*. The time-variancy measures, including centroid variability, also exhibited strong correlation to higher dimensions of the timbral similarity for natural instruments, less strong or nonexistent for many synthetic emulations. This suggests that a significant problem in emulating natural instruments is the failure to match time-variant characteristics in the steady state. For example, we found evidence that omnibus techniques for creating amplitude variance, as employed in emulators, exaggerated variancy relative to natural instruments. Also, combining mismatches among centroid and time variancy in synthesis can create a real aberration (e.g., Yamaha flute). Nowhere was this more apparent than in the identification and categorization experiments, where the natural instruments produced significantly more accurate results than the E-mu, which was based on samples of natural instruments. This was followed by the Roland and Yamaha, which were less accurately identified and categorized than either E-mu or the natural instruments. In addition, the miscategorization of instruments such as the Yamaha flute with French horn found expression in their proximity in the similarity spaces.

Comparing our perceptual spaces to research conducted in other laboratories is complicated by the fact that previous work almost universally included impulse signals in the stimulus set. This produces perceptual spaces that have near-categorical properties on the first dimension correlated strongly with attack time (Iverson & Krumhansl, 1993; McAdams et al., 1995). In contrast, we found only a relatively low positive correlation among our natural continuant signals with attack time (r = .371) compared with Krimphoff et al. (1994; r = .94) and McAdams et al. (1995; r = –.94), who used both continuant and impulse signals based on FM synthesis. The mapping of spectral centroid to our first dimension was as strong (r = .95 for naturals under optimal rotation) as that found for the second dimension of a number of previous studies (Iverson & Krumhansl, 1993, ca. r = –.70; Krimphoff et al., 1994, ca. r = .94) that used both impulse and continuant signals. When only quasi-natural, continuant signals were used, as in the present experiments and those of Grey and Gordon (1978, see also Grey 1975, 1977), the high correlation of centroid is to first dimension coordinates (r = .94 for loudness-function-based centroids; r = .93 for centroids not based on loudness functions). These facts support our hypothesis that impulse signals polarize similarity spaces in the presence of continuant sounds (see Hajda, 1996, for additional experiments comparing impulse and continuant signals and their perceptual properties).

Data are available from three previous studies that used similar sets of instruments. Plomp (1970, 1976) used synthesized instruments based on single periods of the steady state of nine natural tones. Six instruments

were in common with the present study: oboe, clarinet, violin, trumpet, bassoon, and French horn. The stimuli were at 349 Hz (ca. F_4, in contrast to the $B\flat_4$ used in the present experiments). The two dimensional CMDS solution from Plomp's data (Hajda et al., 1997) was correlated with the coordinates from the natural space of the present study. Dimension 1 correlation was .51, and Dimension 2 correlation was .61 for the six instruments.

Next we compared our natural signal results to the work of Wedin and Goude (1972), who used stimuli at 440 Hz (A_4), within a half-step of the tone chosen for our study. Seven instruments were in common with the present study: French horn, clarinet, bassoon, flute, trumpet, oboe, and violin. A two-dimensional CMDS solution based on their data (Hajda et al., 1997) produced a correlation with our study on Dimension 1 of .84 and for Dimension 2 of –.15. We had often reported that the Wedin and Goude (1972) data corresponded well with our results, including those derived from dyads (Kendall & Carterette, 1991), and here we find additional strong evidence for this relative to the first dimension.

Finally, we compare our two-dimensional CMDS solution to the three-dimensional CLASCAL space with latent classes and specificities of McAdams et al. (1995). The stimuli were Yamaha FM-synthesized signals. Six 311 Hz ($E\flat_4$) emulations were in common with the present study: French horn, clarinet, bassoon, trumpet, English horn, and violin (the "bowed string" instrument of their study). The correlation of Dimension 1 (centroid) of the natural instrument space in our study (Figure 4) with Dimension 2 (centroid) of the McAdams et al. (1995) study was .58 for these six instruments, and Dimension 2 in our study (time variancy) had a correlation of –.08 with Dimension 3 (spectral flux) in McAdams et al. (1995).

Therefore, even with some very striking differences in technique and stimulus set characteristics, small to moderate positive correlations are found for Dimension 1, the centroid (*nasality*) dimension. The relations among spectral flux dimensions is less compelling, but we have already elaborated on differences in natural signals and emulations that apply a fixed technique for creating time variancy among all partials of a spectrum. It is probably no accident that the space with the strongest mapping (Wedin & Goude, 1972) is in the same tessitura as the present study, but with such a large number of differences among experimental conditions, caution is advisable.

Our experiments with verbal attributes in this study produced the same factors from principal components analysis as did our work with dyads (Kendall & Carterette, 1993b). Once again, a *power* factor accounted for the most variance, followed by a *stridency* factor containing the verbal attribute *nasal* with its highest loading. The *power* factor does not appear to correlate with the first two or three dimensions of our perceptual spaces.

However, the *stridency* factor, in terms of the verbal attribute *nasal*, has a high positive correlation with the first dimension of the perceptual space for natural instruments ($r = .80$) and with long-time-average spectral centroid as well ($r = .77$). The third factor we called *vibrato* because it dealt with the attributes *tremulous* and *rich,* which are associated with time-variancy. There were modest correlations of *rich* ($r = -.43$), *tremulous* ($r = .40$) and *brilliant* ($r = -.52$) with Dimension 2 of the natural perceptual space (Figure 4).

Within our present study, the stability of the main results across numerous contexts inspires us to extend this work. Clearly, centroid and time variability must be investigated as precisely controlled independent variables. In addition, the issue of tessitura and timbre, within and among instruments, deserves deeper consideration. One possibility is an approach that moves from physical measurements to perceptual spaces and tests validity through resynthesis. This analysis by synthesis is based on a useful ordering of methodological techniques, moving cyclically from perceptual analysis to acoustical analysis to timbral synthesis.[8]

References

Ashby, F., & Maddox, W. (1998). Stimulus categorization. In M. H. Birnbaum (Ed.), *Measurement judgement, and decision making* (pp. 252–301). San Diego: Academic Press.

Bismarck, G. von (1974a). Timbre of steady sounds: A factorial investigation of its verbal attributes. *Acustica, 30,* 146–159.

Bismarck, G. von (1974b). Sharpness as an attribute of the timbre of steady sounds. *Acustica, 30,* 159–192.

Carroll, J. D., & Chang, J. J. (1970). Analysis of individual differences in multidimensional scaling via an N-way generalization of 'Eckart-Young' decomposition. *Psychometrika, 41,* 283-319.

Carterette, E. C., & Kendall, R. A. (1996). Musical communication. In H. Fastl, S. Kuwano, & A. Schick (Eds.), *Recent trends in hearing research* (pp. 131–160). Oldenburg, Germany: Bibliotheks-und Informationssystem der Universtät Oldenburg.

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

Clark, M., Jr., Luce, D., Abrams, R., Schlossberg, H., & Rome, J. (1963). Preliminary experiments on the aural significance of parts of tones of orchestral instruments on choral tones. *Journal of the Audio Engineering Society, 11,* 45–54.

Clark, M., Jr., Robertson, P., & Luce, D. (1964). A preliminary experiment on the perceptual basis for musical instrument families. *Journal of the Audio Engineering Society, 12,* 199–203.

Donnadieu, S., McAdams, S., & Winsberg, S. (1996). Categorization, discrimination and context effects in the perception of natural and interpolated timbres. In B. Pennycook &

8. Preliminary reports of the results of parts of these experiments were presented at the Third International Conference on Music Perception and Cognition, Liège, Belgium (July 23–27, 1994), the International Symposium on Musical Acoustics, Le Normont, Dourdan, France (July 2–6, 1995), and the Fourth International Conference on Music Perception and Cognition, Montreal, Canada (August 11–15, 1996).

E. Costa-Giomi, *Proceedings of the Fourth International Conference on Music Perception and Cognition* (pp. 73–78). Montreal: McGill University.

Gordon, J. W. (1987). The perceptual attack time of musical tones. *Journal of the Acoustical Society of America, 82,* 88–105.

Grey, J. M. (1975). *An exploration of musical timbre.* Doctoral dissertation, Stanford University. [Department of Music Report STAM-M-2. Stanford, CA: Center for Research in Computer Applications in Music and Acoustics.]

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

Grey, J. M., & Gordon, J. W. (1978). Perception of spectral modifications on orchestral instrument tones. *Computer Music Journal, 11,* 24–31.

Hajda, J. M. (1995). The relationship between perceptual and acoustical analyses of natural and synthetic impulse signals. (Master's thesis, University of California, Los Angeles, 1995). *Masters Abstracts International, 33*(6). (University Microfilms International Publications No. 13-61, 681)

Hajda, J. M. (1996). The effect of reverse playback and signal partitioning on the identification of percussive and nonpercussive musical tones. In B. Pennycook & E. Costa-Giomi (Eds.), *Proceedings of the Fourth International Conference on Music Perception and Cognition* (pp. 25–30). Montreal: McGill University.

Hajda, J. M., Kendall, R. A., Carterette, E. C., & Harshberger, M. L. (1997). Methodological issues in timbre research. In I. Deliège & J. Sloboda (Eds.), *The perception and cognition of music* (pp. 253–306). London: L. Erlbaum.

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

Jakobson, R., Fant, C. G. M., & Halle, M. (1951/1961/1963). *Preliminaries to speech analysis: The distinctive features and their correlates.* Cambridge, MA: MIT Press [reprint of Acoustics Laboratory, MIT, Technical report No. 13, with addenda et corrigenda, and supplement on "Tenseness and Laxness."]

Kendall, R. A. (1986). The role of acoustic signal partitions in listener categorization of musical phrases. *Music Perception, 4,* 185–214.

Kendall, R. A., & Carterette, E. C. (1989). Perceptual, verbal, and acoustical attributes of wind instrument dyads. In *Proceedings of the First International Conference on Music Perception and Cognition* (pp. 365–370). Kyoto, Japan: Japanese Society of Music Perception and Cognition.

Kendall, R. A., & Carterette, E. C. (1990). The communication of musical expression. *Music Perception, 8,* 129–164.

Kendall, R. A., & Carterette, E. C. (1991). Perceptual scaling of simultaneous wind instrument timbres. *Music Perception, 8,* 369–404.

Kendall, R. A., & Carterette, E. C. (1992). Convergent methods in psychomusical research based on integrated, interactive computer control. *Behavior Research Methods, Instruments, & Computers, 24,* 226–231.

Kendall, R. A., & Carterette, E. C. (1993a). Verbal attributes of simultaneous wind instrument timbres: I. von Bismarck's adjectives. *Music Perception, 10,* 445–468.

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

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

Kendall, R. A., & Carterette, E. C. (1996) Difference thresholds for timbre related to spectral centroid. In B. Pennycook & E. Costa-Giomi (Eds.), *Proceedings of the Fourth International Conference on Music Perception and Cognition* (pp. 91–95). Montreal: McGill University.

Kendall, R. A., Carterette, E. C., & Hajda, J. M. (1994). Comparative perceptual and acoustical analyses of natural and synthesized continuant timbres. In I. Deliège (Ed.), *Proceedings of the Third International Conference for Music Perception and Cognition* (pp. 317–318). Liège, Belgium: European Society for the Cognitive Sciences of Music.

Kendall, R. A., Carterette, E. C., & Hajda, J. M. (1995). Perceptual and acoustical attributes of natural and emulated orchestral instrument timbres. *Proceedings of the International Symposium on Musical Acoustics, July 2–6, 1995, Le Normont, Dourdan, France* (pp. 596–599). Paris: Societe Francaise d'Acoustique.

Krimphoff, J., McAdams, S., & Winsberg, S. (1994). Caractérisation du timbre des sons complexes. II. Analyses acoustiques et quantification psychophysique. *Journal de Physique IV, Colloque C5, supplément au Journal de Physique III, 4,* 625–628.

Krumhansl, C. L. (1989). Why is musical timbre so hard to understand? In S. Nielzén & Olsson (Eds.), *Structure and perception of electroacoustic sound and music* (pp. 43–53). Amsterdam: Excerpta Medica.

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.

Kruskal, J. B., & Wish, M. (1978) *Multidimensional scaling.* Beverly Hills, CA: Sage Publications.

Lichte, W. H. (1941). Attributes of complex tones. *Journal of Experimental Psychology, 28,* 455–480.

McAdams, S., Winsberg, W., Donnadieu, S., De Soete, G., & Krimphoff, J. (1995). Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes. *Psychological Research, 58,* 177–192.

Miller, J. R., & Carterette, E. C. (1975). Perceptual space for musical structures. *Journal of the Acoustical Society of America, 27,* 337–352.

Plomp, R. (1970). Timbre as a multidimensional attribute of complex tones. In R. Plomp & G. F. Smoorenberg (Eds.), *Frequency analysis and periodicity detection in hearing* (pp. 397–411). Leiden, Netherlands: A. W. Sijthoff.

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

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

Saldanha, E. L., & Corso, J. F. (1964). Timbre cues and the identification of musical instruments. *Journal of the Acoustical Society of America, 36,* 1–15.

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

Schiffman, S. S., Reynolds, M. L., & Young, F. W. (1981). *Introduction to multidimensional scaling: Theory, methods, and applications.* New York: Academic Press.

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

Wessel, D. L. (1973). Psychoacoustics and music: A report from Michigan State University. *PACE: Bulletin of the Computer Arts Society, 30,* 1–2.

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

Appendix A
Description of Keyboards/Sound Modules

FM SYNTHESIS

Tones based on the Yamaha DX/TX series of sound generators have been used in a number of experimental studies on instrument timbre (Krumhansl, 1989; McAdams et al., 1995; Wessel et al., 1987). The Yamaha tones are based on FM synthesis as formulated by Chowning (1973). In our study, we used the Yamaha DX7, one of the early generation of FM keyboard synthesizers (the other studies cited used the Yamaha TX802, a descendent of

the DX7). The DX7, a 12-bit, 50k samples per second, monophonic system, ships with 128 factory presets made by "algorithms" of carriers and modulators. There are six operators, essentially a digital sine-wave generator with a digital envelope generator, which are combined in various ways. The DX7 has 32 linkings of these "algorithms." There is no difference between modulators and carriers, the main distinction being position in the stack. In our study the DX7 used the "E!" expansion circuit board, which gives expanded internal memory, microtonality, more MIDI control parameters, and contains 256 additional preset internal voices.

HYBRID SYNTHESIZER: ROLAND D-50

The Roland D-50 is based on what the manufacturers call "Linear Arithmetic Synthesis." Essentially, there are two classes of 16-bit generator. One, called the synthesizer, is like a traditional analog synthesizer, complete with low frequency and higher frequency oscillators, filters, envelope generators, and amplifiers. The other class of generator uses stored sound samples. A stereo output signal is the combination of up to two of these modules in all possible combinations; postprocessing includes reverberation and chorusing. We chose this instrument again because of its wide use and hybrid character of sampling and synthesis. Samples consist of 47 "one-shot" tones that are not looped and 29 "looped sounds." Examples of one-shot tones are impulsive like xylophone, harp and guitar, breaths, chinks, and noises. Looped sounds include arco violin, saxophone, and singing voice. The other 24 sounds are combinations of these two that are looped. We note that when a looped tone is increased in frequency, its loop rate increases, hence the periodicity of time-variant property of the signal varies with pitch. The idea of using samples in combination with synthesis was widely emulated, for example in the Kawai K1 and Korg M1.

SAMPLING KEYBOARDS

The third class of synthesizer is a sampling module, the E-Mu EmaxII and Proteus/2. Like the Roland D-50, the EmaxII provides facilities for 16-bit sampling and synthesis at 39k samples per second; however, the architecture is different. Whereas in the Roland D-50 the primary function of samples is to merge sampled attacks with a steady state of synthesized signals, the Emax II is a full-featured sampling keyboard with a built-in hard drive for the storage of sample sounds and provides such digital sampling effects as mixing, sound reversal, sample splicing, and individual tuning and attenuation for each sample. Sample presets are created by taking multiple samples across the tessitura of an instrument and then transposing the samples across a smaller range. In addition to digital sample processing, the keyboard provides for filtering, voltage-controlled envelopes, and other dynamic processing often found in traditional synthesis. As with the Roland D-50, a patch can consist of up to two channels either of which may be synthesized or sampled for producing stereo output. Synthesis in the Emax II is based on a "spectrum-space synthesizer," which is a traditional additive synthesizer. The spectrum synthesizer can manipulate up to 24 sinusoidal signals. The Proteus/2 uses read-only memory cartridges made up of various subsets of the full set of samples of the E-Mu EmaxII. It does not provide for additive synthesis but relies rather on the layering of sampled sounds and is equipped with 384 presets.

Appendix B

Table of Voicings

Instrument	Yamaha DX7	Roland D-50	E-mu Proteus/2
Flute	ROM 3A-Patch1, Flute1	FC 41, Flute 1	FP 24, Flute
Violin (Arco)	E!PVL Bank 5-Patch 06, Violin 3	FC 23, Solo Violin	FP 2, Solo Violin
Tenor Sax	E!PVL Bank4-Patch 21, Saxophone 4	FC 31, Tenor Sax	EMAX II-Tenor Sax-Patch 8, Crisp Sax
Alto Sax	None	FC 34, Alto Sax	None
Soprano Sax	None	FC 32	None
French Horn	E!PVL-Bank 4-Patch 14 French Horn	FC 25, Classical Horn	FP 36, French Horn
English Horn	None	3rd Party Patch, English Horn	FP 27, English Horn
Clarinet	ROM 4A-Patch 4, Clarinet	FC 44, Clarinet	FP 28, Clarinet
Bassoon	ROM 4A-5, Bassoon	FC 43, Bassoon	FP 30, Bassoon
Trumpet	E!PVL-Bank 4-Patch 20, Trumpet 1	FC 26, Classical Trumpet	FP 39, Trumpet 1