

A common perceptual space for harmonic and percussive timbres

STEPHEN LAKATOS

*John B. Pierce Laboratory, New Haven, Connecticut
and Institut de Recherche et de Coordination Acoustique/Musique (IRCAM), Paris, France*

The goal of a series of listening tests was to better isolate the principal dimensions of timbre, using a wide range of timbres and converging psychophysical techniques. Expert musicians and nonmusicians rated the timbral similarity of three sets of pitched and percussive instruments. Multidimensional scaling analyses indicated that both centroid and rise time comprise the principal acoustic factors across all stimulus sets and that musicians and nonmusicians did not differ significantly in their weighting of these factors. Clustering analyses revealed that participants also categorized percussive and, to a much lesser extent, pitched timbres according to underlying physical-acoustic commonalities. The findings demonstrate that spectral centroid and rise time represent principal perceptual dimensions of timbre, independent of musical training, but that the tendency to group timbres according to source properties increases with acoustic complexity.

Timbre is a complex and multidimensional perceptual attribute most closely associated with a sound's "quality" or "texture." Most studies of auditory timbre over the past three decades have employed some form of multidimensional scaling (MDS) algorithm to characterize its attributes. The principal goal of such research is to discover the mapping between the physical features of the signal and the mental representation of their corresponding timbral attributes by the listener. MDS algorithms usually make few a priori assumptions about the structural properties of psychological data and are therefore particularly appropriate for the study of complex stimuli whose underlying perceptual or psychophysical characteristics are poorly understood. One assumption is that timbre can be represented by a small number of continuous, orthogonal dimensions. An additional assumption underlying past studies has been that the timbres of two or more sounds should be compared under controlled conditions in which the pitch, loudness, and duration of such sounds have been equalized in order to control for possible interactions between these parameters and timbre.

Consistent reports have emerged in the literature regarding the principal acoustic determinants of timbre perception. For example, using synthetic tones with determined sets of timbral attributes, Miller and Carterette (1975) reported that the number of harmonics comprising

a sound, as well as the shape of the temporal energy envelope, characterized listeners' judgments of timbral similarity. Grey (1977; Grey & Gordon, 1978) used MDS techniques to derive a three-dimensional perceptual model of timbre, with the first dimension related to the spectral energy distribution of the sounds and with the remaining two axes corresponding to the degree of temporal synchronicity in the rise and decay of the upper harmonics and to the degree of spectral fluctuation in the signal envelope. Krimphoff, McAdams, and Winsberg (1994) reanalyzed a three-dimensional space obtained with synthetic sounds by Krumhansl (1989), and they found the following acoustic correlates for the dimensions: (1) the centroid of the sound spectrum, (2) the logarithm of the rise time, and (3) the "spectral flux" corresponding to the standard deviation of the time-averaged harmonic amplitudes from the sound's spectral envelope. Critical dynamic cues need not be present only in the rise portion of a tone but can occur throughout its duration (Iverson & Krumhansl, 1993; Wedin & Goude, 1972).

In general, then, MDS techniques typically yield either two or three interpretable psychophysical dimensions. Although several studies yield a third dimension, relating to temporal variations in either the spectral envelope or spectral fine structure, its psychophysical nature appears to vary with the composition of the stimulus set. Therefore, it remains somewhat unclear whether there are additional timbral dimensions besides those characterized by the central tendency of the amplitude spectrum and the rise portion of a tone, that generalize across a broad range of timbres. The inability to identify consistently a third dimension may reflect the use of only a limited number or range of instrument timbres in previous studies, or it may reflect limitations in the MDS algorithms themselves.

Several conceptual issues stemming from past MDS research are addressed in this paper. First, as a set of tim-

This research was supported in part by a CNRS postdoctoral fellowship and an NIH National Research Service Award to the author. The author thanks Lawrence Marks for providing facilities at the John B. Pierce Laboratory to complete portions of this study, Steve McAdams for contributing to the CLASCAL analyses, and Koei Kudo for assisting in running musically trained participants at IRCAM. Correspondence should be addressed to S. Lakatos, Washington State University, 14204 N.E. Salmon Creek Ave., Vancouver, WA 98686 (e-mail: lakatos@vancouver.wsu.edu).

bres becomes more varied and heterogeneous, do spectral center of gravity and rise time adequately characterize the dimensional structure of timbre, or do additional dimensions of timbre come into play? Furthermore, can interpretations of timbre based on the assumption of a continuous dimensional structure—an assumption made by most, if not all, MDS algorithms—be extended by analyses that do not make such strict assumptions? The most recent MDS algorithms available to researchers achieve such a level of sophistication that it is now possible to examine more complex interpretations of similarity data. The emergence of such new techniques encourages a reexamination of the dimensional nature of timbre by extending past approaches to larger, more varied stimulus sets.

The principal aim of the present study, therefore, was to expand on previous findings of MDS studies by using a recently developed MDS algorithm called *CLASCAL* (Winsberg & De Soete, 1993) to analyze the dimensional structure of timbre across a broader timbral range than previously tested. Stimuli were more heterogeneous than those in past studies, and they sampled a wide range of sound source properties and modes of excitation and included an assortment of pitched and unpitched percussive sounds, as well as standard pitched orchestral instruments. Stimuli were tested both in separate stimulus sets and in a composite set, in order to assess whether the perceptual structures of the separate sets would be preserved in a broader timbral context.

The potential effect of musical training on timbral classification schemes was also examined by testing groups of both professional musicians and nonmusicians, respectively. The number of studies examining the relation between musical expertise and timbre perception is limited. Using speeded and unspeeded classification tasks, Pitt (1994), for example, found that nonmusicians had greater difficulty attending to pitch in the face of timbral variation than did musicians, indicating an overreliance on timbre information when categorizing sounds that vary along both dimensions. Whether musical training differentiates processing of timbre alone remains an open question. It would seem reasonable to assume that years, or decades, of professional experience with musical instrument sounds would give rise to an increasingly sophisticated representation of timbre and its attributes. A finding that musical training has little, if any, effect on listeners' representation of timbre, on the other hand, would support the notion that the basic dimensions of timbre are somehow perceptually primary and relatively immune to alteration through experience. Since musical training has rarely served as an independent variable in past studies of timbre, its inclusion in the present study was expected to shed some preliminary light on this area.

A secondary aim of the present research was to determine whether listeners also grouped or categorized timbres on the basis of common instrumental characteristics or modes of excitation. To investigate this possibility,

similarity data were also analyzed using an extended additive tree model, *EXTREE* (Corter & Tversky, 1986), which can detect both nested and overlapping groupings of timbres. In order to encourage potential grouping on the basis of ecological factors, the musical instruments used to generate the stimulus sets sampled a wide range of source properties and modes of excitation, focusing on those combinations that might elicit categorical perception of source properties. In sum, then, the present study was designed to apply a broad range of recently developed analytic techniques to further isolate the perceptual attributes of timbre.

METHOD

Participants

Thirty-four participants (23 men, 11 women) between the ages of 18 and 40 served as participants. Eighteen of them (17 men, 1 woman) were professional musicians recruited at IRCAM; all had the highest levels of training in the analysis and synthesis techniques currently employed in computer music. This participant group will be referred to as *musicians*. The remaining 16 individuals (6 men, 10 women), henceforth called *nonmusicians*, were Yale students recruited from a group of participants who had participated in psychophysical studies at the John B. Pierce Laboratory. The participants who reported 1 year or less of training on a musical instrument and who were not currently receiving such training were invited to participate. A 1-year allowance for instrumental training was given, because few Yale students reported never having studied a musical instrument in the past. To allow only those without any instrumental training at all in their past to be labeled nonmusicians would have artificially restricted eligible participants to a small minority of students; these students could potentially have been unrepresentative of most students at Yale in other respects, as well. Nine of the 16 participants reported having studied a musical instrument for 1 year or less; none had any formal training in music theory or orchestration. None reported any hearing problems. All participants in the study received monetary compensation for their participation.

Stimuli

Stimuli were 34 sounds selected from the *McGill University Master Samples* (MUMS) compact disk (Opolko & Wapnick, 1987) of digitally recorded musical instruments (see Table 1). An additional sound (tam-tam) was recorded at IRCAM, to bring the total stimulus set to 35. The recording quality of the large tam-tam sample available in the MUMS collection was judged to be unsuitable for the study. The first set of 17 tones were produced by traditional pitched instruments (e.g., flute, trumpet, piano) playing at D#4. For identification purposes, these instruments will henceforth be referred to collectively as the *harmonic* set. The second (*percussive*) set consisted of 18 percussion instruments. Of these, 7 instruments were pitched at D#4 in MUMS (i.e., celesta, marimba, steel drum, tubular bells, tympani, vibraphone [bowed], vibraphone [struck]); the remainder were weakly pitched (e.g., bowed cymbal, log drum) or unpitched (e.g., tam-tam, bamboo chimes). The principal criterion for stimulus selection was that the instrumental sounds should span a broad range of physical sources and manners of excitation. Figure 1 highlights the general acoustic categories of resonance and excitation into which the harmonic and percussive instruments fall: (1) physical sources comprise strings, air columns, bars, plates, membranes, and tubes/ blocks, with additional amplification for strings, bars, and membranes arising from acoustically coupled resonance cavities, and (2) modes of excitation can be continuous (i.e., blown,

Table 1
Instrument Recordings Used in the Present Study

Instrument	MUMS Reference			Centroid (Hz)	Time to Maximum	
	Volume	Track	Index		Amplitude (msec)	Duration (msec)
	Harmonic					
Baroque recorder*	11	60	05	580	88	1,500
B♭ Clarinet*	02	10	14	1950	51	1,500
Tenor crumhorn	11	40	01	2950	35	1,500
English horn	02	09	12	1800	44	1,500
Flute (no vibrato)	09	86	04	1000	234	1,500
Flute flutter-tongued*	02	02	04	600	560	1,500
French horn*	02	19	26	770	75	1,500
Harp*	09	22	04	400	22	1,500
Harpichord*	11	95	06	1470	30	1,500
Piano*	03	02	43	1100	26	1,500
Pipe organ	10	06	03	1480	59	1,500
Alto saxophone	03	16	03	1230	53	1,500
Tenor saxophone (growls)*	08	69	01	2150	66	1,500
B♭ Trumpet	07	16	02	2300	42	1,500
C Trumpet (muted)*	02	17	10	2500	211	1,500
Violin (no vibrato)	09	63	03	1450	65	1,500
Violin (martelé)*	01	05	10	850	50	656
	Percussive					
Bamboo chimes	03	12	28	3900	37	922
Bongo drum	03	11	30	1750	4	845
Castanets	03	12	35	3600	6	1,500
Celesta*	09	38	04	800	20	1,500
Cuica*	03	12	49	850	102	1,500
Cymbals (bowed)*	04	83	11	4900	678	1,500
Cymbals (struck)*	04	83	01	3700	1	1,500
Log drum*	03	12	57	1100	15	951
Marimba	03	04	23	900	78	1,500
Snare drum*	03	11	09	1400	6	897
Steel drum*	03	10	01	900	2	1,500
Tambourine (pop)	03	12	45	7800	5	934
Tam-tam		IRCAM recording		2000	63	1,500
Temple block	03	12	29	1500	7	1,327
Tubular bell*	03	10	04	900	12	1,500
Tympani*	04	49	01	800	10	1,500
Vibraphone (bowed)*	03	07	11	900	339	1,500
Vibraphone (struck)	03	06	11	850	83	1,500

*Instruments also used in the combined stimulus set.

bowed) or impulsive (i.e., struck, plucked) vibration. Additional acoustic factors, such as obstructions (e.g., the mute of the trumpet) or noise components (e.g., flutter-tonguing, growls), were selected to increase the variety in the harmonic stimulus set.

Sound samples were transferred in direct digital format (44.1 kHz) to the hard disk of a NeXT computer. In order to maintain a constant interstimulus onset interval, all sounds were edited to a length of 2,000 msec, either by removing portions of the steady states of sounds with greater duration or by adding an appropriate length of silence to the ends of shorter sounds. A 50-msec linear decay ramp was imposed at the end of each sample for those sounds whose length had originally exceeded 2,000 msec. Ten expert listeners equalized the loudnesses of the 36 sounds by adjusting the intensity of a comparison sound using a sliding scale on a computer screen until its loudness matched that of a referent, the English horn. Averages of these adjustments were then used to either amplify or attenuate each sound digitally. In a similar fashion, seven expert listeners equalized the pitches of all clearly pitched sounds (i.e., all of the harmonic sounds and 7 of the percussive) by means of a matching paradigm in which they selected one sound from among 20 pitch-altered variations, arranged on a computer screen in ten 5-cent steps above and below the original sampled sound, that most closely matched the

pitch of the reference sound (English horn). In addition, two expert listeners and the author made coarse adjustments to the relative pitches of several weakly pitched instruments (i.e., cuica, bowed cymbal, struck cymbal, log drum, temple block) to bring them close in pitch to other, more strongly pitched instruments. Listeners' pitch adjustments were then averaged and applied to the loudness-matched sounds. The final adjusted stimulus set was highly uniform in terms of pitch and loudness.

Apparatus

The reproduction of the sounds, the presentation and timing of stimuli, and the entering of responses were controlled by the psychoacoustic test program *PsiExp* (B. K. Smith, 1994) designed at IRCAM and run on a NeXT computer with integrated digital-to-analog converters. Compact-disk recordings of the MUMS sounds were transferred to the NeXT computer using a Sony Model CDP-295 CD player and a ProPort Model 656 Stereo Audio-DSP Port Interface. Loudness equalization of the MUMS recordings was implemented using a custom version of the *PsiExp* program. Pitch calibration was performed using a pitch-shifting program called *Foo* (Eckel & Gonzalez-Arroyo, 1994), which is based on a band-limited interpolation technique described by J. O. Smith and Gossett

RESONATOR

		String	Air Column		Bar	Plate	Membrane	Tube/Block
Amplification:		plate & resonance cavity			air column		resonance cavity	contains own resonance cavity
Continuous, forced vibration	Blown		air jet	flute, recorder organ				
			lip reed	trumpet French horn				
			single reed	English horn alto sax				
			double reed	clarinet crumhorn				
	Bowed	violin			vibraphone	cymbal	cuica	
Impulsive	Struck	piano			celesta marimba vibraphone	cymbal tam-tam steel drum	bongo drum long drum snare drum tambourine tympani	bamboo tubes castanets temple block tubular bell
	Plucked	harpsichord harp						
	Noise components		flutter-tonguing growls				snares, rattles	
	Obstruction		mute					

Figure 1. The general acoustic categories of resonance and excitation into which the harmonic and percussive instruments fall.

(1984). The algorithm computes an output signal from a set of discrete time samples of the input signal amplitudes to change the pitch and, simultaneously, the duration of a sound.

The tam-tam sound used in the present study was recorded in an anechoic chamber using two Schoeps microphones (Colette series MK6) and a Sony 670 DAT recorder (sampling rate of 48 KHz, 16-bit resolution). The tam-tam itself had a diameter of 98 cm and was struck with a heavy felt mallet. The resulting sound was down-sampled to a rate of 44.1 kHz and transferred to the hard disk of the NeXT computer.

The musicians were tested in an anechoic room at IRCAM. Converted sounds were amplified by means of a Yamaha P2075 power amplifier and reproduced on Yamaha NS-1000M loudspeakers situated in approximately 4 m in front of the participant. The nonmusicians were tested in an IAC single-walled sound isolation booth at the John B. Pierce Laboratory (no anechoic chamber was available for replicating the testing conditions at IRCAM). Sounds for these participants were reproduced on Yamaha MDR-V6 Digital Stereo Headphones connected directly to the integrated headphone output port on the NeXT. The confound between musical experience and stimulus presentation was not deemed important in light of the lack of difference between musicians and nonmusicians obtained here; one can therefore conclude that both musical experi-

ence and mode of presentation were not significant factors in the present experimental context.

Procedure

A similarity rating technique was employed. On each trial, the participants heard two sounds presented in sequence and separated by a 2-sec pause. The participants rated their timbral similarity by adjusting the position of a continuous sliding switch on a similarity scale presented on a computer screen (the switch was positioned at the center of the scale at the beginning of each trial). The left and right endpoints of the scale were labeled *very similar* and *very different*, respectively. The scale comprised approximately 500 discrete positions between these endpoints. The participants were instructed to use the full range of the similarity scale. An unlimited number of stimulus repetitions per trial was permitted.

The participants judged timbral similarity for three stimulus sets, including identity trials, with each set presented during a different experimental session: (1) harmonic sounds (instruments 1–17 in Table 1), (2) percussive sounds (instruments 18–35 in Table 1), and (3) a selection of 10 sounds from each of the previous two sets (indicated by asterisks next to the relevant instruments in Table 1). Sessions were administered in counterbalanced order, except that the combined set always followed the other two; although such a

block design is not fully counterbalanced, it was important for all participants to have equal exposure to the harmonic and percussive sounds prior to the combined set in order to reduce the effect of differential stimulus familiarity in this condition.

Before each session, the participants were permitted to listen for approximately 10 min to the appropriate stimulus set and to make comparisons between pairs or groups of sounds by means of a computer program for playing digitized sounds in sequence. Following this listening period, the participants received 10 practice similarity judgments before the experimental session. In order to compensate to a degree for their limited familiarity with many of the sounds in the stimulus sets, nonmusicians were given a separate 1-h training session during which they were provided the opportunity to listen to each of the three stimulus sets using the sound player program; they then performed 50 practice similarity judgments per set.

After the practice session, the participants completed either 170 trials for experimental sessions involving the harmonic and percussive sets or 210 trials for the combined set. Trials comprised each possible pairing of sounds presented in a randomized order. The three sessions each lasted approximately 1 h.

Psychophysical Analyses

CLASCAL. The data from the 33 participants were analyzed with the CLASCAL model (Winsberg & De Soete, 1993). CLASCAL represents an extension of the INDSCAL or weighted Euclidean distance model (see Carroll & Chang, 1970), in which the distance function between two stimuli per dimension is weighted separately for each of a total of N participants. Instead of assigning weights to individual participants—a process that adds a substantial number of parameters to the scaling model that are rarely beneficial in interpreting a particular solution—CLASCAL assumes that each participant belongs to one and only one of a small number, T , of latent classes or subpopulations of participants, where $T \ll N$, and that the distance function is weighted equally by all participants belonging to a given latent class. In this model, the distance between stimulus j and j' for latent participant class t is given by:

$$d_{jj't} = \left[\sum_{r=1}^R w_{rt} (x_{jr} - x_{j'r})^2 + v_t (s_j + s_{j'}) \right]^{\frac{1}{2}}, \quad (1)$$

where w_{rt} is the weight given by latent class t to dimension r , $(x_{jr} - x_{j'r})$ is the distance between stimulus j and j' along dimension r , s_j and $s_{j'}$ are the specificities associated with stimuli j and j' , respectively, and v_t is the weight given by latent class t to the whole set of specificities ($v_t \geq 0$). When $T = N$, the CLASCAL model is computationally equivalent to that of INDSCAL; when $T = 1$, it is equivalent to the classical Euclidean model proposed by Torgerson (1958). The CLASCAL model removes the rotational invariance of the solution and retains psychologically meaningful dimensions (as does INDSCAL), thus increasing ease of interpretation. The appropriate number of dimensions is determined by finding the lowest value for the information criterion *BIC* (see Schwartz, 1978) for the spatial model. The *BIC* statistic is an index of parsimony in model building based on the asymptotic behavior of Bayes estimators. The preferred spatial model indicated by *BIC* is then used in a Monte Carlo significance testing procedure (Hope, 1968), which determines the optimal number of latent classes.

In addition to its ability to detect latent classes of participants, the CLASCAL model also provides the option of computing a separate specificity measure unique to each stimulus (s_j above). The specificity measure s_j can be conceptualized as the sum of squares of coordinates along those dimensions specific to object j , all of which have nonzero coordinates only for object j . The specificity represents a measure of the “uniqueness” of each stimulus, in that the variance component associated with each specificity is not shared by any other stimulus in the multidimensional space. Specificity measures are useful to obtain for data that are not easily accommodated by a spatial model characterized exclusively by R common

dimensions. When all s_j are constrained to be zero and $T = N$, the CLASCAL model is functionally equivalent to INDSCAL.

CLASCAL analyses were performed on the combined data of musicians and nonmusicians in order to determine whether they behaved as distinct groups in their judgments of timbral similarity. The latent class approach served as a metric to determine whether musicians and nonmusicians differed significantly in their weighting of the Euclidean timbral dimensions. In order to select an appropriate spatial model, the number of latent participant classes was first determined separately for each of the three stimulus sets. Next, the number of common dimensions by which to represent each stimulus space was selected, and the decision whether or not to include specificity measures in our spatial model was made according to the information criterion *BIC* (Schwarz, 1978), derived from maximum likelihood estimations. On the basis of the model selected, the CLASCAL program calculated the coordinates of each timbre along each common dimension, the specificity of each timbre, and the combined specificities for each latent class. The positions of the timbres in each of the three stimulus spaces were then plotted graphically.

EXTREE. In order to examine possible categorical relations among timbres that are not easily represented in a spatial model, the CLASCAL analyses were supplemented by submitting the timbral dissimilarity matrices to the EXTREE computer program (Corter & Tversky, 1986). The extended tree structure generated by EXTREE is an evolution of the additive tree (Sattath & Tversky, 1977) in that it is able to represent both nested and nonnested features within proximity data. An additive, or ultrametric, tree can be considered a feature tree, in which the length of each arc in the tree represents the features shared by all stimuli that follow from that arc. A feature is nested either if any two clusters of objects representing these features are disjoint or if one includes the other. The EXTREE algorithm first constructs an additive tree and then attempts to detect additional nonnested or overlapping clusters of features.

The ability to represent nonnested features is particularly useful for data possessing a nominal factorial structure. For example, similarity data obtained for a 2×2 factorial structure representing the product of two categories of familial relations (e.g., *mother–father* and *daughter–son*) cannot easily be represented by an additive or ultrametric tree, because each factor does not form a unidimensional array (e.g., *mother* clusters with *father* but is also proximal to *daughter*). In this case, EXTREE is able to represent nonnested clusters such as *mother–daughter* by placing marked segments along the arcs corresponding to features contained within such clusters. When two or more stimuli share a marked segment, the distance covered by the segment does not enter into the computation of the path-length distance between the stimuli. In this light, EXTREE interpretations for timbral data would be useful in cases in which participants judge the timbral similarity on several nonorthogonal factors (e.g., categorical distinctions between physical source properties), not easily represented in a hierarchical or additive tree, or in a spatial model of low dimensionality.

Centroid calculations. In order to develop psychophysical interpretations for our analyses, the SNDAN sound analysis/synthesis package (Beauchamp, 1993) was used to calculate the spectral centroid and rise time of each of our stimuli. SNDAN computes centroid values for harmonic spectral components using the following formula (Beauchamp & Horner, 1995):

$$\text{centroid} = f_a \left[\frac{\sum_{k=1}^N A_k F_k}{\sum_{k=1}^N A_k} - 1 \right], \quad (2)$$

where k is the harmonic number, N equals the number of harmonics in the calculations, A_k is the amplitude of harmonic k , and f_a is the mean fundamental frequency. This computation is performed over a series of fixed time frames, so that the resulting centroid is a function of time.

Table 2
Log Likelihood and Values of Information Criterion
BIC for Spatial Models Derived From Similarity
Ratings for 33 Subjects in Comparisons of
Harmonic, Percussive, and Combined Timbral Stimuli

Dimension	Without Specificities		With Specificities	
	LogL	BIC	LogL	BIC
Harmonic				
1	-40	239	1,897	-3,482
2	1,143	-1,982	2,042	-3,629
3	1,681	-2,917	2,057	-3,517
4	1,911	-3,234	2,148	-3,557
5	2,126	-3,520		
6	2,102	-3,330		
Percussive				
1	239	-308	2,364	-4,396
2	2,209	-4,095	2,812	-5,138
3	2,807	-5,136	2,931	-5,222
4	2,876	-5,126	3,000	-5,207
5	3,028	-5,217		
6	3,061	-5,185		
Combined				
1	724	-1,256	3,366	-6,356
2	2,520	-4,674	3,667	-6,783
3	3,475	-6,408	3,749	-6,772
4	3,707	-6,697	3,815	-6,730
5	3,796	-6,700		
6	3,954	-6,841		

Note—The model with the lowest value for the BIC criterion (in bold-face) is considered to be the most appropriate.

RESULTS

Harmonic Set

BIC statistics suggested a CLASCAL space with two dimensions and with specificities (Table 2). Monte Carlo simulation chose two latent classes for this data set (Table 3). Class 1 comprised 11 musicians and 12 non-musicians, and Class 2 comprised 6 musicians and 3 nonmusicians (1 participant was ambiguous with respect to class), indicating that there was no real difference in class belongingness with respect to musical training. The participants in Class 2 weighted the two dimensions, as well as the specificities, of the model space more heavily than did those in Class 1, suggesting that the space captures a greater portion of the variance associated with the timbral ratings of those in Class 2.

The positions of the timbral stimuli within the two-dimensional CLASCAL space are presented graphically in Figure 2. The defining characteristic of the space is the segregation of instruments on the basis of transient properties: Instruments within the left side of the graph are excited continuously (e.g., French horn, tenor sax) and produce timbres consisting primarily of steady-state signals, whereas instruments on the right side are excited impulsively (e.g., harp, piano) and possess stronger transient characteristics. Correspondingly, the positions of timbres along Dimension 1 of the harmonic space correlate significantly and inversely ($r = -.621$) with the logarithm of their rise times (Table 4). Dimension 2, on the

other hand, correlates strongly with the logarithm of the timbres' spectral centroid: There is an orderly progression from timbres with relatively low spectral centroids at the top of Figure 2 (e.g., flute, harp) to those with high centroids (e.g., muted trumpet, tenor crumhorn). The strong psychophysical correlation between the rise times/centroids of timbres and their dimensional coordinates reinforces the findings of previous studies that these two acoustic measures correlate with the principal perceptual dimensions of timbre.

The specificities for the harmonic timbres are listed in Table 5, along with those for the combined set. High specificities reflect distinctive properties of individual stimuli that are not easily represented within the dimensions of a continuous space. Table 5 reveals that timbres with relatively high specificity values (e.g., $>.100$) tend to be ones that have noise components or other nonharmonic characteristics (e.g., the flutter-tongued flute and the tenor sax growls) or ones that have unusual modes of excitation or distinctive source features relative to the other stimuli (e.g., the violin [martelé] and the muted trumpet, respectively). The violin (martelé) has a high specificity since its spectral characteristics are proximal to those of the violin (no vibrato), but its duration allies it more closely with the plucked and struck strings (harp, piano, harpsichord). There is a region between the continuous and the transient timbres in which no timbres are situated (with the possible exception of the violin [martelé]), suggesting that there may be a categorical distinction between these two types of sounds (i.e., no natural timbre can occupy a position intermediate between continuously excited vs. impulsive categories, since it would be acoustically implausible).

Although specificities provide a measure of the extent to which individual timbres possess unique features, they do not indicate where certain subsets of timbres share selected features. The CLASCAL algorithm cannot distinguish between one or many specific dimensions for each stimulus and thus cannot determine whether certain specific dimensions are shared by a small subset of stimuli.

Table 3
Estimated Weights for Latent Subject Classes
in the CLASCAL Spatial Models for Harmonic,
Percussive, and Combined Stimulus Sets

Class	Dimension			Specificities
	1	2	3	
Harmonic				
1	0.91	0.69	-	0.71
2	1.10	1.31	-	1.29
Percussive				
1	0.86	0.80	0.67	-
2	1.14	1.20	1.33	-
Combined				
1	0.82	0.85	-	0.67
2	1.18	1.15	-	1.33

Note—Weights are given separately for each dimension of the model and for specificities (where appropriate).

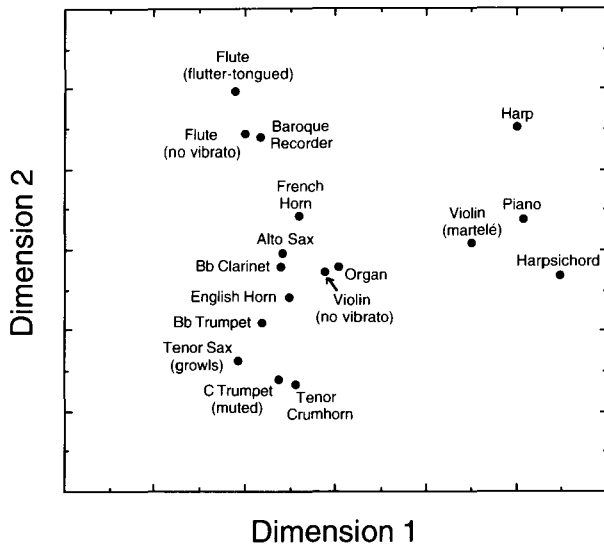


Figure 2. Two-dimensional CLASCAL space for the harmonic timbre set.

In an attempt to detect categorical features that may cut across the continuous dimensions of the CLASCAL model, the harmonic stimulus set was analyzed with the EXTREE program. The resulting tree structure is shown in Figure 3; the marked feature pattern matrix at the right side of the figure organizes the common marked segments along various branches of the tree into vertical columns so that shared features can be more readily interpreted. The major division of the tree corresponds to the distinction between impulsive and continuous sounds and complements the separation between these groups of stimuli in the CLASCAL space. Beyond this overall separation, and contrary to initial expectations, timbres do not cluster strongly according to physical commonalities among the instruments producing the timbres, nor do the nonnested marked segments in the solution reflect any clear categorization of source properties. Rather, features appear to correspond to more proximal spectral and temporal similarities in the respective waveforms of the timbres that are not reflected in their locations within the CLASCAL space. For example, Features 3, 4, 6, and 7 link timbres with similar spectral centroids that are otherwise separated by their transient properties. Feature 1 appears to group timbres with dissimilar centroids (flute, tenor sax, violin [martelé]) on the basis of their distinctive modes of excitation. The remaining features are more difficult to characterize: Feature 5 may reflect a mutual “softness of tone” among the baroque recorder, French horn, and alto sax, and Feature 2 links the organ and the harpsichord to the main cluster of brass and woodwinds with moderate to high centroids. In sum, nonhierarchical features within the harmonic space tend to be those that reflect proximal characteristics, rather than the distal physical-acoustic commonalities among subsets of sounds.

Percussive Set

A three-dimensional CLASCAL space was obtained for the percussive stimuli. Although BIC statistics indicated that specificities should be included, the dissimilarity matrices for musicians and nonmusicians were analyzed separately using EXSCAL (Winsberg & Carroll, 1989)—an MDS program algorithmically identical to CLASCAL except that it does not compute latent classes—and three-dimensional spaces without specificities were obtained for both groups. Therefore, a CLASCAL space without specificities was used on the assumption that any specificities would essentially represent variance arising from small differences between the ratings of the two participant groups. Two latent classes were found: Class 1 consisted of 6 musicians and 10 musicians, whereas Class 2 represented 11 musicians and 6 musicians. Thus, there was a tendency to have more nonmusicians in Class 1 and more musicians in Class 2, although the separation by musical training is not complete. As with the harmonic set, the participants in Class 2 weighted all dimensions of the percussive space, particularly Dimension 3, more heavily than those in Class 1. Perhaps, in this case, the greater exposure musicians have had to percussive sounds permits them to focus to a greater extent on spectral and temporal commonalities among timbres, sounds that might otherwise sound quite different to the untrained ear.

The three-dimensional percussive space is shown in three two-dimensional perspectives in Figures 4A, 4B, and 4C. Similar to the CLASCAL model for the harmonic set, Dimensions 1 and 2 of the percussive space again correlate strongly and negatively with the logarithms of attack time ($r = -.702$) and centroid ($r = -.897$), respectively (see Table 3). This finding is remarkable in light of the extremely broad range of timbres selected for this stimulus set and seems to suggest that, at some fundamental level of auditory processing, timbres as disparate as those pro-

Table 4
Correlations Between the Logarithms of the Rise Time and Centroid for Timbres and Their Coordinates Along Each Dimension (D) of the CLASCAL Spaces for Harmonic (Harm.), Percussive (Perc.), and Combined (Comb.) Stimulus Sets

	Correlation	<i>p</i>
Harm. D1 vs. Log(Rise Time)	-.621	.0066*
Harm. D2 vs. Log(Rise Time)	.312	.2275
Perc. D1 vs. Log(Rise Time)	-.702	.0007*
Perc. D2 vs. Log(Rise Time)	.328	.1871
Perc. D3 vs. Log(Rise Time)	.098	.7020
Comb. D1 vs. Log(Rise Time)	.751	<.0001*
Comb. D2 vs. Log(Rise Time)	.282	.2316
Harm. D1 vs. Log(Centroid)	-.292	.2610
Harm. D2 vs. Log(Centroid)	-.916	<.0001*
Perc. D1 vs. Log(Centroid)	.018	.9449
Perc. D2 vs. Log(Centroid)	-.897	<.0001*
Perc. D3 vs. Log(Centroid)	.065	.8010
Comb. D1 vs. Log(Centroid)	-.240	.3132
Comb. D2 vs. Log(Centroid)	.749	<.0001*

*Significant correlation.

Table 5
Specificity Measures for Timbre
in the Harmonic and Combined Sets

Instrument	Specificity
Harmonic Set	
B \flat Clarinet	.076
Tenor crumhorn	.000
English horn	.021
Flute plutter-tongued)	.100
Flute (no vibrato)	.005
French horn	.096
Harpsichord	.004
Harp	.004
Organ	.263
Piano	.026
Baroque recorder	.032
Alto saxophone	.043
Tenor saxophone (growls)	.096
B \flat Trumpet	.045
C Trumpet (muted)	.117
Violin (martelé)	.408
Violin (no vibrato)	.089
Combined	
B \flat Clarinet	.057
Celesta	.028
Cuica	.140
Flute (plutter-tongued)	.151
Cymbal (bowed)	.188
French horn	.050
Harpsichord	.069
Harp	.000
Cymbal (struck)	.087
Piano	.048
Baroque recorder	.000
Long drum	.119
Tenor saxophone (growls)	.083
Snare drum	.141
C Trumpet (muted)	.105
Violin (martelé)	.267
Steel drum	.035
Tubular bells	.192
Tympani	.206
Vibraphone (bowed)	.041

Note—The specificity magnitude for a given timbre indicates the extent to which it possesses unique attributes not accounted for by the dimensions of the appropriate CLASCAL model.

duced by a cuica, a tam-tam, and bamboo chimes are organized and represented primarily within the context of spectral central tendencies and attack characteristics.

Dimension 3, as with previous studies of timbre, proved difficult to interpret psychophysically. The positions of timbres along this dimension correlate with neither centroid nor attack, and there appears to be little intuitive structure to their sequence; for example, the cuica, a bowed membranophone, and the marimba, a struck steel bar have the same coordinates along Dimension 3 as do the bowed cymbal, the tubular bells, and the tympani. Spectral flux, representing the standard deviation of the time-averaged amplitudes of the partials (see Krimphof et al., 1994), did not correlate significantly with Dimension 3 ($r = .14$).

In contrast to the harmonic space, the EXTREE solution for the percussive space shows a consistent segregation

of instruments along the lines of physical source characteristics and manners of excitation (Figure 5). Timbres are organized in three principal clusters: (1) wooden/metal bars and tubes (celesta, marimba, vibraphone, tubular bells), (2) metal plates (tam-tam, cymbals), (3) a large cluster encompassing membranophones (tympani, snare drum, bongo drum), instruments with wood cavities (log drum, temple block) or other more complex resonator shapes fabricated in large part from wood (castanets, bamboo chimes, tambourine rattles). Two instruments, the steel drum and cuica, cluster outside of these groups, probably due to their hybrid physical structures/excitation modes (the steel drum nonetheless clusters close to the membranophones, perhaps reflecting that it shares the property of a metallic resonating cavity with the tympani). In addition to the principal clusters, several local clusters further differentiate instruments on the basis of source properties. The bamboo chimes and tambourine, for instance, appear to cluster because of the presence of multiple small source components in each instrument; the castanets also cluster nearby since they also consist of small interacting components. Similarly, the metal and wooden bars cluster closely and are thus removed from the nonetheless acoustically related structure of the tubular bells (i.e., bar vs. tube), and the temple block, log drum, and bongo cluster apart from the tympani, snare drum, and steel drum, probably due to their primarily wooden construction.

Several of the marked segments in Figure 5 reflect commonalities in manners of excitation across otherwise disparate clusters, particularly in terms of continuous versus impulsive sounds. Marked Feature 2 denotes the similar bowed style of the vibraphone and cuica, and Feature 7 further links these two instruments to the bowed cymbal. Also, the struck cymbal, tambourine, bamboo chimes, and castanets share a sharp, high-frequency attack (Features 4 and 6). Other features highlight common source properties that cut across clusters. The marimba, log drum, and temple block, for instance, are all constructed from wood (Feature 1), even though their resonator shapes are substantially different. Features 5 and 8 link the bowed cymbal to the struck cymbal and tam-tam. Feature 3 is somewhat more difficult to interpret: The log drum is struck in a gentle manner similar to the struck/bowed bars (all five instruments have approximately the same coordinate along the verbal dimension of playing effort).

In sum, the multidimensional scaling and clustering solutions for the percussive stimuli point to the participants' use of two interrelated criteria to compare timbres. First, the participants judged timbres primarily according to two orthogonal perceptual dimensions: one correlating with the spectral center of gravity, and the other correlating with transient properties of the signal (specifically, rise time). A third dimension did not appear to correlate with any easily interpretable acoustical correlate, although this does not mean that it does not exist. These first two dimensions correspond to those found for

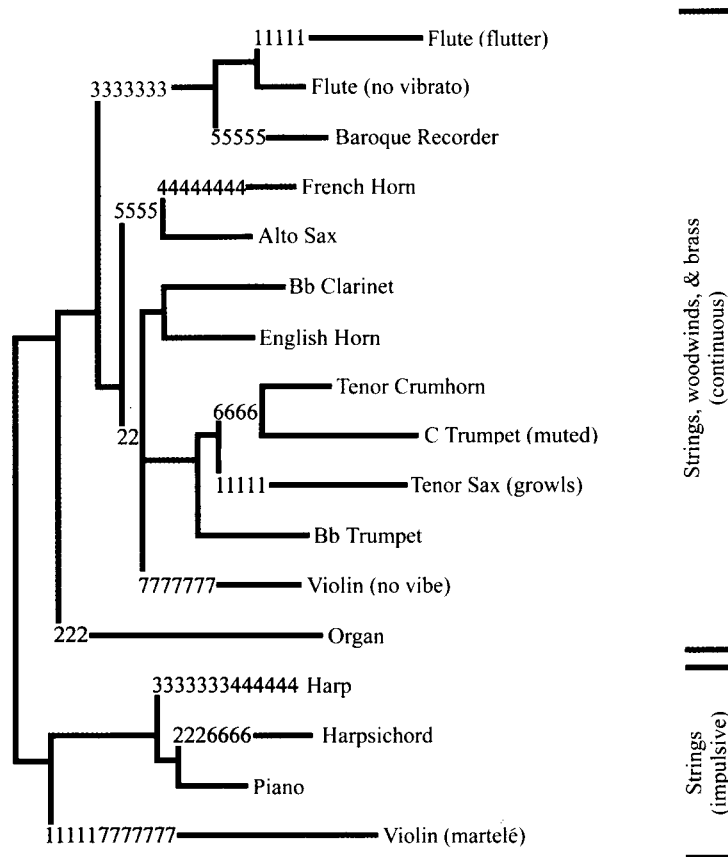


Figure 3. EXTREE clustering solution for the harmonic timbre set. Numbers indicate shared features (e.g., the flutter-tongued flute and tenor sax growls share Feature 1). The greater the number of digits used to indicate a marked segment, the greater the strength of the shared feature (e.g., the shared feature 3333333 has a greater magnitude than the shared feature 6666).

the harmonic stimuli and, when taken in conjunction with findings from previous studies, indicate that, at the most basic level of analysis, our perceptual representation of timbre is bidimensional. Second, the participants were also able to interpret the percussive timbres in terms of shared physical source characteristics and modes of excitation. This points to the mental representation of properties, such as the material and shape of an instrument, as well as the mass, density, and striking force of an exciter such as a mallet. That source characteristics played a lesser role in distinguishing timbres in the harmonic set may have been due to the restricted range of exciter and resonator properties represented in those sounds.

Combined Set

The CLASCAL algorithm selected a two-dimensional space with specificities for the combined set. The Monte Carlo simulation rejected both a one- and a two-class interpretation for the data set, but CLASCAL did not permit us to go higher than two classes given the number of participants in our study. We therefore opted for a two-class interpretation in light of these restrictions, with the

caveat that the interpretation is suboptimal. Seven musicians and 10 nonmusicians composed Class 1, and 8 musicians and 3 nonmusicians composed Class 2, so there was a tendency to have more nonmusicians in Class 1 and roughly equal numbers of musicians in both classes. Five participants were ambiguous with respect to class. Again, the two classes do not correspond strongly to differences in musical training. As with the previous two stimulus sets, the participants in Class 1 weighted both dimensions (and specificities) less strongly than did the participants in Class 2.

Figure 6 shows the CLASCAL space for the combined set. Not surprisingly, the logarithms of each timbre's rise time and spectral centroid correlate with Dimension 1 ($r = .751$) and Dimension 2 ($r = -.749$), respectively. This reinforces the trend seen in the earlier scaling solutions that these two acoustic parameters account for a large portion of the variance for a wide variety of timbres. The harmonic and percussive stimulus subsets in Figure 6 are differentiated primarily with respect to attack time, with the percussive stimuli having the relatively shorter rise times, and fall into separate regions within the

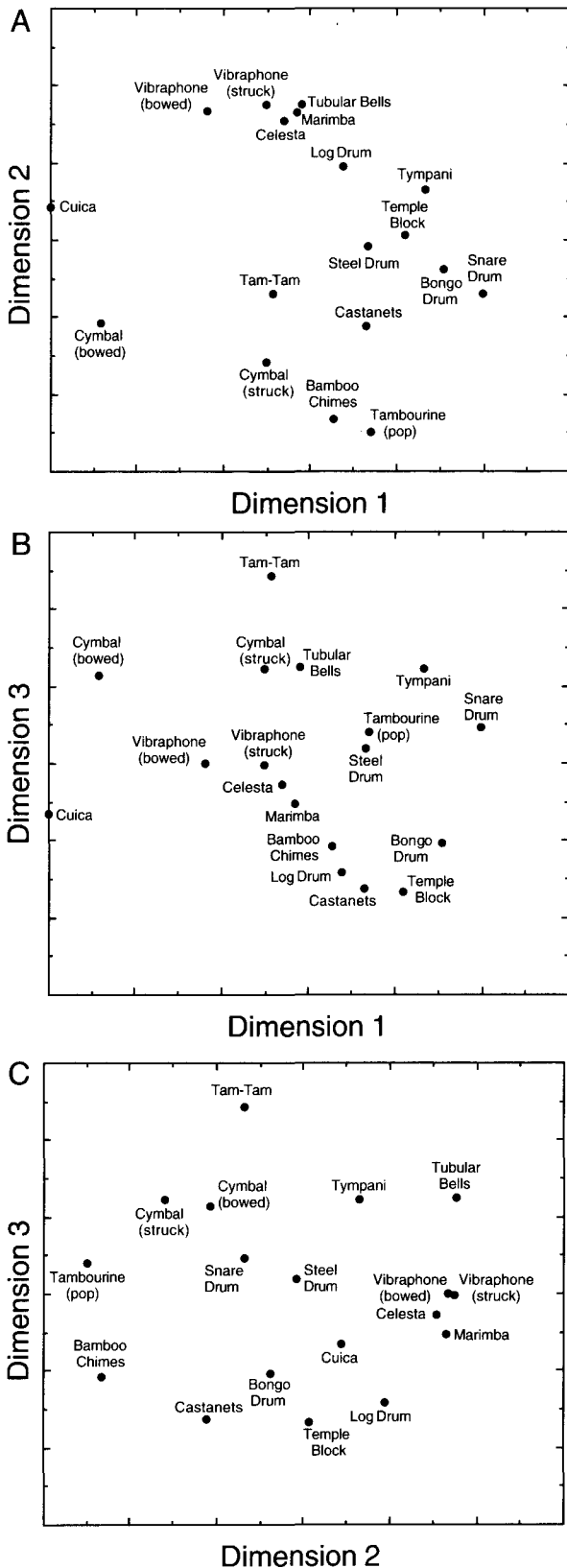


Figure 4. Three-dimensional CLASCAL space for percussive stimuli. Dimension 3 correlates significantly with the participants' VAME ratings for timbral "richness."

space. There is nonetheless a partial overlap between the two subsets, particularly among the string instruments (piano, harp, harpsichord, violin [martelé]) and the percussive instruments with bar and tube resonators (celesta, bowed vibraphone, tubular bells), probably due to the inherently greater similarities in the physical structure of these instruments than in any of the others. The cuica also groups with the trumpet and tenor sax, despite its very different exciter-resonator attributes. Other instruments, such as the cymbals and drums, lie quite far from the main cluster of traditional pitched instruments (e.g., clarinet, flute, French horn). The cymbals, in particular, define the outer boundaries of timbres with high centroids and sharp attacks. The pattern of specificities for the combined set is similar to that of the harmonic set (see Table 4): Instruments with atypical excitations or noise components (e.g., the flutter-tongued flute, bowed cymbal, violin [martelé], and cuica) tend to have higher specificities. The percussive sounds generally have slightly higher specificities, although there is no sharp distinction in this respect between the two subsets.

The EXTREE solution for the combined set is displayed in Figure 7, and indicates a clustering of instruments based on shared physical attributes. There appears to be a range of physical-acoustic properties represented in the four main clusters: (1) the relatively simple resonators (bars, strings, struck tubes), (2) tubes whose air columns are excited by blowing, (3) drums, including both the membranophones and the steel drum, and (4) the metal plates, with relatively complex resonators. Within these main groups, further differentiation occurs. The two bar instruments, the celesta and vibraphone, cluster together closer than they do with the harp; by the same token, the piano and the harpsichord cluster somewhat away from the tubular bells. Similar to the EXTREE results for the harmonic space, no strong, systematic grouping by physical sources among the blown tubes was found. Although the air jet instruments (flute and baroque recorder) cluster closely, neither the lip reeds (trumpet and French horn) nor the single/double reeds (clarinet and tenor sax) do so. In fact, the French horn is most closely linked to the cuica, a bowed membrane, which further illustrates that among the traditional brass and woodwind instruments, similarities among what one might call "surface characteristics" of timbre seem to take precedence over any underlying similarities in instrument properties. The violin (martelé) is again the odd timbre out in this set, clustering separately from the four main groups and with the highest specificity value (.267). The participants both in the combined condition and in the harmonic condition reported great difficulty in deciding whether to group this stimulus with the other strings or with certain percussive sounds (e.g., the drums).

DISCUSSION

The results reported here show that similarity relations among acoustically diverse timbres can be characterized adequately by spectral centroid and rise time alone. Stimulus sets consisting of traditional pitched or-

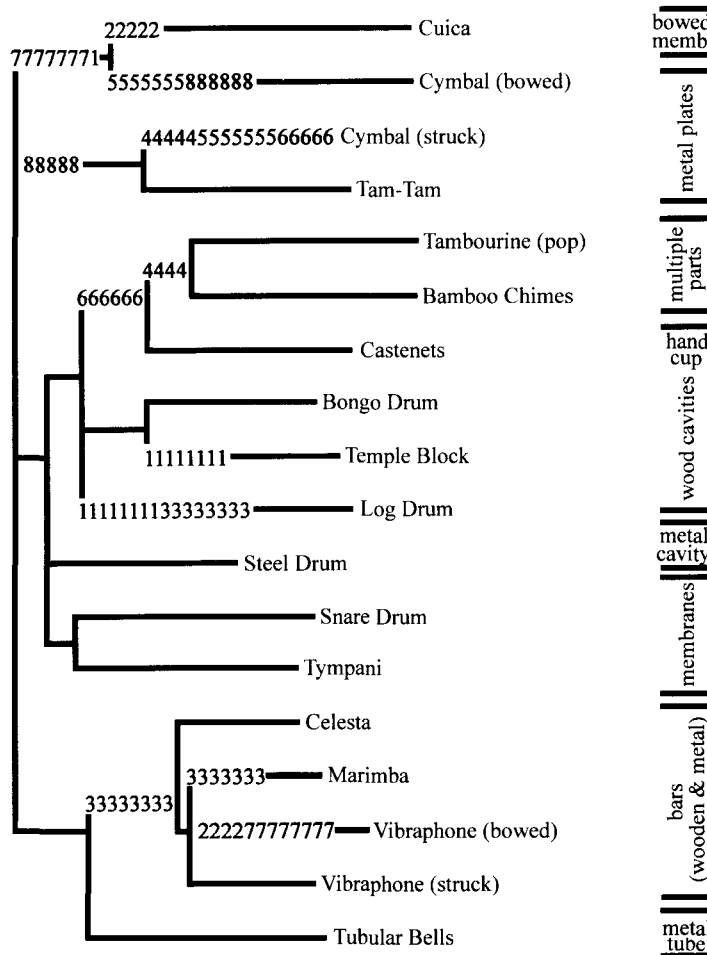


Figure 5. EXTREE clustering solution for percussive timbre set.

chestral instruments, percussive instruments, and their combination all yielded perceptual spaces with spectral centroid and rise time as acoustic correlates when the variance associated uniquely with each timbre (i.e., specificity) was parceled out by the CLASCAL algorithm. For the percussive space, a third dimension associated with timbral “richness” was also found, but it did not correlate significantly with either centroid or rise time. The present study extends past findings by demonstrating that timbres considerably more exotic and diverse than the traditional pitched instruments used in previous MDS studies can nonetheless be modeled within a similar perceptual space.

An additional finding of the present work was that there was little systematic difference in the way in which musicians and nonmusicians weighted the dimensions of the CLASCAL spaces. There is some evidence that musicians weighted all of the dimensions of the percussive space more heavily than did nonmusicians, but there is no evidence that the two groups weighted them differentially; this suggests that, because of their superior training, musicians used the same dimensions as nonmusicians, only to a greater extent. The absence of a significant

training effect suggests that the dimensions composing timbre are somehow perceptually primary and therefore cannot be altered substantially by experience.

A secondary expectation of the present study was that the greater variety of acoustic source properties and modes of excitation in the percussive set, as well as in the combined set, would prompt both musicians and nonmusicians to group timbres according to such shared characteristics. The present study provides evidence that listeners do so, although some caution in interpreting the results is advisable since the source/excitation groupings were not always mutually exclusive. There is a fairly orderly arrangement of acoustic sources at a superordinate level, ranging from the simplest resonators (e.g., bars, tubes) to the moderately complex (e.g., wood or metal cavities) to the highly complex (e.g., metal plates, sources with multiple components). Also, a distinction between continuous and impulsive sounds was made across clusters (e.g., the bowed vibraphone is linked to the bowed cuica and cymbal). At a more basic level, sources also clustered according to material type (e.g., wood, metal, membrane) and more specific shape features (e.g., temple block and

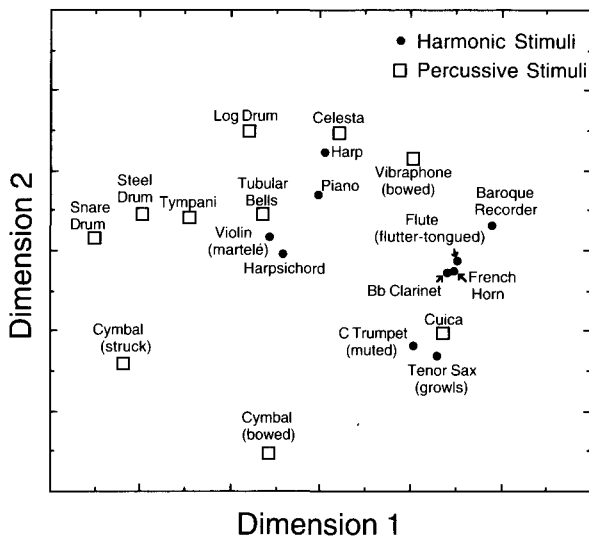


Figure 6. CLASCAL space for combined timbre set. Harmonic stimuli are represented by filled circles; percussive stimuli are represented by open squares.

bongo drum vs. log drum). The reappearance of most of these clusterings in the combined data set strengthens the argument that they not only arise for a specific stimulus set but are replicable across multiple contexts.

Given consistent findings about the dimensional structure of timbre in this work and in others, to what extent does a two-dimensional structure represent a useful descriptive model? From an intuitive perspective, it may seem unsatisfying to accept that two orthogonal dimensions capture most of the variance inherent in our rich acoustic environment, much as it would seem reductionistic to characterize the wide range of visual objects in our environment exclusively by length, width, and height. One potential explanation may stem from the fact that MDS algorithms, including CLASCAL, almost invariably generate low-dimensional solutions because they seek the most parsimonious dimensional fit to the data. It may be that higher dimensions of timbre exist but that the number of timbres (or the variance inherent within the data) would need to be increased in order to extract those higher dimensions. It may also be that attributes of timbre beyond those of spectral centroid and rise time may not be captured adequately by a continuous spatial model, especially if such "higher" attributes arise from nominal/categorical aspects of the sound sources that produce the timbres (such nominal attributes would be represented by increased variance within the MDS solution, and not necessarily by additional dimensions unique to each attribute).

Indirect evidence for additional, perhaps noncontinuous, timbral attributes arises when one generates synthetic timbres that vary exclusively along the dimensions derived from MDS studies. The resulting sounds fail to capture much of the range of timbral variation occurring

in real musical instruments or other sound sources and tend to sound artificial and hollow. Students of computer music, who have attempted to synthesize electroacoustic sounds by manipulating their time-varying spectrum along dimensions similar to those suggested by MDS studies, are familiar with how difficult it can be to produce a realistic-sounding timbre. In a study of timbre using newly developed MDS techniques, McAdams, Winsberg, Donnadieu, De Soete, and Krimphoff (1995) found that musical timbres possess specific attributes—some deriving from the implied method of physical source excitation—that cannot be accommodated by a model postulating shared perceptual dimensions, suggesting that a purely dimensional interpretation of timbre perception may mask other noncontinuous or categorical factors. Thus, it would appear that many of the acoustic components necessary to convey a realistic impression of a physically generated sound may not be represented adequately by a two-dimensional model.

Can additional MDS analyses provide further insight into the acoustic determinants of timbre? As noted above, the answer depends on whether attributes of timbre that are not captured by a two-dimensional model can be represented by additional dimensions or whether such attributes are nominal and therefore are not easily represented by a Euclidean spatial metric. If additional timbral dimensions exist, one strategy would be to select stimuli that have similar spectral centroid and rise times but are perceptually dissimilar; such a strategy would reduce variation along the two principal dimensions of timbre and therefore increase the likelihood that additional dimensions can emerge from the data. Careful preselection of stimuli on this basis might represent a better strategy than, say, increasing the heterogeneity of a stimulus set by adding more stimuli, given that the number of paired comparisons increases geometrically with the number of stimuli presented (e.g., 20 stimuli result in 380 pairings, whereas 36 stimuli would have resulted in 1,260 pairings and exhausted participants). If, on the other hand, additional timbral attributes are nondimensional, it may be necessary to move from the purely descriptive model of timbre that MDS provides to a more predictive, hypothesis-driven approach that attempts to link acoustical properties of sounds and their sources to their perceptual consequences.

An important starting point for the development of a more predictive model would be the articulation of a theory of proximal and distal stimulus processing for timbre perception. Although one can describe timbral attributes in terms of microstructural features (e.g., time-varying amplitudes of individual partials), it is likely that listeners also attend to more categorical acoustic invariants, such as the mechanical characteristics of the instruments themselves (e.g., the coupling of the reed to the air column in a clarinet) or physical commonalities shared by particular instrumental families. Such an ecological component of timbre perception was suggested by Gibson (1966), who theorized that physical processes, such as a

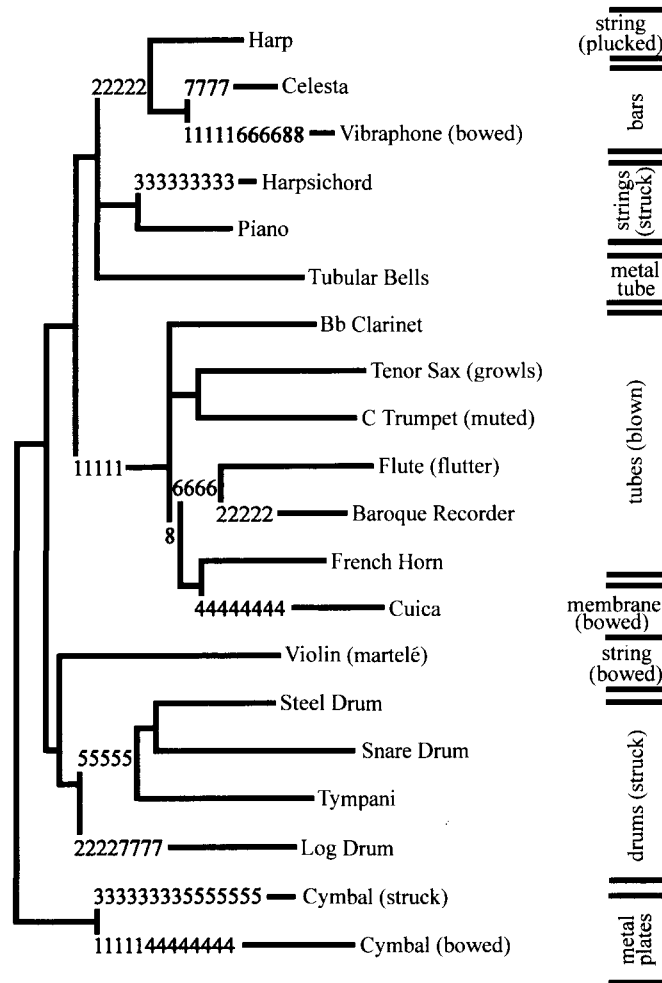


Figure 7. EXTREE clustering solution for the combined timbre set.

mallet striking a bar, are specified in the underlying dynamics of the resulting acoustic signal and therefore represent environmental invariants that are available to the listener. Among several recent studies providing an ecological view of auditory perception, Lakatos, McAdams, and Caussé (1997) found that listeners' ability to distinguish between metal and wood bars of equal length varied directly with their increasing width/thickness ratio, as well as with their spectral centroids, and that listeners could attend to the characteristic vibrational modes of the bars (i.e., torsional, transverse, longitudinal) in the process of discriminating their geometric shapes. Such evidence suggests that, if indeed listeners use the spectral and temporal characteristics of sounds in timbral comparisons, they do so in the context of acoustically relevant features of the sources themselves.

Far from being mutually exclusive, however, the mechanisms required to parse the principal perceptual attrib-

utes of a sound and to infer its sound properties must be complementary, since the primary role of the auditory system is to provide a veridical representation of sound events and sources in the environment. MDS studies, combined with more hypothesis-driven approaches, can therefore continue to contribute to our understanding of timbre by helping to tease apart the different stages of processing that ultimately give rise to a cognitive representation of sound sources.

REFERENCES

BEAUCHAMP, J. W. (1993). *Unix workstation software for analysis, graphics, modification and synthesis of musical sounds* (Audio Engineering Society Preprint No. 3479). New York: Audio Engineering Society.

BEAUCHAMP, J. W., & HORNER, A. (1995). *Wavetable interpolation synthesis based on time-variant spectral analysis of musical sounds* (Audio Engineering Society Preprint No. 3960). New York: Audio Engineering Society.

- CARROLL, J. D., & CHANG, J. J. (1970). Analysis of individual differences in multidimensional scaling via an N-way generalization of Eckart-Young decomposition. *Psychometrika*, **35**, 283-319.
- CORTER, J. E., & TVERSKY, A. (1986). Extended similarity trees. *Psychometrika*, **51**, 429-451.
- ECKEL, G., & GONZALEZ-ARROYO, R. (1994). Musically salient control abstractions for sound synthesis. In *Proceedings of the 1994 International Computer Music Conference* (pp. 256-259). Aarhus: International Computer Music Association.
- GIBSON, J. J. (1966). *The senses considered as perceptual systems*. Boston: Houghton-Mifflin.
- 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). Perceptual effects of spectral modifications on musical timbres. *Journal of the Acoustical Society of America*, **63**, 1493-1500.
- HOPE, A. C. (1968). A simplified Monte Carlo significance test procedure. *Journal of the Royal Statistical Society: Series B*, **30**, 582-598.
- IVERSON, P., & KRUMHANS, C. L. (1993). Isolating the dynamic attributes of musical timbre. *Journal of the Acoustical Society of America*, **94**, 2595-2603.
- KRIMPHOFF, J., MCADAMS, S., & WINSBERG, S. (1994). Characterisation du timbre des sons complexes: II. Analyses acoustiques et quantification psychophysique [Characterization of the timbre of complex sounds: II. Acoustic analyses and psychophysical quantifications]. *Journal de Physique*, **4**, 625-628.
- KRUMHANS, C. L. (1989). Why is musical timbre so hard to understand? In S. Nielzen & O. Olsson (Eds.), *Structure and perception of electroacoustic sound and music* (pp. 43-53). Amsterdam: Elsevier.
- LAKATOS, S., MCADAMS, S., & CAUSSÉ, R. (1997). The representation of auditory source characteristics: Simple geometric form. *Perception & Psychophysics*, **59**, 1180-1190.
- MCADAMS, S., WINSBERG, S., DONNADIEU, S., DE SOETE, G., & KRIMPHOFF, J. (1995). Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent participant 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*, **58**, 711-720.
- OPOLKO, F., & WAPNICK, J. (1987). *McGill University master samples* [Compact disk]. Montreal: McGill University.
- PIIT, M. A. (1994). Perception of pitch and timbre by musically trained and untrained listeners. *Journal of Experimental Psychology: Human Perception & Performance*, **20**, 976-986.
- SATTATH, S., & TVERSKY, A. (1977). Additive similarity trees. *Psychometrika*, **42**, 319-345.
- SCHWARZ, G. (1978). Estimating the dimensions of a model. *Annals of Statistics*, **6**, 461-464.
- SMITH, B. K. (1994). *Psiexp, version 1.3: A psychacoustic experiment environment for the NeXT computer* [Computer program]. Paris: IRCAM.
- SMITH, J. O., & GOSSETT, P. (1984). A flexible sampling-rate conversion method. In *Proceedings of the IEEE Conference on Acoustics, Speech and Signal Processing* (Vol. 2., pp. 19.4.1-19.4.2). New York: IEEE Press.
- TORGERSON, W. S. (1958). *Theory and methods of scaling*. New York: Wiley.
- WEDIN, L., & GOUDE, G. (1972). Dimension analysis of the perception of instrument timbre. *Scandinavian Journal of Psychology*, **13**, 228-240.
- WINSBERG, S., & CARROLL, J. D. (1989). A quasi-nonmetric method for multidimensional scaling via an extended Euclidean model. *Psychometrika*, **54**, 217-229.
- WINSBERG, S., & DE SOETE, G. (1993). A latent class approach to fitting the weighted Euclidean model, CLASCAL. *Psychometrika*, **58**, 315-330.

(Manuscript received January 9, 1998;
revision accepted for publication December 31, 1999.)