Psychol Res (1995) 58: 177–192


ORIGINAL ARTICLE

Stephen McAdams · Suzanne Winsberg · Sophie Donnadieu · Geert De Soete · Jochen Krimphoff

# Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes

Received: 6 September 1994 / Accepted: 18 June 1995

**Abstract** To study the perceptual structure of musical timbre and the effects of musical training, timbral dissimilarities of synthesized instrument sounds were rated by professional musicians, amateur musicians, and nonmusicians. The data were analyzed with an extended version of the multidimensional scaling algorithm CLASCAL (Winsberg & De Soete, 1993), which estimates the number of latent classes of subjects, the coordinates of each timbre on common Euclidean dimensions, a specificity value of unique attributes for each timbre, and a separate weight for each latent class on each of the common dimensions and the set of specificities. Five latent classes were found for a three-dimensional spatial model with specificities. Common dimensions were quantified psychophysically in terms of log-rise time, spectral centroid, and degree of spectral variation. The results further suggest that musical timbres possess specific attributes not accounted for by these shared perceptual dimensions. Weight patterns indicate that perceptual salience of dimensions and specificities varied across classes. A comparison of class structure with biographical factors associated with degree of musical training and activity was not clearly related to the class structure, though musicians gave more precise and coherent judgments than did nonmusicians or amateurs. The model with latent classes and specificities gave a better fit to the data and made the acoustic correlates of the common dimensions more interpretable.

Some of these results were reported at the Third French Conference on Acoustics, Toulouse (Donnadieu, McAdams, & Winsberg, 1994)

S. McAdams (✉)
Laboratoire de Psychologie Expérimentale (CNRS), Université René Descartes, EPHE, 28 rue Serpente, F-75006 Paris, France, and Institut de Recherche et de Coordination Acoustique/Musique (IRCAM), 1 Place Igor-Stravinsky, F-75004 Paris, France

S. Winsberg
IRCAM 1 Place Igor-Stravinsky, F-75004 Paris, France

S. Donnadieu
Laboratoire de Psychologie Experimentale (CNRS) and IRCAM, Paris, France

G. De Soete
Department of Data Analysis, University of Ghent, Henri Dunantlaan 1, B-9000 Ghent, Belgium

J. Krimphoff
IRCAM, Paris, France

## Introduction

Timbre, in contrast to pitch and loudness, remains an auditory attribute, which has been until recently, poorly understood from a psychophysical standpoint. In fact until about 25 years ago, timbre was considered to be a perceptual parameter of sound that was simply complex and multidimensional, defined primarily by what it wasn't: what distinguishes two sounds presented in a similar manner and being equal in pitch, subjective duration, and loudness (American Standards Association, 1960; Plomp, 1970). This multidimensionality makes it impossible to measure timbre on a single continuum such as low to high, short to long, or soft to loud, and raises the problem of determining experimentally the number of dimensions and features required to represent the perceptual attributes of timbre and of characterizing those attributes psychophysically.

Multidimensional scaling (MDS) has been a fruitful tool for studying the perceptual relationships among stimuli and for analyzing the underlying attributes used by subjects when making (dis)similarity judgments on pairs of stimuli (Kruskal, 1964 a,b; Shepard, 1962 a,b). The object of MDS is to reveal relationships among a set of stimuli by representing them in a low-dimensional (usually Euclidean) space so that the distances among the stimuli reflect their relative dissimilarities. To achieve this representation, dissimilarity data arising from $N$ sources, usually subjects, each relating $J$ objects pairwise, is modeled by one of a family of

MDS procedures to fit distances in some type of space, generally Euclidean or extended Euclidean of low dimensionality $R$. The different dimensions are then interpreted as psychologically meaningful attributes that underlie the judgments. An important distinction among different MDS techniques (which we discuss below) is the kind of spatial model used to represent the distances between pairs of stimuli.

Multidimensional scaling of (dis)similarity judgments has been the tool of predilection for exploring the perceptual representation of timbre (e.g. Plomp, 1970; Miller & Carterette, 1975; Grey, 1977; Krumhansl, 1989; Kendall & Carterette, 1991). There are several reasons for this choice: (1) the judgments are relatively easy to make for subjects; (2) the technique makes no apriori assumptions about the nature of the dimensions that underly the perceptual representation used by subjects to compare the timbres of two sound events; (3) the resulting geometric representation of the data can be readily visualized in a spatial model; and (4) the spatial model has been found to have predictive power (Grey & Gordon, 1978; Ehresman & Wessel, 1978; Kendall & Carterette, 1991; McAdams & Cunibile, 1992).

The object of the present paper is to illustrate the use of a new MDS technique in the study of musical timbre. This new technique provides a means for determining a parsimonious number of psychologically meaningful dimensions common to all stimuli as well as dimensions specific to individual stimuli, and to assign the sources (subjects) to a small number of latent classes. Hence in contrast with previous studies of musical timbre, a large number of subjects with widely varying musical experience were employed. Since maximum likelihood estimation was used to determine the parameters of the model, statistical tests were employed to select both the number of latent classes and the appropriate spatial model, including the number of psychologically meaningful common dimensions and whether to include specific dimensions.

Below we present brief surveys of the different MDS distance models and their use in the study of musical timbre. We then present an experimental study of the timbre of complex, synthesized sounds using the new technique. In addition to providing further support for the psychophysical interpretation of certain primary dimensions of musical timbre, this study has three facets that further advance our knowledge of timbre perception: (1) the use of complex, synthetic sounds designed either to imitate acoustic instruments or to create perceptually interpolated hybrids between such instruments; (2) the estimation of specific attributes (denoted specificities) possessed by individual sounds not accounted for by the common dimensions of the Euclidean spatial model; (3) the estimation of latent classes of subjects and the comparison of this class structure with degree of musical training and activity.

## Multidimensional scaling analysis

*Distance models.* In the classical MDS model the objects are assumed to possess collectively a small number of psychological attributes. The classical model was proposed by Torgerson (1958) and Gower (1966) and used by Shepard (1962 a,b) and Kruskal (1964 a,b). This type of model is implemented in programs such as MDSCAL and KYST for nonmetric MDS. The Euclidean distance, $d_{jj'}$, between the stimuli $j$ and $j'$ is given by

$$d_{jj'} = \left[ \sum_{r=1}^{R} \left( x_{jr} - x_{j'r} \right)^2 \right]^{\frac{1}{2}} \tag{1}$$

where $x_{jr}$ is the coordinate of stimulus $j$ on the dimension $r (j, j' = 1, \ldots, J)$.

In this model, the distance between a pair of stimuli does not depend on the data source or subject. In the classical model, the choice of axes is arbitrary as the model distance does not depend upon this choice. Thus this model is rotationally invariant. In the weighted Euclidean model, however, psychologically meaningful dimensions are postulated. These common dimensions are weighted differently by each source or subject. That is, it is assumed that each dimension has a different salience for each source or subject. The INDSCAL, or weighted Euclidean distance, model proposed by Carroll and Chang (1970) removes the rotational invariance that exists in the classical Euclidean distance model. The distance, $d_{jj'n}$, between stimuli $j$ and $j'$ for source $n$ in the weighted Euclidean model is given by

$$d_{jj'n} = \left[ \sum_{r=1}^{R} w_{nr} \left( x_{jr} - x_{j'r} \right)^2 \right]^{\frac{1}{2}} \tag{2}$$

where $x_{jr}$ is the coordinate of stimulus $j$ on dimension $r$ $(j = 1, \ldots, J)$, and $w_{nr}$ is the weight for dimension $r$ associated with source $n$ $(n = 1, \ldots, N; w_{nr} \geq 0)$.

Since psychologically meaningful dimensions are postulated in the weighted Euclidean model and these dimensions are weighted differently by each source or subject, rotational invariance is removed. The lack of rotational invariance makes the interpretation much easier for the user, since it is often difficult to find psychologically interpretable dimensions (by rotation in $R$ space), if the classical distance model is postulated and the dimensionality of the space exceeds two.

In addition to sharing the common dimensions, the stimuli may differ in ways that are specific to each one. A spatial model that is more appropriate in this case (postulating both common dimensions and specificities) has been proposed and tested on several data sets (see Bentler & Weeks, 1978; De Leeuw & Heiser, 1980; Takane & Sergent, 1983; Winsberg & Carroll, 1989a; and De Soete, Carroll, & Chaturvedi, 1993). To distinguish this model from the classical (Euclidean, spatial) MDS model, Winsberg and Carroll (1989a) called it the extended two-way Euclidean model with common and

specific dimensions, or simply the "extended two-way model," for short. In this model, the distance between stimuli $j$ and $j'$ is given by

$$d_{jj'} = \left[ \sum_{r=1}^{R} \left( x_{jr} - x_{j'r} \right)^2 + s_j + s_{j'} \right]^{\frac{1}{2}} \tag{3}$$

where $s_j$ is the square of the coordinate of stimulus $j$ along the dimension specific to that stimulus (one cannot distinguish mathematically between one or many such specific dimensions). This model may be thought of as a hybrid between a Euclidean spatial model and an additive tree (Sattath & Tversky, 1977). The specificity may be thought of as the square of the perceptual strength of a feature possessed by the stimulus. The $s$s may, of course, be constrained to be zero, making the standard Euclidean model in $R$ dimensions a special case of the extended model.

An extension of the common dimensions and specificities model to a weighted Euclidean model was developed by Winsberg and Carroll (1989b). In this last model the distance between stimuli $j$ and $j'$ for source $n$ is given by

$$d_{jj'n} = \left[ \sum_{r=1}^{R} w_{nr} \left( x_{jr} - x_{j'r} \right)^2 + v_n (s_j + s_{j'}) \right]^{\frac{1}{2}} \tag{4}$$

where $v_n$ is the weight given by source $n$ to the whole set of specificities ($v_n \geq 0$).

*Latent class approach.* The modeling of individual differences that results in the rotational uniqueness of the object coordinates is most likely the reason for the popularity of the weighted Euclidean model among users. The $N$ different sources are usually the $N$ subjects and the data consist of dissimilarity judgments from these subjects. In this case (or in any case where $N$ is large), the cost of removing rotational invariance, to obtain ease of interpretation, is the introduction of many nuisance parameters (the individual subject weights $w_{nr}$ and $v_n$). In practice, these weights are rarely interpreted for each individual subject. The weighted Euclidean model is not generally retained because goodness-of-fit measures are radically improved (thus justifying so many additional parameters), but rather because its dimensions are meaningful psychologically. Therefore, Winsberg and De Soete (1993) proposed a latent-class approach to this problem. This approach has the advantage that rotational invariance is removed since the distance model posits psychologically meaningful dimensions, but the number of parameters is reduced considerably.

In the latent-class approach, it is assumed that each of the $N$ subjects belongs to one and only one of a small number of latent classes or subpopulations. The classes are latent because it is not known in advance to which one a particular subject belongs. We postulate $T$ latent classes (with $T \ll N$). The (unconditional) probability that any subject belongs to latent class $t$ is denoted $\lambda_t (1 \leq t \leq T)$, with:

$$\sum_{t=1}^{T} \lambda_t = 1. \tag{5}$$

It is also assumed that for a subject $n$ in latent-class $t$, the data $y_n$ (where $y_n$ is a $J(J-1)/2$-dimensional vector of dissimilarities for source $n$) are independently normally distributed with means $d_t = (d_{21t}, d_{31t}, d_{32t}, \ldots, d_{j(j-1)})$ and common variance $\sigma^2$. Sometimes the dissimilarities from each source or subject are arranged in the lower triangle of a matrix. Here the data for each source are presented as a vector and the entire data set is a $N \times J(J-1)/2$-matrix $\mathbf{Y}$. In the CLASCAL model proposed by Winsberg and De Soete (1993), the distance between stimuli $j$ and $j'$ in latent class $t$ is given by

$$d_{jj't} = \left[ \sum_{r=1}^{R} w_{tr} \left( x_{jr} - x_{j'r} \right)^2 \right]^{\frac{1}{2}} \tag{6}$$

where $w_{tr}$ is the INDSCAL-type weight for latent class $t (w_{tr} \geq 0)$.

To identify fully the latent class-weighted Euclidean model, some constraints must be applied. First, the latent class weights for a given dimension (for $r = 1, \ldots, R$) are constrained to sum up to the number of classes:

$$\sum_{t=1}^{T} w_{tr} = T. \tag{7}$$

Secondly, the coordinates for a given dimension are constrained to sum to zero:

$$\sum_{j=1}^{J} x_{jr} = 0 \tag{8}$$

The first constraint normalizes the weights and the second one centers the solution. The latent class-weighted Euclidean distance model has $T + 1 + J \times R + T \times R$ parameters corresponding to $\lambda$ (class structure vector), $\sigma^2$ (variance parameter), $\mathbf{X}$ (stimulus configuration matrix), and $\mathbf{W}$ (weight matrix), respectively. By subtraction from the number of model parameters, the number of constraints imposed on these parameters via equations 5, 7 and 8, the degrees of freedom of the model are obtained:

$$T + (T + J - 2) \cdot R. \tag{9}$$

When $T = 1$, it is necessary to subtract $R(R-1)/2$ from equation 9 for the rotational indeterminacy that occurs in this case.

For each latent class $t$, a separate set of weights $w_t$ is estimated. These weights are constrained to be nonnegative. The stimulus configuration $\mathbf{X}$ (a $J \times R$ matrix) and the variance parameter $\sigma^2$ are assumed to be the same for all latent classes. Since we do not know in advance to which latent class a particular subject

$n$ belongs, the probability density function of $y_n$ becomes a mixture of multivariate normal densities. Estimates of the parameters $\mathbf{X}$, $\mathbf{W}$ ($a$ $T \times R$ matrix), $\sigma^2$, and $\lambda$ (a $T$ vector) are obtained by maximizing the likelihood function. As in many mixture problems (McLaughlin & Basford, 1988), the likelihood function is most easily optimized by means of an EM (expectation-maximization) algorithm (Dempster, Laird, & Rubin, 1977; for a description of the likelihood function as well as the steps involved in the present application of the EM algorithm, see Winsberg and De Soete, 1993). Once parameter estimates for $\mathbf{X}$, $\mathbf{W}$, $\sigma^2$, and $\lambda$ are obtained, the a-posteriori probability that subject $n$ belongs to latent class $t$ is computed by means of Bayes' theorem. The subject is assigned to that class for which the a-posteriori probability is greatest. In general, this probability is close to one for one of the classes for each subject.

In this paper we also present the application of an extended CLASCAL model which allows for both common dimensions and specificities. In this extended CLASCAL model the distance between stimuli $j$ and $j'$ for latent class $t$ is given by

$$d_{jj't} = \left[ \sum_{r=1}^{R} w_{tr} \left( x_{jr} - x_{j'r} \right)^2 + v_t (s_j + s_{j'}) \right]^{\frac{1}{2}} \tag{10}$$

where $v_t$ is the extended INDSCAL-type weight for the specificities for latent class $t$ ($v_t \geq 0$). To obtain the number of degrees of freedom in the extended CLASCAL model, one must add $J + (T - 1)$ to the degrees obtained for the ordinary CLASCAL model to account for the specificities and their weights.

Latent-class formulations, or more general mixture distribution approaches, have also been used in the context of various uni- and multidimensional-scaling models for paired-comparison data (Bockenholt & Bockenholt, 1990; De Soete, 1990; De Soete & Winsberg, 1993; Formann, 1989), for data obtained in a "pick any $n$ stimuli"-type task (Bockenholt & Bockenholt, 1990; De Soete & De Sarbo, 1991), and for single-preference data (De Sarbo, Howard, & Jededi, 1991; De Soete & Winsberg, 1993; De Soete & Heiser, 1993). In all of these applications, latent-class modeling has proved to be a viable technique for capturing systematic group differences in a parsimonious way.

*Model selection.* In most situations, we do not know the number of latent classes in advance. The usual procedure for deciding on the number of classes involves testing whether a solution for $T + 1$ latent classes gives a significantly better fit than one for the same model with $T$ classes. If a $(T + 1)$-class solution does not significantly improve the solution for $T$ classes, $T$ classes are considered sufficient to describe the data adequately. Unfortunately in the case of finite mixture models, the likelihood ratio statistic for testing $T$ versus $T + 1$ latent classes is not asymptotically distributed as a chi-squared with known degrees of freedom (McLaughlin & Basford, 1988). So likelihood ratio tests and information criteria such as AIC and BIC that rely on the same regularity conditions cannot be used. We therefore use a Monte Carlo significance testing procedure proposed by Hope (1968) and first applied in the context of latent-class analysis by Aitken, Andersen, and Hinde (1981).

The procedure can be summarized as follows. Let $\hat{\theta}_1, \ldots, \hat{\theta}_T$, $\lambda$, denote maximum likelihood estimates of $\theta_1, \ldots, \theta_T, \lambda$ for a $T$-class model, where $\theta_t$ is the parameter vector for class $t$, and $\lambda$ is the class-weight vector. From the $T$-class population with parameters $\hat{\theta}_1, \ldots, \hat{\theta}_T, \lambda$, a number (say $S - 1$) of random Monte Carlo samples $\hat{\mathbf{Y}}$ of size $N$ are drawn. The model is fit with $T$ and $T + 1$ classes for each of these generated samples $\hat{\mathbf{Y}}$ and the likelihood statistic for comparing the $T$-class and $(T + 1)$-class solutions is computed. The $T$-class solution is rejected at significance level $\alpha$ in favor of the $(T + 1)$-class solution, if the value of the likelihood ratio statistic for $\mathbf{Y}$ exceeds $S(1 - \alpha)$ of the values of the statistic obtained for the Monte Carlo samples $\hat{\mathbf{Y}}$. A minimal value of $S$ when a significance level $\alpha = .05$ is used is 20. Hope (1968) showed that the power of the Monte Carlo significance test increases as $S$ becomes larger. We have used $S = 250$ on the null model for paired comparisons in the present study.

One of the advantages of using a maximum likelihood criterion for estimating the model parameters is that it enables statistical model evaluation by means of likelihood-ratio tests and information criteria. Ramsay (1977) was the first to use maximum-likelihood estimation (MLE) in MDS via his program MULTISCALE. Winsberg and Carroll (1989a) also used this criterion in the MDS context. The use of MLE removes the difficulties of choosing an appropriate spatial model using goodness-of-fit measures like stress 1, stress 2, or s-stress (squared stress) and looking for the *elbow* that indicates that the addition of a supplementary dimension does not sufficiently reduce the stress to be worth trying to interpret. In general, this elbow is poorly defined in real data structures. Once the number of latent classes has been determined by means of Hope's (1968) procedure, the appropriate distance model, with or without specificities and with the appropriate number of common dimensions, can be chosen by a comparison of the values of the information criterion. One such criterion is the AIC statistic (Aikake, 1977) which is defined for model $\Omega$ as

$$AIC_\Omega = -2\log \hat{L}_\Omega + 2v_\Omega \tag{11}$$

where $\hat{L}_\Omega$ is the estimate of the likelihood function and $v_\Omega$ is the number of degrees of freedom for model $\Omega$.

The AIC statistic does not take into account sample size and in many situations tends to select a model with too many parameters (see Bogdozan, 1987). The BIC

statistic proposed by Schwarz (1978) takes into account sample size and usually is more parsimonious. In our case (paired-comparisons data), BIC is defined for model $\Omega$ as

$$BIC_{\Omega} = -2\log\hat{L}_{\Omega} + \nu_{\Omega}\log\left(N\frac{J(J-1)}{2}\right). \quad (12)$$

Both statistics explicitly compensate for a goodness-of-fit due to an increased number of model parameters. The model with the smallest value of these statistics is said to give the best representation of the data. From experience with artificial data, Winsberg and De Soete (1993) suggest using the BIC criterion. We shall use this criterion as a basis for model selection here, though AIC values are also reported.

## Multidimensional studies of musical timbre

Plomp was among the first to use the classic Euclidean spatial model for the multidimensional representation of synthesized steady-state spectra derived both from Dutch vowels (Plomp, Pols, & van de Geer, 1967; Pols, van der Kamp, & Plomp, 1969) and from organ-pipe and musical-instrument (wind and bowed-string) tones (Plomp, 1970, 1976). This technique was also used to study the effects on timbre perception of phase relations among frequency components (Plomp & Steenecken, 1969). For vowel and musical-instrument spectra and a variety of phase spectra, three-dimensional solutions were found. A two-dimensional solution was sufficient for organ-pipe spectra. Further, the perceptual dimensionality was much smaller in each solution than the number of degrees of freedom available for the construction of the stimuli. No attempt was made by these authors to interpret quantitatively the psychophysical nature of the individual dimensions. However, the authors computed distance measures on the vector of energy levels in a bank of 1/3-octave filters (a rough estimate of auditory filter bandwidths Zwicker & Scharf, 1965) for the vowel, organ-pipe, and musical-instrument spectra. Analysis of these distances with MDSCAL gave spatial solutions similar to those for the dissimilarity ratings for each stimulus set. This correspondence indicates that the global activity level present in the array of frequency-specific auditory nerve fibers may be a sufficient sensory representation from which a small number of perceptual factors related to the spectral envelope are extracted.

Several studies of recorded musical-instrument tones or of tones synthesized to capture certain acoustic characteristics of instrument tones have obtained two- or three-dimensional spatial solutions. Wedin and Goude (1972) found a clear relation between the three-dimensional perceptual structure of similarity relations among musical-instrument tones (winds and bowed strings) and the spectral-envelope properties. However, whether the tones were presented with the attack portion of the tone or with this portion removed seemed to have only a slight effect on the perceptual structure (the mean dissimilarities for the two conditions were correlated at .92). In one of their experiments on synthesized tones, Miller and Carterette (1975) varied the amplitude envelope (a temporal property), the number of harmonics (a spectral property), and the temporal pattern of onset asynchrony of the harmonics (a spectrotemporal property). They found that the spectral property was represented on two of the three dimensions and that the two other properties combined were organized along the third dimension. These results suggested a perceptual predominance of spectral characteristics in the timbre judgments.

To the contrary, a greater contribution of temporal and spectrotemporal properties has been found by other researchers with recorded wind and bowed-string instrument tones (Grey, 1977; Wessel, 1979; Iverson & Krumhansl, 1993) and with relatively complex synthesized tones meant either to imitate conventional musical instruments (winds, bowed string, plucked strings, mallet percussion) or to represent a hybrid of a pair of these instruments (Krumhansl, 1989). In these studies, one dimension generally seemed to correspond to the centroid of the amplitude spectrum (Grey & Gordon, 1978; Iverson & Krumhansl, 1993; Krimphoff, McAdams & Winsberg, 1994) and another either to properties of the attack portion of the tone (Grey, 1977; Krimphoff et al., 1994) or to properties of the overall amplitude envelope (Iverson & Krumhansl, 1993). The psychophysical nature of the third dimension seemed to vary with the stimulus set used, corresponding either to temporal variations in the spectral envelope (Grey, 1977) or to spectral fine-structure (Krimphoff et al.'s, 1994, analysis of Krumhansl's, 1989, stimuli).

MDS techniques have also been applied to judgments on instrument dyads in which two instruments played either single tones (in unison or at an interval of a musical third) or melodies (in unison or in harmony) (Kendall & Carterette, 1991). The dimensional structures obtained remained relatively stable over the different contexts for the first two dimensions (labeled verbally as *nasality* and *brilliance/richness*), but attempts were not made to characterize the dimensions psychophysically. What this study did demonstrate is that a quasi-linear vector model may be able to explain the perception of timbre combinations on the basis of the dimensional structure of individual timbres, i.e., the position of timbre dyads in a given space can be predicted on the basis of the vector sum of the positions of the constituent timbres. This hypothesis of a vector-like representation has also been applied to the perception of relations between timbres (Ehresman & Wessel,

1978; McAdams & Cunibile, 1992). These studies showed that listeners can to a certain extent make judgments of the similarity of intervals between pairs of timbres on the basis of a representation analogous to a multidimensional vector.

It seems likely that timbre can be defined not only in terms of a certain number of continuous dimensions shared by a set of sound events, but also in terms of distinguishing features or dimensions that may be specific to a given timbre. Only one study to date (Krumhansl, 1989) has tested this notion with an extended Euclidean model (eq. 3; Winsberg & Carroll, 1989a). The sounds tested were synthesized imitations and hybrids of conventional western musical instruments. Judgments of dissimilarity from professional musicians gave rise to a three-dimensional solution, with nonzero specificities on about 60% of the timbres. The three common Euclidean dimensions of this study have been characterized quantitatively by Krimphoff et al. (1994) in terms of rise time, spectral centroid, and irregularity of the spectral envelope. The specificities were quite strong on timbres such as the harpsichord and the clarinet, and especially on some of the hybrid timbres such as the *pianobow* (bowed piano), the *guitarnet* (guitar/clarinet hybrid) and the *vibrone* (vibraphone/trombone hybrid). In some of these cases, it seems obvious that acoustic "parasites" such as the clunk at the end of the harpsichord (the return of the hopper) or the raspy double attack on the vibrone may have been perceived as discrete features distinguishing these sounds from the others in a unique way. The relative perceptual strength of these unique features may have been captured by the specificities in the extended Euclidean model, but they have yet to be systematically related to particular acoustic properties.

Analyses with weighted Euclidean models are also of interest in order to determine whether the weights on different dimensions and specificities correspond to biographical factors such as the level of musical training or cultural origin. Most of the timbre spaces described above were derived exclusively from musician listeners (Wessel, 1979; Grey, 1977; Krumhansl, 1989). A few studies have used individual-differences scaling (INDSCAL) and recruited subjects of varying degrees of musical training (Wedin & Goude, 1972; Miller & Carterette, 1975), but have found no systematic differences in the dimensional weights between subject groups. Serafini (1993), on the other hand, tested two groups of western musician listeners on a set of Javanese percussion sounds (xylophones, gongs, metallophone) and a plucked-string sound. One group had never played or listened to Indonesian gamelan music and the other was composed of people who had played Javanese gamelan for at least two years and had knowledge and experience of Javanese culture. Listeners heard pairs of either single notes or melodies played by these instruments and their dissimilarity judgments were analyzed with INDSCAL. A two-dimensional solution was found, the dimensions of which corresponded to the spectral centroid in the attack portion of the tone (a timbral dimension) and the mean level in the resonant portion of the tone (a dimension more properly characterized as related to loudness). Differences between the two groups were only found for the melodic condition: gamelan players appeared to focus their judgments more on the attack dimension, whereas nonplayers appeared to accord equal weight to the two dimensions.

No studies of musical-timbre scaling have been conducted to date that have employed a large number of listeners of varying levels of musical training with an analysis of latent class structure. Only one study (Krumhansl, 1989) has analyzed the specific weights on timbres. The experiment reported below fills this gap.

## Method

*Subjects.* Ninety-eight subjects were recruited who had varying degrees of musical training, ranging from nonmusicians to professional musicians. Each subject completed a questionnaire at the end of the experiment concerning the amount and kind of musical training (compositional, instrumental, and theoretical) they had received, the number of years of music making (composing, conducting, playing an instrument) in which they engaged, and the amount and type of musical listening they did regularly. The subjects were assigned to one of three groups according to the degree of musical training they had obtained, the number of years of music making in which they had engaged, and their self-identification as one of professional musician, amateur musician, or nonmusician. The 24 composers, performers, and musicologists making a living from music were assigned to the professional group. Their ages ranged from 21 to 55 years ($M = 30$) and they had from 8 to 51 years of music making ($M = 18.0$, $SD = 9.5$). This group included 7 females and 17 males. The amateur group was composed of subjects who identified themselves as amateurs and either had engaged in at least 5 years of music making and still played on at least an occasional basis or had recently taken up music and played on a regular basis. The 46 subjects assigned to this group included music students from City University in London as well as students, staff, and associates of the Institut de Psychologie at the Université René Descartes. Their ages ranged from 18 to 57 years ($M = 22$) and they had from 1 to 18 years of music making ($M = 9.1$, $SD = 4.9$). This group included 27 females and 19 males. The remaining 28 subjects composed the nonmusician group and were students, staff, and associates of the Institut de Psychologie at the Université René Descartes. They had engaged in less than five years of music making, played rarely or not at all, and had no formal music training beyond childhood music lessons. Their ages ranged from 21 to 53 years ($M = 26$) and they had from 0 to 4 years of music making ($M = 0.5$, $SD = 1.1$). This group included 18 females and 10 males. Two subjects identifying themselves as amateurs were classed as nonmusicians and one self-identified nonmusician was classed as amateur. All subjects were paid a token fee for their participation. The subjects were tested at three sites: IRCAM or the Université René Descartes in Paris or City University in London. All subjects were tested under similar conditions.

*Stimuli.* The basic task of the study was to compare the timbres of pairs of complex musical sounds and to rate their degree of dissimilarity. The set of sounds used included 18 of the 21 digitally synthesized instruments developed by Wessel, Bristow, and Settel (1987)

**Table 1** Names, labels in Figure 1, maximum level, and total durations for the 18 sounds used. All sounds had a fundamental frequency of 311 Hz (E-flat4).

| Name (origins of hybrids in parentheses) | Label | Max Level (dBA) | Total Duration (ms) |
|---|---|---|---|
| French horn | hrn | 72 | 569 |
| Trumpet | tpt | 60 | 520 |
| Trombone | tbn | 64 | 563 |
| Harp | hrp | 61 | 707 |
| *Trumpar* (trumpet/guitar) | tpr | 56 | 635 |
| *Oboleste* (oboe/celesta) | ols | 62 | 716 |
| Vibraphone | vbs | 59 | 770 |
| *Striano* (bowed string/piano) | sno | 61 | 775 |
| Harpsichord | hcd | 53 | 521 |
| English horn (cor anglais) | ehn | 67 | 507 |
| Bassoon | bsn | 65 | 495 |
| Clarinet | cnt | 64 | 496 |
| *Vibrone* (vibraphone/trombone) | vbn | 62 | 1096 |
| *Obochord* (oboe/harpsichord) | obc | 63 | 544 |
| Guitar | gtr | 57 | 569 |
| Bowed string | stg | 58 | 1071 |
| Piano | pno | 60 | 1008 |
| *Guitarnet* (guitar/carinet) | gnt | 63 | 557 |
| Mean | | 61.5 | 673 |
| Standard deviation | | 4.1 | 200 |

and employed in the study by Krumhansl (1989)[1] (see Table 1). These sounds were synthesized on a Yamaha TX802 FM Tone Generator with the frequency modulation technique (Chowning, 1973). Twelve of the instruments were designed to imitate traditional western instruments (e.g., trumpet, guitar, vibraphone, bowed string) and six were designed as hybrids of two traditional instruments (e.g. the trumpar aimed to capture perceptual characteristics of both the trumpet and the guitar).

The pitch, subjective duration, and loudness of all these sounds were equalized so that subjects' ratings would only concern the differences in their timbres (see Table 1). The pitch was fixed at E-flat4 (a fundamental frequency of approximately 311 Hz). Two listeners (authors SM and SD) equalized the loudnesses and subjective durations of the sounds by adjustment – independently at first and then by consensus in the case of differences in adjustment. The loudness was adjusted by changing the MIDI[2] velocity value in the synthesizer. This parameter normally controls the intensity and spectrum of the sound as a function of the speed with which a key is struck. The adjusted values varied between 45 and 70 on a scale of 127 to attain an equal impression of loudness when the sounds were played at a mean level of 62 dB SPL. The maximum physical level attained by each sound was then measured at the earphone on a Bruel and Kjaer 2209 sound level meter (A-weighting, fast response) with a flat-plate coupler. The tone durations were adjusted around a mean value of about 670 ms by a change in the duration between the MIDI *note-on* and *note-off* points in the evolution of the tone. The tone starts physically within 1 or 2 ms of the *note-on* in a monophonic situation, and the tone begins to decay more or less rapidly within 1 or 2 ms of the *note-off* command. The actual physical durations required to obtain subjective equality varied between 495 and 1,096 ms because of the various shapes of the onset and offset ramps.

*Procedure.* The experimental session consisted of a familiarization phase, a training phase, and an experimental phase. The subject read the experimental instructions and asked any questions necessary for clarification. Then the 18 sounds were presented in a random order to familiarize the subjects with the range of variation among timbres that was to be rated on a 9-point scale. On each experimental trial, the subject's task was to compare the pairs of instrument sounds and rate directly their degree of dissimilarity on a scale of 1 (very similar) to 9 (very dissimilar). The pair could be played as many times as was desired before the rating was entered into the computer keyboard. Subjects were asked to use the full scale in making their judgments. Fifteen trials were chosen at random from the 153 pairs for each subject to train them in making the dissimilarity rating. Subjects were informed that these ratings would not be included in the analysis. Once this phase was completed, all pairs ($J(J-1)/2 = 153$) of the 18 sounds (excluding identical pairs) were presented for dissimilarity ratings in a different random order for each subject. Each pair was presented once over the course of the experiment and the order of presentation of the sounds within the pair was chosen at random for each subject. Subjects were allowed to take a break at any time during the experimental session, which lasted from 30 to 45 minutes.

The subject was seated in a quiet room in front of the computer. The experiment was controlled by a LISP program running on a Macintosh SE/30 computer which commanded the Yamaha TX802 via a MIDI interface. The stimuli were presented diotically via Sony Monitor K240 earphones connected directly to the output of the synthesizer.

## Results

Each subject's data consisted of a vector of 153 paired comparisons among 18 sounds. The analysis proceeded in two stages. Inter-subject correlations on the dissimilarity matrices were computed and a cluster analysis of the correlations was performed to detect subjects who performed very differently from the others. Data sets that were systematically uncorrelated with all other sets may have indicated subjects who had not adopted a systematic rating strategy or those who misunderstood the instructions. These subjects were eliminated from further analysis. Subsequently, multidimensional scaling of the selected data sets was performed by an extended version of CLASCAL (Winsberg & De Soete, 1993). The number of latent classes was determined by Hope's (1968) procedure and then the appropriate spatial model was selected.

[1] Three timbres were eliminated from Krumhansl's (1989) set in order to reduce the number of comparisons. All three were very close to some other timbre in the Euclidean space of her three-dimensional solution. Two of them (oboe, sampled piano) had specificities of zero and one (bowed piano) had a moderately strong specificity. As such, their removal should not have had much of an effect on the global structure of the space

[2] MIDI = Musical Instrument Digital Interface: an international industry standard for communication between computers and musical instruments that use microprocessors. It includes information about pitch, timing, key velocity and various musical control parameters. For a sound whose spectrum does not vary with the key velocity, the MIDI velocity scale corresponds to a roughly linear loudness scale

**Table 2** Log likelihood, degrees of freedom, and values of information criteria AIC and BIC for spatial models with 5 latent classes of subjects obtained from dissimilarity ratings by 88 subjects on 18 timbres. Values for the null model (no dimensional structure) are shown for comparison

| | Without Specificities | | | | With Specificities | | | |
|---|---|---|---|---|---|---|---|---|
| # Dim. | logL | *df* | AIC | BIC | logL | *df* | AIC | BIC |
| 2 | −23010 | 47 | 46115 | 46468 | −21505 | 69 | 43148 | 43666 |
| 3 | −21546 | 68 | 43228 | 43738 | −20990 | 90 | 42159 | 42835 |
| 4 | −21077 | 89 | 42331 | 42999 | −21054 | 111 | 42331 | 43164 |
| 5 | −20876 | 110 | 41973 | 42799 | | | | |
| 6 | −20735 | 131 | 41732 | 42716 | | | | |
| 7 | −20940 | 152 | 42183 | 43324 | | | | |
| Null | −19666 | 770 | 40872 | 46653 | | | | |

## Cluster analysis

The correlations between the dissimilarity vectors of all pairs of subjects were computed. This vector was submitted to a hierarchical cluster analysis using the nearest neighbor (single link) algorithm. A subset of 10 subjects formed a group of clusters that were clearly isolated from the rest of the subjects. Among these 10, 9 were nonmusicians. The tenth subject was a music student whose data were negatively correlated with the majority of the other subjects, indicating that she had perhaps inverted the rating scale. The data for these 10 outliers were eliminated from the subsequent multidimensional-scaling analysis.[3]

## Multidimensional analysis

*Determination of the number of classes.* The data from the 88 selected subjects were analyzed with the CLASCAL program. Hope's (1968) procedure was used to determine the number of latent classes in the subject population. According to these analyses, five classes were sufficient to account for the data.

*Determination of the number of dimensions and inclusion of specificities.* Selection of the appropriate model for the data set requires a determination of the number of dimensions and whether or not to include the specificities. The parameters for models consisting of from two to seven dimensions without specificities (eq. 6) and from two to four dimensions with specificities (eq. 10) were estimated for five classes of subjects. The BIC values indicated that the most parsimonious model had six dimensions without specificities (see Table 2). The model for three dimensions with specificities was a close contender. The AIC criterion for the 88 subjects selected the null model (mean dissimilarity ratings on all pairs without spatial structure). This result indicates that the data for the entire group were quite noisy. We opted for the three-dimensional solution with specificities because the psychophysical interpretation of the underlying dimensions was more coherent than for the six-dimensional solution (see Discussion) and its BIC value was close to optimal.

For the selected spatial model, the CLASCAL program provides the coordinates of the timbre of each sound along each common dimension (Table 3), the specificity value for each timbre (Table 3), and the weights for each dimension and the set of specificities for each latent class of subjects (Table 4). The positions of the timbres in the three-dimensional space are shown graphically in Figure 1.

*Estimation of class weights on dimensions and specificities.* The weights for each of the three dimensions and the set of specificities in our selected model were estimated for each class (see Table 4). These weights signify that some classes of subjects accorded more importance to certain attributes of timbre in their judgments. Multiplication of the coordinates in Table 3 by the appropriate weights in Table 4 for a given class yields the spatial model for that class. These varying patterns of weights are also what determines the unique orientation of the axes in this model. Classes 1 and 2, which contain the majority of subjects, gave approximately equal weight to all dimensions and the set of specificities, though the weights were slightly higher than the mean for Class 1 and slightly lower than the mean for Class 2. This difference can be attributed to the use of the rating scale since the mean rating for Class 1 was 4.0 and that for Class 2 was 5.5, unpaired $t(304) = -8.73$, $p < .0001$. The other three classes gave less homogeneous patterns of weights which means that the orientation of the axes is primarily determined by the subjects in Classes 3–5. Class 3 weighted dimension 2 quite strongly and the specificities weakly compared to dimensions 1 and 3. Class 4

[3] A separate MDS analysis of the nine nonmusicians with EXSCAL yielded a one-dimensional solution with specificities. The coordinates of the timbres along the lone dimension correlated strongly with the spectral centroid of the tones (see Discussion section). However, the specificity values were quite strong indicating that this group of subjects had not used systematic perceptual factors shared by the timbres in making their dissimilarity ratings

**Table 3** Timbre coordinates along common dimensions and corresponding specificities (square root) for a 3-dimensional spatial solution with specificities and 5 latent classes of subjects derived from dissimilarity ratings by 88 subjects on 18 timbres

| Timbres | Dimension 1 | Dimension 2 | Dimension 3 | Specificities$^{1/2}$ |
|---|---|---|---|---|
| French horn | − 3.3 | 1.3 | − 1.5 | 1.4 |
| Trumpet | − 2.6 | − 1.9 | 0.4 | 1.6 |
| Trombone | − 2.4 | 1.7 | − 1.2 | 1.4 |
| Harp | 3.0 | 1.7 | − 0.4 | 0.8 |
| *Trumpar* | − 0.1 | − 2.7 | 0.1 | 1.9 |
| *Oboleste* | 3.0 | 1.7 | 0.7 | 1.4 |
| Vibraphone | 3.8 | 1.8 | 1.3 | 1.9 |
| *Striano* | − 1.4 | − 0.9 | 1.6 | 1.8 |
| Harpsichord | 3.6 | − 2.8 | 0.5 | 2.2 |
| English horn | − 1.9 | − 1.5 | − 1.9 | 1.9 |
| Bassoon | − 2.4 | − 1.8 | − 2.0 | 1.4 |
| Clarinet | − 2.4 | 1.9 | 0.5 | 2.5 |
| *Vibrone* | 0.7 | 2.3 | − 1.6 | 2.5 |
| *Obochord* | 2.5 | − 2.3 | − 2.7 | 0.0 |
| Guitar | 2.9 | 0.2 | 2.4 | 0.0 |
| String | − 2.4 | − 1.4 | 1.4 | 1.1 |
| Piano | 1.3 | 1.3 | 0.2 | 2.0 |
| *Guitarnet* | − 1.8 | 1.2 | 2.0 | 1.4 |
| Range | 7.1 | 5.0 | 5.1 | 2.5 |

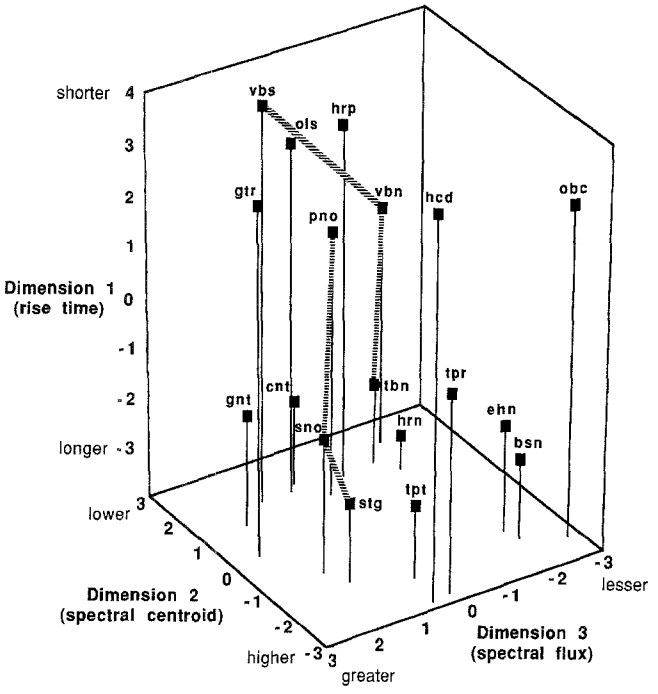


**Fig. 1** Timbre space in three dimensions: a spatial model with specificities and five latent classes derived from dissimilarity ratings on 18 timbres by 88 subjects. The acoustic correlates of the perceptual dimensions are indicated in parentheses. Hashed lines connect two of the hybrid timbres (vbn and sno) to their progenitors. Two others can be examined in the same way in this figure (tpr and gnt) (see Table 1 for timbre labels)

weighted more strongly dimension 1 and the specificities, which were weaker for Class 5, whereas dimensions 2 and 3 were stronger for Class 5 and weaker for Class 4.

*Estimation and analysis of class belongingness.* A-posteriori probabilities that each subject belonged to a given latent class were computed according to Bayes' theorem. Four subjects (three nonmusicians and one student musician) could not be assigned unequivocally to a given class as their posterior probabilities were distributed over all of the classes. So they did not fit into any one class and were removed from subsequent analyses of class structure. Four other subjects had ambiguous assignments to two classes with the preferred class having a probability of less than .65. The probability for the preferred class for 12 other subjects was between .65 and .95 and that for the remaining 68 subjects was greater than .95.

**Table 4** Estimated weights in the selected 3-dimensional model with specificities for 5 latent classes of subjects obtained from dissimilarity ratings by 88 subjects for 18 timbres

| Dim | 1 | 2 | 3 | Specif |
|---|---|---|---|---|
| Class | | | | |
| 1 | 1.14 | 0.94 | 1.18 | 1.72 |
| 2 | 0.81 | 0.69 | 0.73 | 0.74 |
| 3 | 1.05 | 1.77 | 1.22 | 0.58 |
| 4 | 1.24 | 0.44 | 0.51 | 1.09 |
| 5 | 0.76 | 1.15 | 1.36 | 0.88 |

The distribution across latent classes of the 84 subjects for whom a preferred class could be determined was analyzed according to our original grouping by degree of musical training as well as according to three items from the questionnaire that could be conceived as ordinal scales: years of music making (composition, conducting, performance), habitual amount of music playing, and habitual amount of music listening. These data are shown in Table 5. Two of the professional musicians (one each from Classes 1 and 4) did not fill out the questionnaire and so their data are absent from the last three factors in the table.

**Table 5** Distribution of subjects in each latent class according to degree of musical training, number of years of music making, amount of music playing and music listening (the bottom panel has data for only 82 subjects since two professional musician subjects did not fill out the questionnaire)

| Musical training | Class 1 | Class 2 | Class 3 | Class 4 | Class 5 |
|---|---|---|---|---|---|
| Professional | 8 | 5 | 0 | 7 | 4 |
| Amateur | 9 | 17 | 2 | 3 | 7 |
| Nonmusician | 7 | 4 | 1 | 7 | 3 |
| **Music making** | **Class 1** | **Class 2** | **Class 3** | **Class 4** | **Class 5** |
| > 10 years | 12 | 14 | 1 | 7 | 10 |
| 5–10 years | 3 | 6 | 0 | 0 | 1 |
| 0–4 years | 8 | 6 | 2 | 9 | 3 |
| **Music playing** | **Class 1** | **Class 2** | **Class 3** | **Class 4** | **Class 5** |
| Every day | 12 | 14 | 1 | 5 | 8 |
| Occasionally | 5 | 5 | 1 | 4 | 3 |
| None | 6 | 7 | 1 | 7 | 3 |
| **Music listening** | **Class 1** | **Class 2** | **Class 3** | **Class 4** | **Class 5** |
| > 3 times/day | 8 | 8 | 0 | 6 | 4 |
| 2–3 times/day | 11 | 11 | 0 | 6 | 6 |
| < 2 times/day | 4 | 7 | 3 | 4 | 4 |

The musical-training categories are defined above in the Methods section. The music-making categories were defined by the number of years of musical activity. While these two factors certainly covary, we felt that they might reveal different tendencies. The music-playing categories were defined by the amount of regular instrumental practice and the music-listening categories by the frequency of daily listening. In the analysis, Classes 1 and 2 were combined since their weight patterns were similar and Class 3 was removed since there were too few subjects for the analysis to be reliable. A table of counts for each analysis class and category was constructed from the data in Table 5. The null hypothesis was that the proportional distribution of categories for each analysis factor is constant across classes. Differences in distribution may indicate a relation between these biographical factors and class belongingness.

An exploratory data-analysis technique based on counted fractions (Tukey, 1977, chap. 15) was used to evaluate differences between classes for each factor. The folded log (or flog) represents the difference between the log of the proportion of data below a cutoff point between two categories on an ordinal scale and the log of the proportion above that point. This statistic is preferable for comparing classes to the raw proportion in each category as it is symmetric about a mid-point of equal proportion, because of the folding or differencing part, and increases the importance of smaller differences near the endpoints of the scale (0 and 1), due to the log transformation. This analysis shows that the distributions of categories of a given biographical factor are not the same for all classes for the factors musical training and music making, indicating differences across classes, while they are similar or parallel for the factors music playing and music listening, indicating a lack of difference. However, as can be seen in Table 5, it is generally the case that each class contains some people of each category of a given biographical factor, indicating that each category of a factor can give each of the weighting types revealed by the class structure.

## Discussion

The analyses presented in this section have several goals: (1) to evaluate the stability of the timbre spaces obtained on the same set of synthetic sounds in two different studies (the present one and that of Krumhansl, 1989) using different subject populations; (2) to determine the psychophysical nature of the common dimensions found in the present study on the basis of acoustic parameters derived by Krimphoff et al. (1994); (3) to report the results of informal listening to the distinctive aspects of sounds that are indicated by the specificities; and (4) to discuss the relation between class belongingness and musical training and activity.

### Comparison of the two spatial solutions obtained with Krumhansl's (1989) solution

Krumhansl (1989) used a set of 21 sounds that included the 18 employed in the present study. Her subjects were nine musicians on the staff at IRCAM. As is reported in that paper, her data were analyzed by Winsberg using the spatial models described by equations 3 and 4. A three-dimensional model with specificities was

selected. The weighted Euclidean model was rejected in favor of the unweighted model because adding weights only slightly improved the log likelihood. However, using a Procrustian rotation algorithm, Winsberg rotated the unweighted solution to the weighted model to yield meaningful dimensions. Correlations between dimensions in the resulting model and our model were computed on the coordinates of individual timbres as well as on the specificities (see Table 6).[4]

The first two dimensions of our timbre space were strongly correlated with the dimensions that Krumhansl (1989) labeled "temporal envelope" and "spectral envelope." However, Krumhansl's "spectral flux" dimension was not significantly correlated with any of the dimensions in our model, which suggests differences between the subject populations for this dimension. Although the specificities were significantly correlated between the two studies (see Table 6), there were important differences between them: the harp and *guitarnet* had much higher specificities in Krumhansl's model than in the present one, whereas the trombone, trumpet, and *trumpar* had moderate specificities in the present model and specificities of zero in Krumhansl's model.

For comparison, we computed the correlations ($df = 16$ in all cases) between the coordinates on the common dimensions of Krumhansl's (1989) model and each of the dimensions of the six-dimensional model selected by BIC. Krumhansl's "temporal envelope" and "spectral envelope" dimensions were well correlated with dimensions 1, $r = .98$, $p < .0001$,[5] and 3, $r = .77$, $p < .01$, of the six-dimensional space, respectively. Her "spectral flux" dimension, however, was significantly correlated with dimensions 3, $r = .60$, $p < .01$, 4, $r = -.53$, $p < .05$, and 6, $r = -.57$, $p < .05$. So our dimension 3 correlated with two of her three dimensions and her dimension 2 correlated with three of our six dimensions. The most coherent relation thus exists between the two models with three dimensions and specificities.

## Quantitative analysis of the psychophysical nature of the common dimensions

Only two of the previous studies that found two or three perceptual dimensions of timbre in MDS analyses attempted a quantitative description of their acoustic correlates. Grey and Gordon (1978) found that the spectral centroid of the tones correlated significantly and strongly with the coordinate along the first dimension of spatial models for Grey's (1977) original tones, for tones interpolated acoustically between these originals (Grey, 1975), and for spectral modifications of some of the original tones (Grey & Gordon, 1978). The correlation coefficients for comparisons between the dimension coordinates for these three spaces and a spectral centroid measure derived from a loudness function (Zwicker & Scharf, 1965) were .94, .92, and .92, respectively. Iverson and Krumhansl (1993) characterized the second dimension of their three spatial models (complete tones, attack portion only, attacks removed) in terms of spectral centroid, $rs = -.70, -.61, -.75$, respectively. The first dimension of the attack-only space was characterized in terms of the rise time from the start of the tone to maximum amplitude, $r = .79$.[6]

More recently, however, Krimphoff et al. (1994) have quantified satisfactorily all three common dimensions of Krumhansl's (1989) model. The first dimension correlated very strongly, $r = .94$, with the logarithm of the rise time (measured from the time the amplitude envelope reaches a threshold of 2% of the maximum amplitude to the time it attains maximum amplitude). The second dimension correlated very strongly, $r = .94$, with the spectral centroid (measured as the average over the duration of the tone of the instantaneous spectral centroid within a running time window of 12 ms). The third dimension correlated well, $r = .85$, with a measure of spectral irregularity (log of the *SD* of component amplitudes from a global spectral envelope derived from a running mean of the amplitudes of three adjacent harmonics) rather than with any of a number of measures of spectral variation over time as was presumed by Krumhansl (1989) in originally naming this dimension "spectral flux."

One of the aims of the current study was to validate the acoustic correlates described by Krimphoff (1993; Krimphoff et al., 1994) for a timbre space based on a large set of dissimilarity ratings given by subjects with

**Table 6** Correlations ($df = 16$) between coordinates in Krumhansl's (1989) and our three-dimensional models with specificities for 18 timbres

| | Our Model | | | |
|---|---|---|---|---|
| Krumhansl's Model | Dim 1 | Dim 2 | Dim 3 | Specif |
| Temporal Envelope | .98† | .09 | .27 | |
| Spectral Flux | −.33 | −.20 | .24 | |
| Spectral Envelope | −.01 | −.95† | −.07 | |
| Specificities | | | | .58* |

*$p = .01$, †$p < .0001$ (Fisher's $r$-to-$z$).

[4] Negative correlation coefficients indicate that the two axes being compared were inverted with respect to one another. Since we were interested in distances between objects, the sign of the coordinate system on each axis is of no particular importance

[5] All $p$ values shown for correlation coefficients are those for Fisher's $r$-to-$z$ transform

[6] Although acoustic measurements were compared directly to similarity ratings, equivalent characterizations of the coordinates of this dimension were not reported for their other two stimulus spaces (complete tones and attacks removed)

**Table 7** Correlations ($df = 16$) between acoustic parameters (Krimphoff, 1993; Krimphoff et al., 1994) and the coordinates of 18 timbres along the three common dimensions of our spatial model (5 latent classes and specificities derived from dissimilarity ratings by 88 subjects)

| Acoustic Correlate | Dim 1 | Dim 2 | Dim 3 |
|---|---|---|---|
| Log-Attack Time | −.94† | −.12 | −.16 |
| Spectral Centroid | −.04 | −.94† | −.21 |
| Spectral Irregularity | .41 | .31 | .13 |
| Spectral Flux | −.07 | .13 | .54* |

*$p < .05$, †$p < .0001$ (Fisher's $r$-to-$z$).

varying degrees of musical training. We therefore correlated these acoustic parameters with the coordinates of the 18 sounds ($df = 16$ in all cases) of the present study (see Table 7). Log-attack time accounted for 88% of the variance along Dimension 1 of the perceptual model, $r = -.94$, $p < .0001$. Spectral centroid accounted for 88% of the variance along Dimension 2, $r = -.94$, $p < .0001$. The third dimension (as in most previous studies) presented more of a difficulty in deriving its psychophysical interpretation. The spectral irregularity measure that accounted for 72% of the variance along Krumhansl's (1989) second dimension was not significantly correlated with the third dimension in the present spatial model. The label *spectral flux* given to her third dimension would suggest a parameter measuring the degree of variation of the spectral envelope over time. One such measure developed by Krimphoff (1993) described spectral flux as the average of the correlations between amplitude spectra in adjacent time windows: the smaller the degree of variation of the spectrum over time, the higher the correlation. This parameter correlated significantly with the third dimension of our spatial model, but only accounted for 29% of the variance along this dimension, $r = .54$, $p < .05$. This variance increased to 39% when four of the timbres (clarinet, trombone, *guitarnet*, and *vibrone*) were removed from the correlation, $r = .63$, $df = 12$, $p < .05$. Their removal did not affect the correlations of attack time and spectral centroid with dimensions 1 and 2.

Given the high degree of variation in duration and level among the stimuli (obtained by perceptually equalizing the sounds for loudness and subjective duration), we also correlated various measures of these parameters with the coordinates on the common dimensions. For duration, we computed the energy envelope of each sound (rms amplitude of the waveform over a 10 ms running window that advanced in 5-ms steps). The maximum point of this envelope was determined and the duration encompassing the part of the signal exceeding thresholds of 3, 10, and 20 dB below this maximum were computed. For level, we also determined the rms amplitude across the entire duration of each sound (expressed in dB). These values, as well as the total physical duration and maximum SPL recorded on a sound-level meter (see Methods section) were correlated with the coordinates of each timbre on the common dimensions. After using Bonferroni's correction for multiple tests, the only correlation that attained significance was between the −3 dB threshold duration and the coordinates of dimension 1, $r = -.82$, $df = 16$, $p < .0001$. For the set of synthesized instrument sounds used, the rise time was also strongly correlated with this duration measure, $r = .82$, $df = 16$, $p < .0001$. This correlation reflects the fact that, in general, impulsive sounds tend both to have sharp attacks and to begin decaying immediately, since there is no sustained excitation of the instrument. A similar interpretation was advanced by Iverson and Krumhansl (1993) for one of their dimensions.

For comparison, we also computed the correlation of Krimphoff et al's (1994) parameters with the coordinates on the dimensions of the six-dimensional solution selected by BIC ($df = 16$ in all cases). An equivocal result was found here, as with the correlation of this high-dimensional solution with Krumhansl's (1989) model. The log-rise time and spectral flux parameters correlated significantly only with dimensions 1, $r = -.94$, $p < .0001$, and 2, $r = .51$, $p < .05$, respectively. The spectral fine-structure parameter correlated significantly with dimensions 3, $r = -.55$, $p < .05$, 4, $r = .68$, $p < .01$, and 6, $r = .52$, $p < .05$; and the spectral centroid correlated significantly with dimensions 2, $r = -.74$, $p < .01$, and 3, $r = -.75$, $p < .01$. So two of our dimensions each correlated significantly with two acoustic parameters and two of the acoustic parameters correlated with several dimensions. We conclude that the psychophysical interpretation of this high-dimensional solution is rather ambiguous compared with the three-dimensional solution.

In contrast to the six-dimensional solution, Table 7 shows that each of the acoustic parameters that correlated significantly with a given dimension of the three-dimensional model with specificities was correlated with that dimension only. This orthogonality of the acoustic parameters associated with our perceptual dimensions is what makes a psychophysical interpretation possible. Further, an analysis for three dimensions without specificities was performed to evaluate the effect of removing specificities on the correlations of the acoustic parameters with the coordinates of the resulting solutions. If the specificities were removed, the correlation of spectral centroid with dimension 2 was reduced from .94 to .79, and that for spectral flux with dimension 3 was reduced from .55 to .27. The inclusion of specificities thus improved the psychophysical interpretation of the dimensions.

A similar additional analysis for three dimensions with specificities and only one latent class was performed to evaluate the effect of removing latent-class structure on the correlations of the acoustic parameters with the spatial configuration. If only one latent class

was used, the correlation of spectral flux with dimension 3 was slightly reduced from .55 to .49. These results indicate that the fit of the model to acoustic variables was slightly enhanced by including latent classes.

## Informal analysis of specificities

To make an informal attempt to identify the kinds of characteristics that are captured by the specificities in our model, we listened to the timbres individually and in comparison to all the others and noted verbally what seemed to be unique about each timbre. We noted no distinguishing features for timbres with specificities below 2.0 nor for the trumpet-like sound which had, nonetheless, a specificity of 2.7. For the rest of the timbres with specificities above 2.0, unique features were noted. These are summarized in Table 8. Our general impression is that the perceptual strength of these distinguishing features increases monotonically with the specificity value, but this correspondence needs to be verified under more controlled conditions in future work. One important fact to remark here, however, is that the features noted seem to be of two types: attributes that vary in degree (such as raspiness of attack, inharmonicity, graininess, deviation of pitch glide, hollowness of tone color) and attributes of a more discrete nature that vary in perceptual strength (such as a high frequency *zzzit*! on the offset, a suddenly damped or pinched offset, the presence of a clunk or thud). These reports suggest that what is captured by the specificities may include both additional continuous dimensions of variation, as well as discrete features of variable perceptual salience.

One might have imagined at the outset that hybrid instruments, being unfamiliar to listeners, would have a novelty that would distinguish them perceptually from the more traditional instruments. On average, however, the hybrid timbres do not have greater specific weights than the conventional instrument imitations, neither in Krumhansl's (1989) study nor in the present one. In fact, three of the six hybrids have lower than average specific weights. The highest specific weights found systematically in both studies were for the *vibrone*, the clarinet, the harpsichord, and the piano. The lowest weights in both studies were found for the *obochord*. The specificity of the piano-like sound argues strongly against the relation between specificity and familiarity since this instrument is probably one of the most familiar to the primarily European listeners who participated in this study. It is possible that in the case of certain instruments, such as the harpsichord, the properties suggested by the specificities are related to the simulation of specific mechanical properties of the object. In this case, the acoustic result of the return of the hopper in the harpsichord mechanism is perceptually important and should certainly play an important role in an identification task. Similarly, the timbre of the clarinet has a specific acoustic property, related to the predominance of odd harmonics in its spectrum, due to the conical geometry of the air column.

These results suggest that subjects did indeed make dissimilarity judgments on the basis of criteria related to structural characteristics of the stimuli. Certain of these criteria incited the subjects to analyze the relatively global and common degree of dissimilarity of all the stimuli based on continuous dimensions. The goal in this case was to determine the relations among stimuli along these common dimensions. Some stimuli, though, would seem to possess certain unique structural characteristics that cannot be accounted for by the Euclidean spatial model alone. These specific features or dimensions would be sufficiently salient perceptually to influence the dissimilarity of some timbres with respect to others. An indication of the presence of such features could lead to more systematic psychophysical analyses whose orientation would be quite different from an analysis based only on a Euclidean spatial model.

**Table 8** Eleven of the 18 timbres having specificities greater than 2.0 in the 3-dimensional spatial model with specificities derived from dissimilarity judgments of 88 subjects. Distinguishing features noted in informal comparisons among the sounds are described

| Instrument | Specificity | Description of distinguishing characteristics |
|---|---|---|
| Trombone | 2.1 | slightly raspy attack |
| *Guitarnet* | 2.1 | slight high frequency *zzzit*! on offset |
| Trumpet | 2.7 | nothing remarkable |
| *Striano* | 3.1 | downward pitch glide at end of tone |
| English horn | 3.5 | nasal formant, very sudden offset |
| *Trumpar* | 3.8 | noisy and/or rough attack, roughness in resonance of sound, low frequency thud on onset and offset |
| Vibraphone | 3.8 | metallic sound |
| Piano | 4.2 | slight inharmonicity and soft graininess |
| Harpsichord | 4.7 | very sharp, pinched offset with clunk |
| Clarinet | 6.4 | hollow timbre (very distinctive) |
| *Vibrone* | 6.4 | wobbly double attack |

## Class structure and musical activity

One final aim of this study was to examine the relation between class structure and the musical training of listeners. We hypothesized at the outset that there would be a richer dimensional structure for musicians and that the weights on the dimensions would be more evenly distributed, in line with results found for the multidimensional structure of musical pitch (Shepard, 1982). Although the folded-log analysis showed some differences in the proportional distribution of biographical factors among classes, there was no clear division of musicians, amateurs, and nonmusicians among the latent classes revealed by the CLASCAL analysis.

Recall that Classes 1 and 2 gave roughly equal weights across dimensions and specificities, while Classes 4 and 5 gave high weights on two dimensions, or on one dimension and the specificities, and low weights on the others, respectively. It is these patterns that our analysis sought to explain by the biographical factors. One interpretation of the patterns is that the equal weights for Classes 1 and 2 reflect a shifting of attention among dimensions and specificities over the course of an experimental session, which averages out over trials. The subjects of Classes 4 and 5 may have adopted more consistent strategies of judgment that focussed on a smaller number of dimensions and stuck to them throughout the experimental session. Another interpretation is that members of Classes 1 and 2 were able to focus on more dimensions at a time than could the members of the other classes, and one might predict a priori that these would be principally musicians. At any rate, the factor responsible for making Class 4 focus on the attack time dimension and the specificities, while Class 5 focussed on the spectral centroid and spectral flux dimensions is difficult to tease out from the analysis of the biographical factors presented in the Results section. Overall, both musicians and nonmusicians were able either to weight all dimensions equally (Classes 1 and 2) or to give special attention to some dimensions (Classes 4 and 5). Nor does the degree of musicianship or the amount of training, playing or listening mean that one factor or another will be given preferential weight. The pattern of weighting of a given subject cannot be predicted simply from the biographical data related to that subject. It would thus seem difficult to extract any clear picture of the factors that influence the weight patterns from biographical factors related to musical training and activity.

Separate CLASCAL analyses (three dimensions with specificities and one latent class) were performed for the professional and nonmusician groups as well as for each individual latent class. The variance about the model distances was much greater for the nonmusicians (3.53) and amateurs (3.64) than for the professionals (2.75). The variances for the individual latent classes (containing mixtures of professionals, amateurs, and nonmusicians) were less than the variance for the professional group (range 2.41–2.72). The inclusion of class weights in the dimensional model is thus justified in terms of model fit since it reduces the overall variance. This pattern of results suggests that the effect of musicianship is, among other things, one of variance. Latent classes do not differ with respect to variance, but musicians and nonmusicians do. So musicianship affects judgment precision and coherence.

## Conclusion

A group of 98 listeners of varying degrees of musical training rated the dissimilarities among a set of 18 synthesized instrument sounds. Of these, 88 gave dissimilarity matrices that were sufficiently coherent to be analyzed with an extended version of the CLASCAL MDS algorithm (Winsberg & De Soete, 1993). Monte Carlo simulations indicated that five latent classes of subjects were sufficient to represent the data. About half of the subjects were in the first two classes which had very similar weight patterns, being distinguished only by a scale factor. These subjects gave approximately equal weight to all dimensions and the specificities. The other classes' weight patterns suggested that certain dimensions, or the set of specificities, had greater perceptual salience for each group of listeners. The class structure, however, had an ambiguous relation to the degree of musical training and activity. Timbre, being composed of many of the sensory qualities that specify the identity of a sound source, may likely be used as an important auditory cue for monitoring the environment on a continual basis by listeners in their everyday lives (McAdams, 1993). It would not therefore be surprising that musical training as such did not play an important role in defining the class structure revealed in this study. What is suggested by our study, however, is that musicians make more coherent and precise judgments of timbral dissimilarity and may, by virtue of their training, have adopted a more consistent judgment strategy.

The CLASCAL analysis suggested a six-dimensional model without specificities for individual timbres with a three-dimensional model with specificities being a close contender. Psychophysical quantification of the three-dimensional model was achieved, whereas only one dimension of the six-dimensional solution was unequivocally correlated with one of the acoustic parameters derived by Krimphoff (1993; Krimphoff et al., 1994). Further, the first two dimensions and the specificities of the three-dimensional model correlated significantly with a similar spatial solution found by Krumhansl (1989), who employed a group of professional musician subjects and a set of stimuli including all of ours.

The acoustic correlates of the three common dimensions in our spatial model were log-rise time, spectral centroid, and spectral flux. The first dimension was also well correlated with the duration during which the sound's amplitude envelope remained within 3 dB of the maximum, suggesting that this dimension distinguishes impulsively from continuously excited sound sources. In most multidimensional-scaling studies of musical timbre, dimensions qualitatively related to the first two parameters have been found. The third dimension seems to be less stable across subject populations (in a comparison of this study with that of Krumhansl, 1989) and/or stimulus sets (Grey, 1977; Grey & Gordon, 1978; Krimphoff et al., 1994).

That abstract parameters, such as spectral centroid, spectral irregularity, and spectral flux, seem to explain some of the dimensions used to compare timbres in a dissimilarity-rating task, may suggest that such

judgments are based in part on raw sensory qualities. A dimension related to the manner of excitation of the instrument would suggest that the judgments also include inferences about the nature of the sound sources involved. According to this view, differences between latent classes of subjects would reflect differences either in sensitivity to these qualities or in the importance accorded to them in the comparisons made by the subjects. This notion is further supported by the fact that similar predictive variables are found for synthetic sounds of varying degrees of resemblance to acoustic sources (Miller & Carterette, 1975; and the present study), for recorded instrument tones (Iverson & Krumhansl, 1993; Serafini, 1993; Wedin & Goude, 1972), or for analyzed, modified, and resynthesized instrument tones (Grey, 1977; Grey & Gordon, 1978; Iverson & Krumhansl, 1993). Nonetheless, none of these studies really presented a broad and balanced set of instrument sounds deriving both from different types of resonating structures (strings, bars, plates, air columns) and means of excitation (blowing, bowing, striking, plucking). Such a set would allow systematic variation of the many types of physical properties that instruments possess, perhaps giving rise to judgments more classificatory than continuous. Work in progress in our laboratory intends to clarify this issue.

The specificities that were suggested by the model were explored informally in the present study. This exploration suggested that distinguishing features of the timbres indicated by the specificities in the CLASCAL analysis are of two types: additional perceptual dimensions on which only certain sounds vary and discrete features of varying degrees of perceptual salience. Further work in both acoustic analysis and psychophysical experimentation is needed to verify and develop this notion.

The CLASCAL algorithm (Winsberg & De Soete, 1993), and in particular the extended version employed here, promises to be a powerful tool for the analysis of timbre perception. Specificities provide a way to representing systematic variation in dissimilarities that can't be accounted for by shared dimensions, and may indicate additional dimensions along which only a single or a small number of timbres vary or unique attributes with varying degrees of perceptual salience. Further, the model captures certain systematic variations in judgments that are accounted for by different weighting of the common dimensions and the specificities by latent classes of subjects. Taken together, these added modeling features give a better fit to the data and render the resulting model more interpretable in terms of its acoustic correlates. This approach provides a much needed tool for the analysis of complex perceptual representations and for suggesting orientations for their psychophysical quantification.

**Acknowledgments** The authors would like to thank Eric F. Clarke for help in recruiting subjects at the Music Department, City University, London, UK, and two anonymous reviewers for helpful comments.

## References

American Standards Association (1960). Acoustical Terminology, S1.1–1960. New York: American Standards Association.

Aitken, M., Andersen, D., & Hinde, J. (1981). Statistical model of data on teaching styles. *Journal of the Royal Statistical Society, Series A, 144*, 419–461.

Aikake, H. (1977). On entropy maximization. In P. R. Krishniah (Ed.), *Applications of statistics* (pp. 27–41). Amsterdam: North-Holland.

Bentler, P. M., & Weeks, D. G. (1978). Restricted multidimensional scaling methods. *Journal of Mathematical Psychology, 17*, 138–151.

Bockenholt, U., & Bockenholt, I. (1990). Modeling individual differences in unfolding preference data: A restricted latent class approach. *Applied Psychological Measurement, 14*, 257–269.

Bogdozan, H. (1987). Model selection and Aikake's information criterion (AIC) : The general theory and its analytic extensions. *Psychometrika, 52*, 345–370.

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.

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

Dempster, A. P., Laird, N. M., & Rubin, D. R. (1977). Maximum likelihood estimation from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B, 39*, 1–38.

De Leeuw, J., & Heiser, W. (1980). Multidimensional scaling with restrictions on the configuration. In P. R. Krishniah (Ed.), *Multivariate analysis, vol. 5* (pp. 501–522). Amsterdam: North Holland.

De Sarbo, W. J., Howard, D. J., & Jededi, K. (1991). MULTICLUS: A new method for simultaneously performing multidimensional scaling and cluster analysis. *Psychometrika, 56*, 121–136.

De Soete, G. (1990). A latent class approach to modeling pairwise preferential choice data. In M. Schader & W. Gaul (Eds.), *Knowledge, data, and computer-assisted decisions* (pp. 103–113). Berlin: Springer-Verlag.

De Soete, G., Carroll, J. D., & Chaturvedi, A. D. (1993). A modified CANDECOMP method for fitting the extended INDSCAL model. *Journal of Classification, 10*, 75–90.

De Soete, G., & De Sarbo, W. (1991). A latent class probit model for analyzing pick any n data. *Journal of Classification, 8*, 45–63.

De Soete, G., & Heiser, W. J. (1993). A latent class unfolding model for analyzing single stimulus preference ratings. *Psychometrika, 58*, 545–565.

De Soete, G., & Winsberg, S. (1993). A Thurstonian pairwise choice model with univariate and multivariate spline transformations. *Psychometrika, 58*, 233–256.

Donnadieu, S., McAdams, S., Winsberg, S. (1994). Caractérisation du timbre des sons complexes. I: Analyse multidimensionnalle. *Journal de Physique 4*(C5), 593–596.

Ehresman, D., & Wessel, D. L. (1978). Perception of timbral analogies. *Rapports IRCAM, 13*, Paris: IRCAM.

Formann, A. K. (1989). Constrained latent class models: Some further applications. *British Journal of Mathematical Psychology, 42*, 37–54.

Gower, J. C. (1966). Some distance properties of latent root and vector methods using multivariate analysis. *Biometrika, 53*, 325–338.

Grey, J. M. (1975). An exploration of musical timbre. Unpublished Ph.D. dissertation, Stanford University, Stanford, CA. Stanford University, Dept. of Music Report STAN-M-2.

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

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

Krimphoff, J. (1993). Analyse acoustique et perception du timbre. Unpublished DEA thesis. Université du Maine, Le Mans, France.

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

Krumhansl, C. L. (1989). Why is musical timbre so hard to understand? In S. Nielzén & O. Olsson (Eds.), *Structure and perception of electroacoustic sound and music* (pp. 43–53). Amsterdam: Elsevier (Excerpta Medica 846).

Kruskal, J. B. (1964a). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika, 29*, 1–27.

Kruskal, J. B. (1964b). Non-metric multidimensional scaling: a numerical method. *Psychometrika, 29*, 115–129.

McAdams, S. (1993). Recognition of auditory sources and events. In S. McAdams & E. Bigand (Eds.), *Thinking in sound: The cognitive psychology of human audition* (pp. 146–198). Oxford: Oxford University Press.

McAdams, S., & Cunibile, J-C. (1992). Perception of timbral analogies. *Philosophical Transactions of the Royal Society, London, Series B, 336*, 383–389.

McLaughlin, G. J., & Basford, K. E. (1988). *Mixture models*. New York: Marcel Dekker.

Miller, J. R., & Carterette, E. C. (1975). Perceptual space for musical structures. *Journal of the Acoustical Society of America, 58*, 711–720.

Plomp, R. (1970). Timbre as a multidimensional attribute of complex tones. In R. Plomp & G. F. Smoorenburg (Eds.), *Frequency analysis and periodicity detection in hearing* (pp. 397–414). Leiden: Sijthoff.

Plomp, R. (1976). *Aspects of tone sensation. A psychophysical study*. London: Academic Press.

Plomp, R., Pols, L. C. W., & van de Geer, J. P. (1967). Dimensional analysis of vowel spectra. *Journal of the Acoustical Society of America, 41*, 707–712.

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

Pols, L. C. W., van der Kamp, L. J. T., & Plomp, R. (1969). Perceptual and physical space of vowel sounds. *Journal of the Acoustical Society of America, 46*, 458–467.

Ramsay, J. O. (1977). Maximum likelihood estimation in multidimensional scaling. *Psychometrika, 42*, 241–266.

Sattath, S., & Tversky, A. (1977). Additive similarity trees. *Psychometrika, 42*, 319–345.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics, 6*, 461–464.

Serafini, S. (1993). Timbre perception of cultural insiders: A case study with Javanese gamelan instruments. Unpublished Master's thesis, University of British Columbia, Vancouver, BC.

Shepard, R. N. (1962a). The analysis of proximities: Multidimensional scaling with an unknown distance function. Part I. *Psychometrika, 27*, 125–140.

Shepard, R. N. (1962b). The analysis of proximities: Multidimensional scaling with an unknown distance function. Part II. *Psychometrika, 27*, 219–246.

Shepard, R. N. (1982). Structural representations of musical pitch. In D. Deutsch (Ed.), *The psychology of music* (pp. 344–390). New York: Academic Press.

Takane, Y., & Sergent, J. (1983). Multidimensional scaling models for reaction times and some different judgments. *Psychometrika, 48*, 329–424.

Torgerson, W. S. (1958). *Theory and methods of scaling*. New York: Wiley.

Tukey, J. W. (1977). *Exploratory data analysis*. Reading, MA: Addison–Wesley.

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

Wessel, D. L. (1979). Timbre space as a musical control structure. *Computer Music Journal, 3(2)*, 45–52.

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). Computer Music Association, San Francisco.

Winsberg, S., & Carroll, J. D. (1989a). A quasi-nonmetric method for multidimensional scaling via an extended Euclidean model. *Psychometrika, 54*, 217–229.

Winsberg, S., & Carroll, J. D. (1989b). A quasi-nonmetric method for multidimensional scaling of multiway data via a restricted case of an extended INDSCAL model. In R. Coppi & S. Bolasco (Eds.), *Multiway data analysis* (pp. 405–414). Amsterdam: North-Holland.

Winsberg, S., & De Soete, G. (1993). A latent class approach to fitting the weighted Euclidean model, CLASCAL. *Psychometrika, 58*, 315–330.

Zwicker, E., & Scharf, B. (1965). A model of loudness summation. *Psychological Review, 72*, 3–26.