Perspectives on the Contribution of Timbre to Musical Structure

Author(s): Stephen McAdams

Source: *Computer Music Journal*, Vol. 23, No. 3, Recent Research at IRCAM (Autumn, 1999), pp. 85-102
Published by: The MIT Press
Stable URL: http://www.jstor.org/stable/3681242

Stephen McAdams
Institut de Recherche et Coordination
Acoustique/Musique (IRCAM)
1 Place Igor-Stravinsky
Paris F-75004, France
http://www.ircam.fr
and
Laboratoire de Psychologie Expérimentale, Centre
Nationale de Recherche Scientifique (CNRS)
Université René Descartes
EPHE, 28 rue Serpente
Paris F-75006, France
smc@ircam.fr

Perspectives on the Contribution of Timbre to Musical Structure

Timbre is a misleadingly simple word that encompasses not only a very complex set of auditory attributes, but also a plethora of important psychological and musical issues (Risset and Wessel 1999; McAdams 1993; Hajda, Kendall, Carterette, and Harshberger 1997). It covers many parameters of perception that are not accounted for by pitch, loudness, spatial position, and duration. It is thus, by definition, multidimensional. Musical timbre has been a major component of research carried out by the Music Perception and Cognition team at IRCAM since the team's inception in 1984, and even before in previous work by David Wessel. Our approach has been multifarious, including work on perception and recognition of musical sound sources and sequences, as well as a consideration of the psychological and musical implications of timbre as a set of form-bearing dimensions in music (McAdams 1989). This article presents an overview of some (but not all) of the work on timbre performed at IRCAM, and is not meant to provide an exhaustive review of the literature on timbre perception and modeling.

A goal of the research program has been to determine the structure of the multidimensional perceptual representation of timbre (the so-called timbre space) for individual notes played by musical instruments, and then to attempt to define the acoustic and psychoacoustic factors that underlie this representation. The combination of a quantitative model of perceptual relations among timbres and the psychophysical explanation of the parameters of the model is an important step in gaining predictive control of timbre in several domains such as sound analysis and synthesis and intelligent search in sound databases. Of course, such representations are only useful to the extent that they are: (1) generalizable beyond the set of sounds actually studied, (2) robust with respect to changes in musical context, and (3) generalizable to kinds of listening tasks other than those used to construct the model. To the degree that a representation has these properties, it may be considered a genuine model of musical timbre, the main feature of a good model being predictive power.

The main thrust of the early work on timbre, consisting of determining this multidimensional representation for isolated sound and explaining it psychophysically, is addressed in the first part of this article. Some issues concerning the robustness of the representation across various experimental (and by extrapolation, musical) contexts are also considered. The last three parts of the article address the role of timbre in the creation and perceptual organization of musical materials. The primary questions to be addressed are: Can a timbre-space model predict anything about the perception of abstract relations between pairs of timbres (timbre intervals, by analogy with pitch intervals)? And is the perception of such relations invariant across changes in timbre that preserve the relative positions of the timbres with respect to one another in the space? Can a timbre-space representation be used to organize sequences of timbres into coherent auditory streams, or inversely, to determine how a sequence of timbres

Computer Music Journal, 23:3, pp. 85–102, Fall 1999

will be organized perceptually into streams? And finally, can timbre play a role in larger-scale aspects of musical form, such as the movement between tension and relaxation that has been discussed a great deal in the realm of melodic/harmonic and rhythmic/metric structure?

Multidimensional Timbre Spaces and Their Acoustic Correlates

There are several aims in performing multidimensional scaling (MDS) to derive a timbre space. The main aim is to determine the components of the mental representation of musical timbre. A second one is to quantify the acoustical correlates of the dimensions and features of this representation. And a third one is to develop predictive models on the basis of the first two results for sound-synthesis control and intelligent search in sound databases.

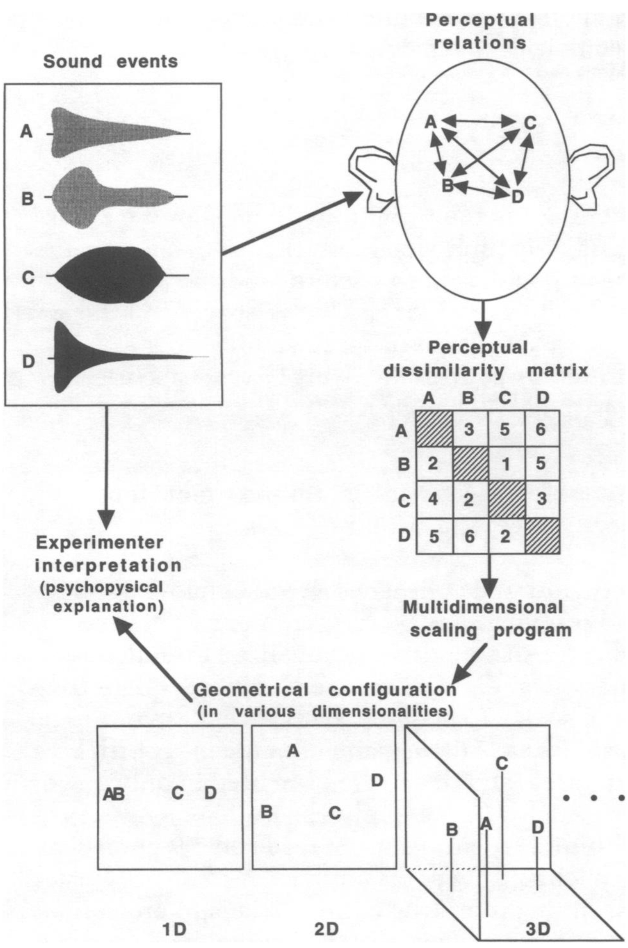
Multidimensional Scaling Techniques

The development of techniques for multidimensional scaling of proximity data in the 1950s and 1960s have provided a tool for exploring complex sensory representations (for a review, see the work by McAdams, Winsberg, Donnadieu, De Soete, and Krimphoff [1995]). These techniques have several advantages, as well as a few limitations. The primary advantage is that from a relatively simple task—judging the degree of similarity or dissimilarity between all pairs of items from a fixed set—an ordered structure is obtained that can often lend itself to psychophysical quantification. Applied for the first time to musical timbre by Plomp (1970) and subsequently by Wessel (1973) and Miller and Carterette (1975), this kind of analysis searches for structure in the perceptual data without obliging the experimenter to make any a priori assumptions about the nature of that structure. To critics of this technique that complain, "You only get out what you put in," the only possible response is, "Well thank goodness, since I wasn't sure what I put in in the first place!" Often, we are interested in discovering the perceptual structure of a set of complex sound events, the nature of which we do not know in advance. These techniques are quite useful for this kind of exploratory data analysis, although they can also be used for more confirmatory analyses, once one has a more clear idea of the relations among acoustic and perceptual parameters.

The basic principle of MDS is illustrated in Figure 1. A set of sounds (equalized in pitch, loudness, duration, and spatial position) is presented to a group of listeners in all possible pairs. The listeners are asked to rate the degree of dissimilarity between each pair of timbres on a numerical scale or with a continuous slider. This scale gives high similarity at one end and high dissimilarity at the other. The basic assumption is that there exists a mental representation of each timbre that has certain prominent components, and the number or slider position reflects a comparison based on these components. Furthermore, this representation is assumed to be relatively similar across listeners (perhaps with some variations that are discussed below), so the structure in the data should somehow reflect the perceptual structure. The data set for each listener has the form of a matrix, each cell corresponding to a pair of timbres. In Figure 1, pairs of identical timbres were not presented, so the diagonal of the matrix is missing. The upper and lower triangular matrices are different only in the order of presentation of the timbres (A, B versus B, A). Any systematic differences between the two half-matrices across listeners may indicate important asymmetries. In studies where we have presented both orders, systematic asymmetries have never been found. The set of matrices are submitted to an MDS program, the main task of which is to fit a distance model to the dissimilarity data so that a monotonic relation exists between the two (i.e., the greater the dissimilarity, the greater the distance); so the distance models the dissimilarity data.

Distance models come in many flavors. The earliest models were Euclidean models or Minkowski generalizations of Euclidean distance. The distance d_{ij} between any two timbres i and j thus

Figure 1. Schema illustrating the derivation and psychophysical quantification of a timbre space.



takes the following form:

$$d_{ij} = \left[\sum_{k=1}^{K} \left(x_{ik} - x_{jk} \right)^r \right]^{\frac{1}{r}}, \qquad (1)$$

where x_{ik} is the coordinate of timbre i on dimension k, K is the total number of dimensions in the model, and r determines the Minkowski metric.

If $r = 2$, a simple Euclidean distance is obtained, and this is what is usually used in timbre studies. In this case, the model incorporates the hypothesis that the distance between timbres is in a Euclidean space with a number of dimensions that is much smaller than the number of stimuli. Each timbre thus has a certain value along each perceptual dimension and could be considered to be specified by this set of values (at least in terms of the small number of salient perceptual dimensions recovered by the analysis technique). Some studies (Wessel 1973, 1979) have used this model as embodied in the MDSCAL program (Kruskal 1964a, 1964b) or other similar programs.

The model also presumes that this set of dimensions and their relative salience are the same for each listener, which may seem a bit constraining given the amount of variation in sensitivity to different auditory parameters that we know exists among listeners. So a possible extension of this spatial model is to presume that each listener may accord more or less perceptual "weight" to each dimension, which gives the following form for the Euclidean distance model:

$$d_{ij} = \left[\sum_{1}^{K} w_{nk} \left(x_{ik} - x_{jk} \right)^2 \right]^{\frac{1}{2}}, \qquad (2)$$

where w_{nk} is the weight [0–1] given to dimension k by listener n.

The INDSCAL program (Carroll and Chang 1970) employs this model, and has also been used in the realm of timbre research (Miller and Carterette 1975; Grey 1977; Grey and Gordon 1978). This distance model has a number of inconveniences from a statistical standpoint, since with each new listener added to the data set the number of parameters in the model increases, which poses problems for validating model selection. One way to get around this problem is to consider that the listeners actually form a small number of "latent classes" that can be determined on the basis of their data. The classes are "latent" in the sense that they are not predetermined, but are derived from the structure of the data. This latent-class approach was implemented in the CLASCAL program by Winsberg and De Soete (1993). In this model, the individual subject weights are replaced by weights for each class of subjects. The number of latent classes is determined at the outset on the basis of the raw data, and statistical tests are performed at the end to estimate the probability that each subject belongs to each class. In general, subjects are assigned to a single class, although class "belongingness" can be ambiguous for some subjects.

Both of these models presume that the timbres share all the perceptual dimensions. However, intuitively it seems likely that some sounds may have characteristics that no other sounds in the set have (like weak even-numbered harmonics in a clarinet sound, or the rapid damping of a harpsichord sound). One might say that these sounds have "specificities" that make them dissimilar to all the other timbres, but such features cannot be accounted for by the shared dimensions along which vary all the timbres of the tested set. This extended Euclidean model has been implemented in the EXSCAL program (Winsberg and Carroll 1988), which uses the following distance model:

$$d_{ij} = \left[\sum_{1}^{K} \left(x_{ik} - x_{jk} \right)^2 + s_i + s_j \right]^{\frac{1}{2}} \qquad (3)$$

where s_i is the specificity for timbre i. Note that $\sqrt{s_i}$ is comparable in magnitude to the coordinates x_k. There are two possible sources for such specificities. Either a given specificity represents an additional dimension along which only timbre i varies, and $\sqrt{s_i}$ is the coordinate along this dimension, or it represents a discrete feature present in timbre i and absent in all the others, and $\sqrt{s_i}$ is the perceptual salience of that feature. The first timbre study to employ this model was conducted by Krumhansl, Wessel, and Winsberg on a set of 21 FM instruments developed by Wessel, Bristow, and Settel (1987). These instruments were intended either to imitate conventional orchestral instruments or to constitute chimeric hybrids between them (e.g., the *trumpar* has the head of a trumpet and the tail of a guitar). We will refer to this space below as the KWW space. In addition to three common dimensions, the EXSCAL analysis also revealed prominent specificities on many of the instrument sounds (Krumhansl 1989; Winsberg and Carroll 1989).

Finally, a combination of the extended Euclidean model and the latent-class approach has resulted in an extension of the CLASCAL model developed by Winsberg and De Soete. This distance model has both specificities and class weights; the weights are applied to each dimension and to the set of specificities taken collectively:

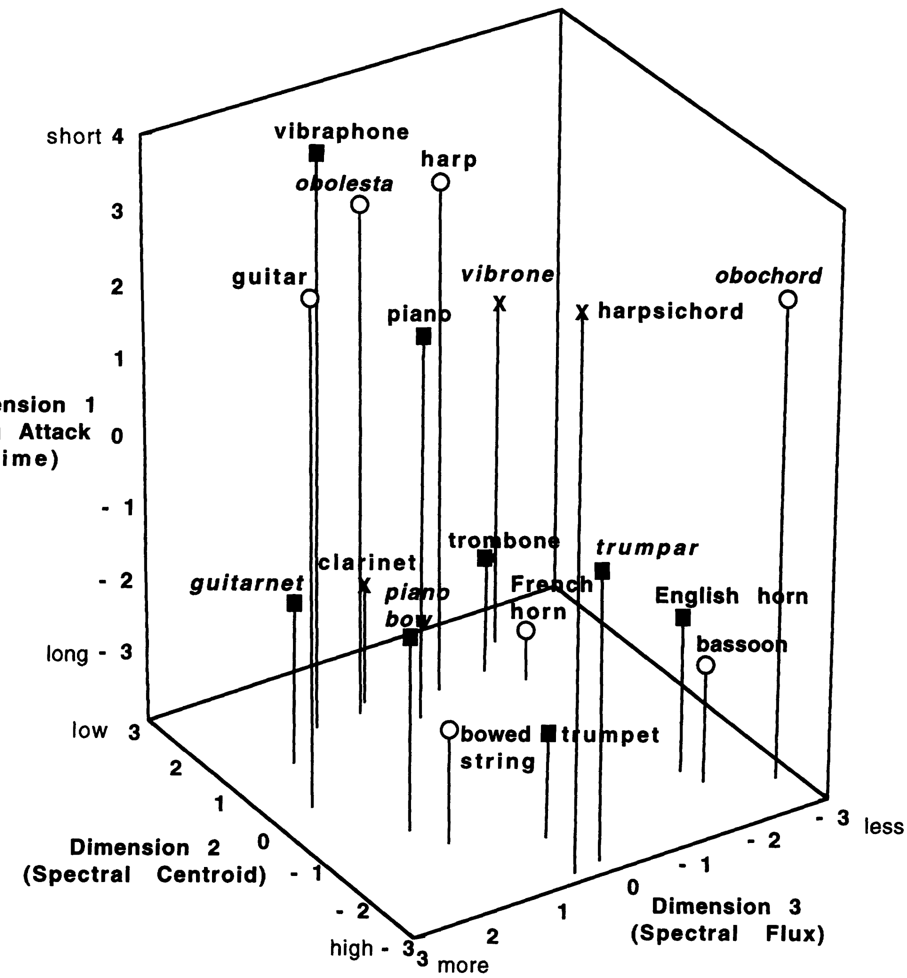
$$d_{ij} = \left[\sum_{1}^{K} w_{kc} \left(x_{ik} - x_{jk} \right)^2 + v_c \left(s_i + s_j \right) \right]^{\frac{1}{2}}, \qquad (4)$$

where w_{kc} is the weight on dimension k for class c, and v_c is their weight on the set of specificities. This model has been used to study a subset of the FM timbres used for the KWW space, and those results are discussed in some detail below (McAdams et al. 1995). This latter space is referred to as the MWDDK space.

A Timbre Space with Common Dimensions, Specificities, and Latent Classes

McAdams and colleagues (1995) employed 18 FM timbres, including both instrument imitations and hybrids. All sounds were equalized for loudness, duration, and pitch, and were played straight from the Yamaha TX802 that synthesized them through headphones. All 153 pairs of nonidentical timbres were used, and they were presented in only one order, which was randomly chosen for each listener. Dissimilarity ratings were made on a 9-point scale, with 9 being "very dissimilar." A total of 84 listeners varying from true nonmusicians to professional musicians participated. They completed a questionnaire concerning musical training and music listening habits. They were classed on the basis of the information in the questionnaire as nonmusician, amateur, or professional. The CLASCAL analysis revealed five latent classes with most of the listeners falling into the first two classes. In no case could class belongingness be related to musical training, musical expertise, or music-listening habits. What suggests that these classes are real is that the error variance of the raw dissimilarity data compared to the model distances between timbres was smaller within a given class than it was even for the group of professional musicians. This variance was the largest for the nonmusician group. This result suggests that while listeners' judgments are more coherent among professional musicians than among nonmusicians, they are even more co-

Figure 2. A three-dimensional timbre space obtained from 18 FM timbres with data from 84 listeners. The dimension labels correspond to the best acoustic correlates. The points indicating the position of each timbre correspond to the relative weight of the specificity: X = high, filled square = moderate, and open circle = low or null.



herent within a given latent class of listeners. What remains a mystery is the set of factors that determine to which class a person belongs.

The analysis also selected a three-dimensional model with specificities as shown in Figure 2. The acoustic correlates of these dimensions are discussed below, but note that while the timbres are distributed in a relatively homogeneous manner along dimensions 2 and 3, they form two large clusters along dimension 1. Also shown in the figure by the form of the point representing the position of each timbre in the space is a rough representation of the strength of the specificity. It is interesting to note that specificity has nothing to do with familiarity, since some hybrids have low specificities while some instrument imitations have high specificities (particularly the harpsichord, which is one of the better imitations in the set). Informal listening and verbal descriptions tend to support the notion that this component of the analysis captures some distinguishing feature of certain of the timbres, such as the pinched offset with the sound of the return of the hopper in the harpsichord, the hollow timbre of the clarinet, the metallic sound of the vibraphone, and so on. Another aspect of the MDS analysis suggests that both specificities and latent classes are real features of this data set: when they are not included in the distance model, the correlation of the common dimensions with various acoustic parameters is much lower, suggesting that their addition cleans up the space and leads to a more reliable interpretation of the common dimensions. Let us now turn to the acoustic correlates.

Acoustic Correlates for Common Dimensions

Our approach to determining the acoustic correlates of timbre space focused initially on the KWW and MWDDK spaces using the FM timbres, and has expanded more recently to include several other spaces, using analyzed/resynthesized or recorded sounds, that have been published in the literature or are currently submitted for publication (McAdams and Winsberg in preparation; McAdams et al. in preparation). The FM spaces will suffice to demonstrate the approach's advantages and limitations. We tend to use an empirical loop consisting of listening to the sounds in front of a visual representation of the timbre space, and trying to get an auditory sense of what changes systematically as one plays a timbre trajectory across a given dimension. The initial impression then leads to the development of signal-processing algorithms, usually based on a time-frequency representation derived from a short-term Fourier analysis (phase vocoder and the like). We have used both the additive environment developed at IRCAM (Depalle, Garcia, and Rodet 1993) and Beauchamp's (1993) Sndan environment. The goal is to find a parameter that varies in a linear relation with the coordinates of the timbres along a given dimension in the timbre space. So we try various algorithms that provide a single parameter per sound, and then either reject them or progressively refine them until the correlations are as high as possible. This approach was first applied by Krimphoff, McAdams, and Winsberg (1994) to the KWW space. The main four correlates are specified in Equations 5–8 (LAT = log attack time, SC = spectral centroid, SS = spectral smoothness, and SF = spectral flux). Attack time is the time it takes to progress from a threshold energy level to the maximum in the rms amplitude envelope. Spectral centroid is the center of gravity of the long-term amplitude spectrum. Spectral smoothness is related to the degree of amplitude difference between adjacent partials in the spectrum computed over the duration of the tone. A trumpet often has a smooth spectrum and a clarinet a jagged one, so the former would have a low value of SS and the latter a higher one. Spectral flux is a measure of the degree of variation of the spectrum over time.

$$\mathrm{LAT} = \log_{10}\left(t_{\max} - t_{\mathrm{threshold}}\right) \tag{5}$$

$$\mathrm{SC} = \frac{1}{T}\int_0^T \mathrm{B}(t)dt \text{ with } \mathrm{B}(t) = \left[\frac{\sum_{k=1}^{N} k\mathrm{A}_k(t)}{\sum_{k=1}^{N} \mathrm{A}_k(t)}\right] \tag{6}$$

for a given analysis window

$$\mathrm{SS} = \sum_{k=1}^{N}\left|20\log(\mathrm{A}_k) - \frac{20\log(\mathrm{A}_{k-1}) + 20\log(\mathrm{A}_k) + 20\log(\mathrm{A}_{k+1})}{3}\right| \tag{7}$$

$$\mathrm{SF} = \frac{1}{\mathrm{M}}\sum_{p=1}^{M}\left|r_{p,p-1}\right| \text{ with } \mathrm{M} = \frac{T}{\Delta t} \text{ and } \Delta t = 16 \text{ msec} \tag{8}$$

where

t_{max} = the instant in time at which the rms amplitude envelope attains its maximum,

$t_{threshold}$ = the time at which the envelope exceeds a threshold value ($0.02 * t_{max}$ in our case),

T = the total duration of the sound,

t = the begin time of the sliding short-term Fourier analysis window,

A_k = the amplitude of partial k,

N = the total number of partials, and

$r_{p,p-1}$ = the Pearson product-moment correlation coefficient between the amplitude spectra at times t_p and t_{p-1}.

We found very high correlations with log attack time (LAT) and spectral centroid (SC) for two dimensions, and a relatively high one with spectral smoothness (SS) for the third dimension. Note that one dimension was temporal and two were spectral in nature in the KWW space. Nearly identical correlations were found for LAT with dimension 1 and SC with dimension 2 in the MWDDK space. However, the third dimension in this latter space was spectrotemporal in nature, and was correlated (somewhat more weakly) with SF. This last set of acoustic correlates is used for the dimension labels in Figure 2. This approach has been used with considerable success (1) on several other musical timbre spaces (McAdams et al. in preparation), (2) on a space of bar sounds generated by physical-model synthesis (Roussarie, McAdams, and Chaigne 1998), as well as (3) on several spaces of interior car sounds (Susini, McAdams, and Winsberg in press).

A Distance Model in Physical Parameter Space

Recently, in collaboration with the Studio OnLine project at IRCAM, one of the research objectives of the Music Perception and Cognition team was to develop a multidimensional distance model from the relations among the perceptual dimensions obtained by MDS analyses and the calculated acoustic parameters (Misdariis et al. 1998). The main aim was to have a quantifiable measure of similarity between sounds that was founded on perceptual principles. First, the subjective data, i.e., the coordinates of the sounds along each perceptual axis, were normalized to make the results of the KWW and MWDDK studies comparable by applying a multiplicative factor to the KWW space. Afterward, a mathematical relation between these subjective data and the values of the corresponding parameter was obtained by a linear regression method: the slopes of the resulting six straight lines (three dimensions for each space) could then be considered as proportionality coefficients between the perceptual and the physical dimensions. Finally, these coefficients were used as weighting factors for the four acoustic correlates in a formula representing the perceptual distance:

$$d = \sqrt{k_{\mathrm{LAT}} \cdot \mathrm{LAT}^2 + k_{\mathrm{SC}} \cdot \mathrm{SC}^2 + k_{\mathrm{SS}} \cdot \mathrm{SS}^2 + k_{\mathrm{SF}} \cdot \mathrm{SF}^2}, \quad (9)$$

where d is the resulting distance between two sounds in this acoustic parameter space, and k is the regression coefficient for each; SC, LAT, SS, and SF are computed as in Equations 5–8.

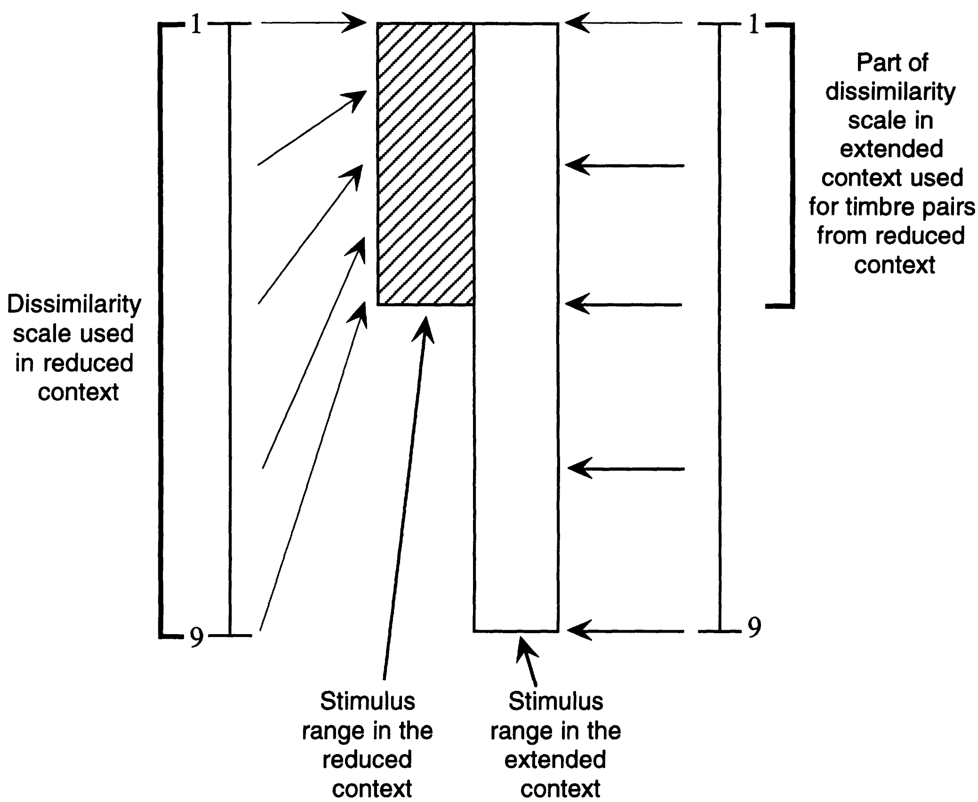
The units for LAT, SC, and SS are $\log_{10}$(sec), Hz, and dB, respectively, whereas SF is unitless. The coefficients for SC and LAT are derived from the mean of the values from the KWW and MWDDK spaces, whereas SS was derived from the (normalized) KWW space and SF from the MWDDK space.

Now comes the real test, however. These parameters were developed on clean, synthesized sounds equalized for pitch, loudness, and duration. In order to apply them to the acoustic sounds recorded in a reverberant environment and varying in pitch and dynamics across a wide variety of playing techniques (normal, flutter tonguing, staccato, multiphonics), a validation step was necessary both in the definition of the perceptual dimensions and in the calculation method of the parameters. This work is still in progress, but for the time being we have focused on the problems of parameter extraction when processing acoustically produced sounds rather than synthesized ones, such as a non-negligible signal-to-noise ratio, natural or forced inharmonicity, high nonstationarity, etc. These factors often cause the algorithms as originally conceived to make enormous computation errors for certain sounds or during certain portions of a given sound. Our current effort is in making the parameter-extraction algorithms more robust in the face of inherently noisy, variable, and sometimes inharmonic, signals. Obviously, in an ideal world we should do an MDS study on the whole database, but over 16,000 sounds (in its current state) would make an experiment with over 250 million trials and, quite frankly, it is hard enough to keep a subject in an experiment that lasts 2 hours, much less 120 years!

Context Effects on Timbre Spaces

One question of interest is whether the perceived or judged relations among certain timbres are affected by the set of timbres included in the experiment. That is, does the dissimilarity judgment for a given pair of timbres change when this pair is presented within the context of different sets of timbres? Some researchers have proposed that the perceived similarity should vary as a function of the context. According to Tversky (1977), each object is represented by a set of feature or attributes. The degree of similarity between two objects is de-

Figure 3. Schema illustrating the scale effect between reduced and extended timbre sets.



fined by a matching function between their common and distinctive features. He proposed that the various features could be weighted differentially according to the stimulus context. When presented in a reduced set, all the dissimilarities have a certain value, but when the same timbres are presented within a larger set that varies along different dimensions and features, the original dissimilarities should decrease owing to the contrast with the new, more widely varying set.

To test this idea, we used the FM timbres described above (Donnadieu, McAdams, and Winsberg 1994; Donnadieu and McAdams 1996). In a given experiment, three sets were presented: two reduced sets with timbres clustered on a small range of values along a given dimension (one high and one low), and an extended set composed of their union. Three clustering conditions were used: (1) fast vs. slow attacks, (2) low vs. high spectral centers of gravity (but spanning the full range of attacks in each reduced set), and (3) low vs. high spectral centers of gravity with all sounds having relatively slow attacks. For the attack condition, dissimilarity judgments were significantly smaller in the extended condition than in either of the reduced conditions. No such effect was observed for the first of the spectral centroid conditions, and a weak effect was found for the second. However, what is important to note—and what does not support Tversky's theory—is that these differences in dissimilarity judgment are primarily due to a change in the judgment scale used, and do not reflect a fundamental change in perception. This can be more easily understood in reference to Figure 3. Note that in all conditions, the same 9-point dissimilarity scale was used, and in each case listeners were asked to use the full scale to make their judgments. In addition, they were always presented with all of the timbres in a set to get a sense of the overall variation available, to know how to use the whole scale. So if we have a reduced set in one condition that is included in an extended set in another condition, the amount of physical change is necessarily smaller in the reduced set. But if this amount of change is mapped onto the same judgment scale, in the reduced condition it covers the whole scale, and in the extended condition it covers only part of the scale. If only the rating strategy changed, and not the actual perceived dissimilarity, one would predict that a scaling factor could relate the judgments on the timbres of the reduced set presented by themselves and in the extended context. This is more or less what is found. What this suggests is that with changes in timbre context, the perceptual relations among the timbres do not change. They thus appear to be relatively robust in the face of context change.

Summary of the MDS Approach

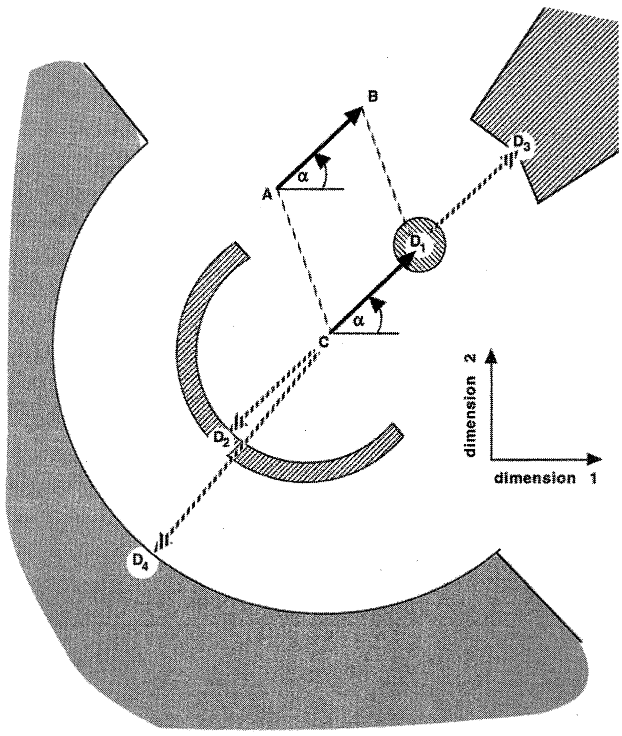
To summarize, we have powerful data-analysis techniques that allow us to extract the structure of perceptual relations among timbres in terms both of dimensions that are shared by a set of tested timbres and of specific dimensions or features that uniquely distinguish timbres from one another. The perceived dissimilarities among timbres are relatively robust, and not affected by changes in timbral context. We can also show that the relative perceptual importance of the individual dimensions and specificities varies among individuals, but that rather than each individual having a completely different set of perceptual weights, listeners tend to form classes that have similar perceptual sensitivities or listening and response strategies. We have been quite successful in determining the acoustic correlates of the continuous shared dimensions, and are in the process of generalizing these findings to robust distance models that can be used in database search engines. However, there remains work to do in refining the parameter extraction, on determining the acoustic nature of the specificities, and on determining the sources of class belongingness among listeners.

Perception of Timbre Intervals

One question we might pose is whether a timbre-space representation can predict anything about timbre perception beyond a simple dissimilarity judgment situation. One successful attempt at showing that this representation has predictive power within a dissimilarity paradigm was made by Grey and Gordon (1978). They wanted to test the psychological reality of the dimension apparently related to spectral distribution that was obtained in Grey's (1977) 3-D space. To do this, they exchanged spectral envelopes between four pairs of sounds, and reinserted the modified sounds in a timbre-space experiment with eight unmodified sounds from the original set. In all cases, as predicted, the relative positions of the modified pairs were inverted with respect to one another along the spectral axis, and (even more interestingly) when the exchange procedure modified other properties of the sounds, they also changed position along some of the other axes, particularly one related to what we now call spectral flux. This was very encouraging for establishing the timbre space as a model of perception. However, we would like to go beyond the dissimilarity paradigm with this representation.

A first step forward in this direction was taken by Ehresman and Wessel (Ehresman and Wessel 1978; Wessel 1979) at IRCAM. Based on previous work on semantic spaces and analogical reasoning (Henley 1969; Rumelhart and Abrahamson 1973), they developed a task in which listeners were asked to make judgments on the similarity of intervals formed between pairs of timbres. The basic idea was that timbre intervals may have properties similar to pitch intervals; that is, a pitch interval is a relation along a well-ordered dimension that retains a degree of invariance under certain kinds of transformation, such as translation along the dimension, or what musicians call "transposition." Ehresman and Wessel wondered whether timbral relations were invariant under transposition as well. But what does transposition mean in a multidimensional space? Figure 4 demonstrates their approach, borrowed from Rumelhart and Abrahamson's work. A timbre interval can be con-

Figure 4. Schema illustrating the choice of different timbres for the end point of the second timbre interval (C-D) to be compared to the first timbre interval (A-B).



sidered as a vector in space connecting two timbres. It has a specific length (the distance between the timbres) and a specific orientation. Together these two properties define the amount of change along each dimension of the space that is needed to move from one timbre to another. If we assume these dimensions to be continuous and linear from a perceptual point of view, then any pair of timbres characterized by the same vector relation should have the same relative perceptual relation, and thus embody the same timbre interval. Transposition thus consists of translating the vector anywhere else in the space as long as its length and orientation are preserved.

Ehresman and Wessel tested this hypothesis on the length (or distance) component. They selected a pair of timbres (A and B) from the set of tones developed by Grey (1977), determined the vector that joined them in the timbre space, and then translated this vector so that its origin corresponded to another tone in the space (C). They hypothesized that an ideal point for comparing the two (D_1) would be close to the end point of the vector, and that the further one moved from this point, the lower the degree of correspondence between the A-B interval and the C-D interval. The interval-comparison task was presented to listeners as an analogy-resolution problem: timbre A is to timbre B as timbre C is to timbre D; rank the various timbre Ds according to how well they fulfill this analogy. They essentially found that the closer timbre D was to the ideal point, the higher the ranking, which supports the vector model of timbre intervals.

Subsequent to this work, McAdams and Cunibile (1992) tested simultaneously the distance and orientation components of the vector using the 3-D KWW space described above (ignoring the specificities). Four cases were considered as shown in Figure 4: (1) timbre D is close to the ideal position (right distance, right orientation); (2) timbre D is about the right distance from C, but at least 90 degrees off in orientation on at least one of the three common dimensions; (3) the C-D vector is in about the right direction, but its length is at least a factor of 1.8 longer than that of vector A-B; and (4) both the orientation and distance are wrong. In each trial, listeners were presented with two sets of four timbres: A-B—C-D; A-B—C-D′, where D and D′ were drawn from two of the four possible cases listed above. Five sets of timbres at different places in timbre space were chosen for each comparison to test for the generality of the results. Both electroacoustic composers and nonmusicians were tested to see if musical training and experience had any effect.

All listeners found the task rather difficult to do, which is not surprising given that even professional composers and computer musicians have had almost no experience with music that uses timbre intervals in a systematic way. The main result is encouraging in that the data globally support the vector model, although this support was much stronger for composers than for nonmusicians. However, when one examines in detail the five different versions of each comparison type, it is clear that not all timbre comparisons go in the direction of the model predictions. Why might this be? Should we abandon the idea of ever developing musical structures based on timbre intervals and timbre melodies in general, or was there some confounding factor in this study? Consider once again the timbre spaces (KWW and MWDDK) derived from the FM timbres that were used in this study.

One factor that came out of the MWDDK study was a difference in the relative perceptual weight accorded to different dimensions by different classes of listeners. While such individual and group differences were found to be very strong in the study of McAdams and colleagues (1995), they could not explain the problem with different timbre sets. Note that the CLASCAL model applies a weighting factor to a whole dimension. This means simply that the space within which the vectors are defined would be uniformly expanded or contracted for certain dimensions, and would not affect the structural relations among the different timbres or the intervals between them. We can therefore rule out latent class belongingness as a problem.

However, the CLASCAL and EXSCAL models also include specificities on certain timbres. These, quite to the contrary, would necessarily distort the vectors that were used to choose the timbres in the McAdams and Cunibile study, since only the positions of the timbres along the three common dimensions were used to calculate the various positions and regions illustrated in Figure 4. As such, certain timbre intervals correspond well to what is predicted since specificities are absent or low in value, whereas others would be seriously distorted and thus not perceived as similar to other intervals owing to moderate or high specificity values. What this line of reasoning suggests is that the use of timbre intervals as an integral part of a musical discourse runs the risk of being difficult to achieve with complex and idiosyncratic sound sources, since they will in all probability have specificities of some kind or another. The use of timbre intervals may, in the long run, be limited to synthesized sounds in which the dimensions of variation can be controlled and tested to ensure that no specificities are present to distort the perceptual relations among them.

Timbre and Auditory Stream Formation

Another aspect of timbre that can contribute to the organization of musical structure is related to the tendency of listeners to perceptually connect sound events that arise from the same sound source. In general, a given source will produce sounds that are relatively similar in pitch, loudness, timbre, and spatial position from one event to the next (see the works of McAdams and Bregman [1979] and Bregman [1990, 1993] for reviews). The perceptual connection of successive sound events into a coherent "message" through time is referred to as auditory stream integration, and the separation of events into distinct messages is called auditory stream segregation (Bregman and Campbell 1971). One guiding principle that seems to operate in the formation of auditory streams is the following: successive events that are relatively similar in their spectrotemporal properties may have arisen from the same source, and should be grouped together; individual sources do not tend to change their acoustic properties suddenly and repeatedly from one event to the next. In the acoustic world, different sources have different timbres, as we have seen with the timbre-space analyses above. Early demonstrations of auditory streaming on the basis of timbre (Wessel 1979) led to suggestions of a link between timbre-space representations and the tendency for auditory streaming on the basis of the spectral differences that were created (McAdams and Bregman 1979). For some time, researchers were convinced that it was primarily the spectral aspects of timbre (such as spectral centroid) that were responsible for auditory streaming, and that temporal aspects (such as attack time) had little effect (Hartmann and Johnson 1991).

Recently the picture has changed significantly, and several studies indicate an important role for both spectral and temporal attributes of timbre in auditory stream segregation. Iverson (1995) used sequences alternating between two recorded-instrument tones with the same pitch and loudness, and asked listeners to judge the degree of segregation. An MDS analysis of the segregation judgments treated as a dissimilarity scale was performed to determine which acoustic attributes contributed to the impression of auditory stream segregation. A comparison with previous timbre-space work using the same sounds (Iverson and Krumhansl 1993) showed that both static acoustic cues (such as spectral centroid) and dynamic acoustic cues (such as attack time and spectral flux) were implicated in segregation. A second experiment used a melody-recognition paradigm in which a target melody with one timbre was to be recognized when interleaved in a sequence with distractor tones played with another timbre. The results of this latter experiment extended those of the first to a situation where tones also varied in pitch.

Bey and McAdams (1997) also used a melody-recognition paradigm. In an initial sequence, an unfamiliar, six-tone target melody was interleaved with a specially constructed distractor melody. The distractor melodies were designed to render the task impossible if the mean pitch difference between the melodies was zero and the timbres and loudnesses of the tones were identical. Then, a second sequence was played consisting either of the target melody alone, or that same melody with two notes changed. Listeners had to decide if the isolated melody was present or not in the preceding mixture. The mixture was presented first to force people to organize it without knowing what they were listening for, and then to use the perceptual representation thus formed to compare with the isolated melody. A number of carefully controlled experiments showed that the probability of correctly recognizing the target is a monotonic function of the mean pitch or timbre difference between the two melodies. All sounds were played with the FM timbres of the KWW space.

In one experiment, the timbre of the distractor melody was chosen to be progressively further away from the target along the spectral centroid axis, keeping the other two dimensions as close as possible. To our great surprise, this difference was insufficient to get high recognition. We then tried varying the timbres along all three axes (attack time, spectral centroid, spectral irregularity), and got good results when the two timbres were in opposite corners of the space. Singh and Bregman (1997) have reported results consistent with the studies of both Iverson (1995) and Bey and McAdams (1997). All of these results are important for auditory stream segregation theory on the one hand, because they show that several of a source's acoustic properties are taken into account when forming auditory streams. On the other hand, they are important for music making (whether it be with computer or acoustic instruments), since they show that many aspects of timbre can be used to organize the musical surface into streams. Timbre thus strongly affects the basic organization of the musical surface into "voices" that will then affect how one perceives the relations among those voices. Different "orchestrations" of a given pitch sequence can completely change what is heard as "melody," as has been demonstrated by Wessel (1979).

Timbre and Musical Tension

The questions posed in this last section concern the contribution of timbre to larger-scale musical form, and in particular to the sense of movement between tension and relaxation. This movement has been considered by many music theorists as one of the primary bases for the perception of larger-scale form in music. It has traditionally been tied to harmony in Western music, and plays an important role in Lerdahl and Jackendoff's *Generative Theory of Tonal Music* (1983). Experimental work on the role of harmony in the perception of musical tension and relaxation (or inversely, in the sense of tension that accompanies a moment at which the music must continue and the sense of relaxation that accompanies the completion of the musical phrase) has suggested that auditory roughness is an important component of perceived tension (Bigand, Parncutt, and Lerdahl 1996). Roughness is an elementary timbral attribute based on the sensation of rapid fluctuations in the amplitude envelope. It can be generated by proximal frequency components that beat with one another. Dissonant intervals tend to have more such beating than consonant intervals. As such, a fairly direct relation between sensory dissonance and roughness has been demonstrated (see the works of Parncutt [1989] and Plomp [1976] for reviews).

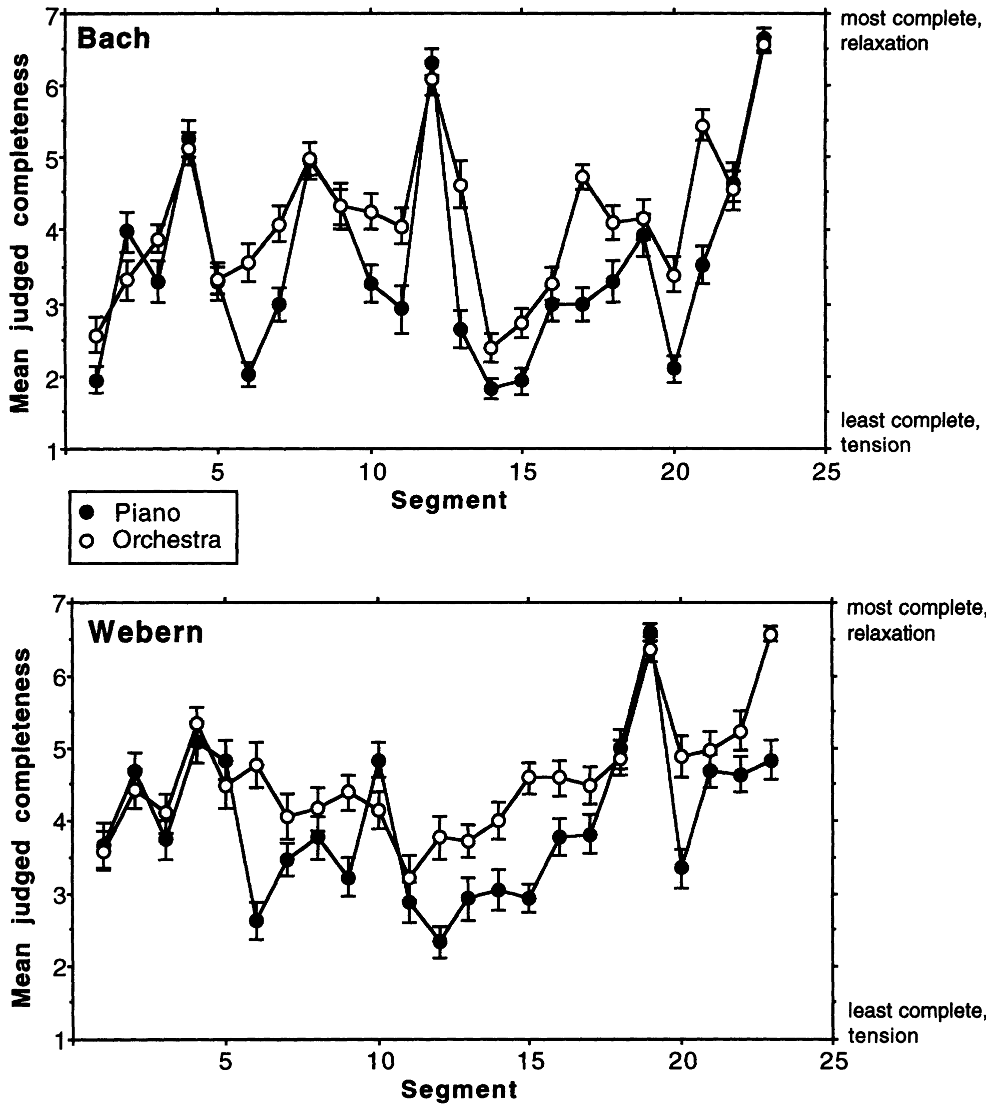
However, for contemporary composers, the interest would be more oriented toward understanding the extent to which movements between tension and relaxation can be created in nontonal sonorities, and the conditions under which such tension can be explained by changes in basic psychoacoustic attributes such as roughness. To address this issue, we used several vertical sonorities that were extracted from the piece *Streamlines* (1995) by Joshua Fineberg and adapted to the needs of our experiment by the composer (Pressnitzer, McAdams, Winsberg, and Fineberg in press). Each chord was played by five instruments (two flutes, clarinet, violin, and viola), and recorded both with a stereo pair at some distance from the ensemble and with directional mikes close to each instrument. The chords were clearly nontonal, including many quarter-tones. In the first experiment, the stereo recording was used. Listeners heard all pairs of the eight chords in both orders and for each pair had to judge whether the movement went from tension to relaxation or from relaxation to tension. The proportion of listeners that judged the movement to be tension-relaxation was computed for each pair, and from this matrix a scale of relative tension was constructed. The same procedure was then applied to roughness (i.e., was the second chord more or less rough than the first?), and a corresponding relative roughness scale was derived. First of all, there were significant differences between the chords in terms both of perceived tension and of perceived roughness, indicating that a sequence of the chords would not be perceived as having a "flat" tension or roughness profile. Second, there was not a perfect correspondence between roughness and tension profiles. A principal components analysis suggested that certain of the chords had unique features (such as an emergent piercing flute note that did not fuse into the sonority, or a prominent perfect fifth that should have been less so). These features apparently affected the tension, but were unrelated to roughness.

A second experiment was thus performed using individually miked tracks that were remixed by the composer to achieve a greater degree of perceptual fusion among the tones, creating a more unified orchestral timbre, as is often done in studio remixes. In particular, he sought to minimize the emergence of certain instruments and of certain tonally charged pitch intervals. With the remixed chords, the same experiment was rerun with a new group of listeners. Again a good distinction among the chords was found giving scales of tension and roughness. However, this time, there was a very strong correlation between the two, indicating that if vertical sonorities (even nontonal ones) are played such as to avoid the emergence of specific features, roughness provides a good explanation for what is perceived as musical tension.

One thing this previous study suggests is that there are numerous timbral factors aside from roughness that may contribute to musical tension. As a first step toward understanding what these may involve, Paraskeva and McAdams (1997) conducted an experiment, the aim of which was both to measure the inflection of musical tension and relaxation due to timbral change and to determine the psychological factors that allow tension and relaxation schemas to be developed. We were particularly interested in: (1) whether orchestration (or synthesis) can be used to inflect the perception of musical tension and release that is otherwise created on the basis of harmonic, rhythmic, and melodic factors; (2) whether there are differences between tonal and nontonal music in this respect; and finally, (3) whether musical training has any influence on these effects.

We used a technique originally developed by Palmer and Krumhansl (1987) and refined by Bigand (1993). In this kind of experiment, a part of a musical phrase is played up to a certain point, and then the music stops. The listener is asked to make a judgment on a seven-point scale concerning the perceived degree of completion of the music at that point. Then the piece is started over again, and the stopping point is a bit further along in the next trial. This continues until the entire piece (or fragment thereof) has been covered. What results is a sort of "completion" profile, which Bigand used to infer musical tension by equating completion with release and lack of completion with tension. A judgment of completion seems easier to make for listeners than a direct judgment of tension. We applied this approach to two pieces: a fragment from the *Ricercar* for six voices by J.S. Bach (tonal) and

Figure 5. Musical tension profiles for piano and orchestra versions of pieces by Bach and Webern.



the first movement of the *Six Pieces for Orchestra* by Anton Webern (nontonal). Each piece was played in an orchestral version (the famous Webern instrumentation was used for the Bach), and a direct transcription of this orchestral version for piano. Both versions were transcribed as MIDI scores, and the same sequence was used to play a sampled piano or a sampled orchestra. Each piece had 23 stopping points. Both professional musicians and nonmusicians participated in the study.

The main result is shown in Figure 5. The stopping point (segment) is indicated on the abscissa and the mean judged completeness (or degree of relaxation) is shown on the ordinate. The first thing to note is that nearly the whole scale was used for both the tonal and the nontonal pieces, indicating that although the degree of fluctuation in tension is generally greater in the tonal work, points of high tension and strong relaxation are experienced in the nontonal work as well, which belies claims by some music psychologists and theorists that such is impossible in nontonal music! Second, note that there are significant differences between the piano versions (filled circles) and orchestral versions (open circles), indicating a significant effect of timbre change on perceived musical tension. However, also note that when they are significantly different, the orchestral version is always more relaxed than the piano version.

Our interpretation of this result is purely speculative at the moment, and needs to be supported by acoustical analyses, modeling, and additional experimentation. This caveat being made, the hypothesis that we advance is that the higher relaxation of the orchestral version may be owing to processes involved in auditory stream formation and the dependence of perceived roughness on the results of such processes (Wright and Bregman 1987). Roughness, or any other auditory attribute of a single sound event, is computed after auditory organization processes have "decided" which bits of acoustic information go together. Piano sounds have a rather sharp attack. If several notes occur at the same time in the score and are played with a piano sound, they will be relatively synchronous (within the limits of MIDI, in our case). Since they all start at the same time and have similar amplitude envelopes, they will tend to be fused together, and the computed roughness will result from the interactions of all the frequency components of all the notes.

The situation may be quite different for the orchestral version for two reasons. The first is that the same timing is used for piano and orchestra versions. In the latter, many instruments are used that have slow attacks, while others have faster attacks. There could then be a great deal of asynchrony between the instruments in terms of perceived attack time (Gordon 1987). In addition, since the timbres of these instruments are often quite different, several different voices with different timbres arrive momentarily at a given vertical sonority, but the verticality is not perceived because the listener would more likely continue to track individual instruments horizontally. So the attack asynchrony and the decomposition of verticalities into horizontalities would concur to reduce the degree of perceptual fusion. Reduced fusion would mean greater segregation. And thus the roughness in the orchestral version would be computed on each individually grouped auditory event, rather than on the whole sound mass. These individual roughnesses in the orchestral version would most likely be much less than that of the piano version. So once again, timbral composition has a very tight interaction with auditory stream formation processes. Here, however, multiplying the timbres used to orchestrate a given pitch and duration structure affects the way it ends up being organized and perceived by the listener, in terms of the auditory attributes computed and their contribution to musical tension and relaxation.

Conclusion

Taken as a whole, the work reported here presents convincing evidence that timbre can play an important role in music perception. It also suggests that we can acquire knowledge about timbre in a rigorous fashion that allows for the development of predictive models for its perceptual effects and possibilities.

We have shown that musical timbre is a combination of continuous perceptual dimensions and discrete features to which listeners are differentially sensitive. The continuous dimensions often have quantifiable acoustic correlates that can then be incorporated into predictive models. The timbre-space representation is a powerful model that allows predictions to be made about timbre perception in situations beyond those used to derive the model. Timbre intervals, for example, can be conceived as vectors within the space of common dimensions. Although the modeling of the interval relations can be perturbed if the sounds have specificities, it would not be affected by differential sensitivity of individual listeners to the common dimensions, since these would expand and contract all relations in a systematic way.

Timbre space also makes at least qualitative predictions about the magnitude of timbre differences that will provoke auditory stream segregation: the further apart the timbres are in the space, the greater the probability that interleaved pitch sequences played with them will form separate streams, thereby allowing independent perception and recognition of the constituent sequences.

Timbre can also play a role in larger-scale movements of tension and relaxation, and thus contribute to the expression inherent in musical form. Under conditions of high blend among instruments composing a vertical sonority, timbral roughness is a major component of musical tension. However, it strongly depends, as do all auditory attributes, on the way auditory grouping processes have parsed the incoming acoustic information into events and streams. And finally, orchestration can play a major role in addition to pitch and rhythmic structure in the structuring of musical tension and relaxation schemas that are an important component of the esthetic response to musical form.

Acknowledgments

Portions of this article were presented in an invited address in the Music and the Brain Symposium at ICMC97 in Thessaloniki. The work reported here has benefited from collaborations over a period of about 10 years with several colleagues and students in the Music Perception and Cognition team at IRCAM and in the Auditory Perception team at the Laboratoire de Psychologie Expérimentale: Caroline Bey, Jean-Christophe Cunibile, Sophie Donnadieu, Joshua Fineberg, Christian Kluxen, Jochen Krimphoff, Nicolas Misdariis, Stella Paraskeva, Daniel Pressnitzer, Vincent Roussarie, Bennett Smith, Patrick Susini, David Wessel, and Suzanne Winsberg.

References

Beauchamp, J. W. 1993. "UNIX Workstation Software for Analysis, Graphics, Modifications, and Synthesis of Musical Sounds." Presented at the 94th Convention of the Audio Engineering Society. Berlin, New York: AES.

Bey, C., and S. McAdams. 1997 (Marseille). "Etudes des Processus de Formation des Flux Auditifs par une Méthode Objective [Studies on the process of auditory stream formation with an objective method]." *Actes du 4e Congrès Français d'Acoustique*, vol. 1. Paris: Societe Français d'Acoustique.

Bigand, E. 1993. "The Influence of Implicit Harmony, Rhythm, and Musical Training on the Abstraction of Tension-Relaxation Schemas in Tonal Musical Phrases." *Contemporary Music Review* 9:123–137.

Bigand, E., R. Parncutt, and F. Lerdahl. 1996. "Perception of Musical Tension in Short Chord Sequences: The Influence of Harmonic Function, Sensory Dissonance, Horizontal Motion, and Musical Training." *Perception and Psychophysics* 58:125–141.

Bregman, A. S. 1990. *Auditory Scene Analysis: The Perceptual Organization of Sound.* Cambridge, Massachusetts: MIT Press.

Bregman, A. S. 1993. "Auditory Scene Analysis: Hearing in Complex Environments." In S. McAdams and E. Bigand, eds. *Thinking in Sound: The Cognitive Psychology of Human Audition.* Oxford: Oxford University Press.

Bregman, A. S., and J. Campbell. 1971. "Primary Auditory Stream Segregation and Perception of Order in Rapid Sequences of Tones." *Journal of Experimental Psychology* 89:244–249.

Carroll, J. D., and J. J. Chang. 1970. "Analysis of Individual Differences in Multidimensional Scaling via an N-way Generalization of Eckart-Young Decomposition." *Psychometrika* 35:283–319.

Depalle, P., G. García, and X. Rodet. 1993. "Tracking of Partials for Additive Sound Synthesis Using Hidden Markov Models." *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing.* Piscataway, New Jersey: IEEE

Donnadieu, S., and S. McAdams.1996. "Effects of Context Change on Dissimilarity, Discrimination, and Categorisation Tasks on Timbre Perception." *Fechner Day 96: 12th Annual Meeting of the International Society for Psychophysics.* Padova: International Society for Psychophysics.

Donnadieu, S., S. McAdams, and S. Winsberg. 1994. "Context Effects in 'Timbre Space.'" *Proceedings of the 3rd International Conference on Music Perception and Cognition.* Liège: European Society for the Cognitive Sciences of Music.

Ehresman, D., and D. L. Wessel. 1978. "Perception of Timbral Analogies." IRCAM Report 13, Paris.

Gordon, J. W. 1987. "The Perceptual Attack Time of Musical Tones." *Journal of the Acoustical Society of America* 82:88–105.

Grey, J. M. 1977. "Multidimensional Perceptual Scaling of Musical Timbres." *Journal of the Acoustical Society of America* 61:1270–1277.

Grey, J. M., and J. W. Gordon. 1978. "Perceptual Effects of Spectral Modifications on Musical Timbres." *Journal of the Acoustical Society of America* 63:1493–1500.

Hajda, J. M., R. A. Kendall, E. C. Carterette, and M. L. Harshberger. 1997. "Methodological Issues in Timbre Research." In I. Deliège and J. Sloboda, eds. *Perception and Cognition of Music*. Hove: Psychology Press.

Hartmann, W. M., and D. Johnson. 1991. "Stream Segregation and Peripheral Channeling." *Music Perception* 9:155–184.

Henley, N. M. 1969. "A Psychological Study of the Semantics of Animal Terms." *Journal of Verbal Learning and Verbal Behavior* 8:176–184.

Iverson, P. 1995. "Auditory Stream Segregation by Musical Timbre: Effects of Static and Dynamic Acoustic Attributes." *Journal of Experimental Psychology: Human Perception and Performance* 21:751–763.

Iverson, P., and C. L. Krumhansl. 1993. "Isolating the Dynamic Attributes of Musical Timbre." *Journal of the Acoustical Society of America* 94:2595–2603.

Krimphoff, J., S. McAdams, and S. Winsberg. 1994. "Caractérisation du Timbre des sons Complexes. II: Analyses Acoustiques et Quantification Psychophysique [Characterizing the timbre of complex sounds. II: Acoustic analyses and psychophysical quantification]." *Journal de Physique* 4(C5):625–628.

Krumhansl, C. L. 1989. "Why is Musical Timbre So Hard to Understand?" In S. Nielzén and O. Olsson, eds. *Structure and Perception of Electroacoustic Sound and Music*. Amsterdam: Excerpta Medica.

Kruskal, J. B. 1964a. "Multidimensional Scaling by Optimizing Goodness of Fit to a Nonmetric Hypothesis." *Psychometrika* 29:1–27.

Kruskal, J. B. 1964b. "Nonmetric Multidimensional Scaling: A Numerical Method." *Psychometrika* 29:115–129.

Lerdahl, F., and R. Jackendoff. 1983. *The Generative Theory of Tonal Music*. Cambridge, Massachusetts: MIT Press.

McAdams, S. 1989. "Psychological Constraints on Form-Bearing Dimensions in Music." *Contemporary Music Review* 4(1):181–198.

McAdams, S. 1993. "Recognition of Sound Sources and Events." In S. McAdams and E. Bigand, eds. *Thinking in Sound: The Cognitive Psychology of Human Audition*. Oxford: Oxford University Press.

McAdams, S., and A. S. Bregman. 1979. "Hearing Musical Streams." *Computer Music Journal* 3(4):26–43.

McAdams, S., and J.-C. Cunibile. 1992. "Perception of Timbre Analogies." *Philosophical Transactions of the Royal Society London*, Series B 336:383–390.

McAdams, S., P. Susini, J. Krimphoff, N. Misdariis, and B. K. Smith. In preparation. A Meta-Analysis of Timbre Space. II: Acoustic Correlates of Common Dimensions.

McAdams, S., and S. Winsberg. In preparation. A Meta-Analysis of Timbre Space. I: Multidimensional Scaling of Group Data with Common Dimensions, Specificities, and Latent Subject Classes.

McAdams, S., S. Winsberg, S. Donnadieu, G. De Soete, and J. Krimphoff. 1995. "Perceptual Scaling of Synthesized Musical Timbres: Common Dimensions, Specificities, and Latent Subject Classes." *Psychological Research* 58:177–192.

Miller, J. R., and E. C. Carterette. 1975. "Perceptual Space for Musical Structures." *Journal of the Acoustical Society of America* 58:711–720.

Misdariis, N., B. K. Smith, D. Pressnitzer, P. Susini, and S. McAdams. 1998 (Seattle, Washington). "Validation of a Multidimensional Distance Model for Perceptual Dissimilarities among Musical Timbres." *Proceedings of the 16th International Congress on Acoustics*. Woodbury, New York: ASA.

Palmer, C., and C. L. Krumhansl. 1987. "Independent Temporal and Pitch Structures in the Determination of Musical Phrases." *Journal of Experimental Psychology: Human Perception and Performance* 13:116–126.

Paraskeva, S., and S. McAdams. 1997 (Thessaloniki). "Influence of Timbre, Presence/Absence of Tonal Hierarchy and Musical Training on the Perception of Tension/Relaxation Schemas of Musical Phrases." *Proceedings of the 1997 International Computer Music Conference*. San Francisco: ICMA.

Parncutt, R. 1989. *Harmony: A Psychoacoustical Approach*. Berlin: Springer-Verlag.

Plomp, R. 1970. "Timbre as a Multidimensional Attribute of Complex Tones." In R. Plomp and G. F. Smoorenburg, eds. *Frequency Analysis and Periodicity Detection in Hearing*. Leiden: Sijthoff.

Plomp, R. 1976. *Aspects of Tone Sensation: A Psychophysical Study*. London: Academic Press.

Pressnitzer, D., S. McAdams, S. Winsberg, and J. Fineberg. "Perception of Musical Tension for Nontonal Orchestral Timbres and Its Relation to Psychoacoustic Roughness." *Perception and Psychophysics* (in press).

Risset, J.-C., and D. L. Wessel. 1999. "Exploration of Timbre by Analysis and Synthesis." In D. Deutsch, ed. *The Psychology of Music*, 2nd ed. San Diego: Academic Press.

Roussarie, V., S. McAdams, and A. Chaigne. 1998 (Seattle, Washington). "Perceptual Analysis of Vibrating Bars Synthesized with a Physical Model." *Proceed-*

ings of the 16th International Congress on Acoustics. Woodbury, New York: ASA.
Rumelhart, D. E., and A. A. Abrahamson. 1973. "A Model for Analogical Reasoning." *Cognitive Psychology* 5:1–28.
Singh, P. G., and A. S. Bregman. 1997. "The Influence of Different Timbre Attributes on the Perceptual Segregation of Complex-Tone Sequences." *Journal of the Acoustical Society of America* 120:1943–1952.
Susini, P., S. McAdams, and S. Winsberg. "Perceptual Characterisation of Vehicle Noises." *Acta Acustica* (in press).
Tversky, A. 1977. "Features of Similarity." *Psychological Review* 84:327–352.
Wessel, D. L. 1973. "Psychoacoustics and Music: A Report from Michigan State University." *PACE: Bulletin of the Computer Arts Society* 30:1–2.
Wessel, D. L. 1979. "Timbre Space as a Musical Control Structure." *Computer Music Journal* 3(2):45–52.
Wessel, D. L., D. Bristow, and Z. Settel. 1987. "Control of Phrasing and Articulation in Synthesis." *Proceedings of the 1987 International Computer Music Conference.* San Francisco: ICMA.
Winsberg, S., and J. D. Carroll. 1988. "A Quasi-Nonmetric Method for Multidimensional Scaling via an Extended Euclidean Model." *Psychometrika* 53:217–229.
Winsberg, S., and J. D. Carroll. 1989. "A Quasi-Nonmetric Method for Multidimensional Scaling of Multiway Data via a Restricted Case of an Extended INDSCAL Model." In R. Coppi and S. Bolasco, eds. *Multi-Way Data Analysis.* Amsterdam: North-Holland.
Winsberg, S., and G. De Soete. 1993. "A Latent Class Approach to Fitting the Weighted Euclidean Model: CLASCAL." *Psychometrika* 58:315–330.
Wright, J. K., and A. S. Bregman. 1987. "Auditory Stream Segregation and the Control of Dissonance in Polyphonic Music." *Contemporary Music Review* 2(1):63–92.