Musical Timbre Perception

Chapter *in* Psychology of Music · December 2013

DOI: 10.1016/B978-0-12-381460-9.00002-X

CITATIONS
103

READS
2,329

1 author:

Stephen Mcadams
McGill University
420 PUBLICATIONS **11,290** CITATIONS

SEE PROFILE

The Psychology of Music

Third Edition

Edited by

Diana Deutsch
Department of Psychology
University of California, San Diego
La Jolla, California

AMSTERDAM • BOSTON • HEIDELBERG • LONDON • NEW YORK
OXFORD • PARIS • SAN DIEGO • SAN FRANCISCO • SINGAPORE
SYDNEY • TOKYO
Academic Press is an imprint of Elsevier

Sound examples available on
http://www.music.mcgill.ca/~smc/McAdams_2013_PsychMus_Sounds.pptx

2 Musical Timbre Perception

Stephen McAdams

McGill University, Montreal, Quebec, Canada

Timbre is a misleadingly simple and exceedingly vague word encompassing a very complex set of auditory attributes, as well as a plethora of intricate psychological and musical issues. It covers many parameters of perception that are not accounted for by pitch, loudness, spatial position, duration, or even by various environmental characteristics such as room reverberation. This leaves myriad possibilities, some of which have been explored during the past 40 years or so.

We now understand timbre to have two broad characteristics that contribute to the perception of music: (1) it is a multitudinous set of perceptual attributes, some of which are continuously varying (e.g., attack sharpness, brightness, nasality, richness), others of which are discrete or categorical (e.g., the "blatt" at the beginning of a sforzando trombone sound or the pinched offset of a harpsichord sound), and (2) it is one of the primary perceptual vehicles for the recognition, identification, and tracking over time of a sound source (singer's voice, clarinet, set of carillon bells) and thus is involved in the absolute categorization of a sounding object (Hajda, Kendall, Carterette & Harshberger, 1997; Handel, 1995; McAdams, 1993; Risset, 2004).

Understanding the perception of timbre thus covers a wide range of issues from determining the properties of vibrating objects and of the acoustic waves emanating from them, developing techniques for quantitatively analyzing and characterizing sound waves, formalizing models of how the acoustic signal is analyzed and coded neurally by the auditory system, characterizing the perceptual representation of the sounds used by listeners to compare sounds in an abstract way or to categorize or identify their physical source, to understanding the role that timbre can play in perceiving musical patterns and forms and shaping musical performance expressively. More theoretical approaches to timbre have also included considerations of the musical implications of timbre as a set of form-bearing dimensions in music (cf. McAdams, 1989). This chapter will focus on some of these issues in detail: the psychophysics of timbre, timbre as a vehicle for source identity, the role of timbre in musical grouping, and timbre as a structuring force in music perception, including the effect of sound blending on the perception of timbre, timbre's role in the grouping of events into streams and musical patterns, the perception of timbral intervals, the role of timbre in the building and release of musical tension, and implicit learning of timbral grammars. A concluding section will examine a number of issues that have not been extensively studied yet concerning the role of timbre

The Psychology of Music. DOI: http://dx.doi.org/10.1016/B978-0-12-381460-9.00002-X

characterization in music information retrieval systems, control of timbral variation by instrumentalists and sound synthesis control devices to achieve musical expressiveness, the link between timbre perception and cognition and orchestration and electroacoustic music composition, and finally, consideration of timbre's status as a primary or secondary parameter in musical structure.[1]

I. Psychophysics of Timbre

One of the main approaches to timbre perception attempts to characterize quantitatively the ways in which sounds are perceived to differ. Early research on the perceptual nature of timbre focused on preconceived aspects such as the relative weights of different frequencies present in a given sound, or its "sound color" (Slawson, 1985). For example, both a voice singing a constant middle C while varying the vowel being sung and a brass player holding a given note while varying the embouchure and mouth cavity shape would vary the shape of the sound spectrum (cf. McAdams, Depalle & Clarke, 2004). Helmholtz (1885/1954) invented some rather ingenious resonating devices for controlling spectral shape to explore these aspects of timbre. However, the real advances in understanding the perceptual representation of timbre had to wait for the development of signal generation and processing techniques and of multidimensional data analysis techniques in the 1950s and 1960s. Plomp (1970) and Wessel (1973) were the first to apply these to timbre perception.

A. Timbre Space

Multidimensional scaling (MDS) makes no preconceptions about the physical or perceptual structure of timbre. Listeners simply rate on a scale varying from very similar to very dissimilar all pairs from a given set of sounds. The sounds are usually equalized in terms of pitch, loudness, and duration and are presented from the same location in space so that only the timbre varies in order to focus listeners' attention on this set of attributes. The dissimilarity ratings are then fit to a distance model in which sounds with similar timbres are closer together and those with dissimilar timbres are farther apart. The analysis approach is presented in Figure 1. The graphic representation of the distance model is called a "timbre space." Such techniques have been applied to synthetic sounds (Miller & Carterette, 1975; Plomp, 1970; Caclin, McAdams, Smith & Winsberg, 2005), resynthesized or simulated instrument sounds (Grey, 1977; Kendall, Carterette, & Hajda, 1999; Krumhansl, 1989; McAdams, Winsberg, Donnadieu, De Soete & Krimphoff, 1995; Wessel, 1979), recorded instrument sounds (Iverson & Krumhansl, 1993; Lakatos,

[1] In contrast to the chapter on timbre in the previous editions of this book, less emphasis will be placed on sound analysis and synthesis and more on perception and cognition. Risset and Wessel (1999) remains an excellent summary of these former issues.

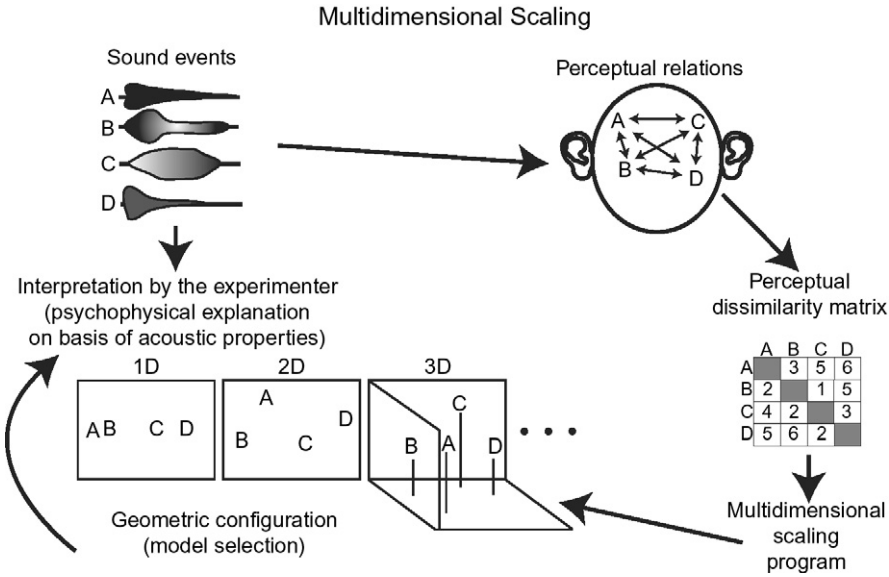

Figure 1 Stages in the multidimensional analysis of dissimilarity ratings of sounds differing in timbre.

2000; Wessel, 1973), and even dyads of recorded instrument sounds (Kendall & Carterette, 1991; Tardieu & McAdams, in press).

The basic MDS model, such as Kruskal's (1964a, 1964b) nonmetric model, is expressed in terms of continuous dimensions that are shared among the timbres, the underlying assumption being that all listeners use the same perceptual dimensions to compare the timbres. The model distances are fit to the empirically derived proximity data (usually dissimilarity ratings or confusion ratings among sounds). More complex models also include dimensions or features that are specific to individual timbres, called "specificities" (EXSCAL, Winsberg & Carroll, 1989) and different perceptual weights accorded to the dimensions and specificities by individual listeners or latent classes of listeners (INDSCAL, Carroll & Chang, 1970; CLASCAL, Winsberg & De Soete, 1993; McAdams et al., 1995). The equation defining distance in the more general CLASCAL model is the following:

$$d_{ijt} = \left[\sum_{r=1}^{R} w_{tr}(x_{ir} - x_{jr})^2 + v_t(s_i + s_j)\right]^{\frac{1}{2}}, \qquad \text{(Eq. 1)}$$

where d_{ijt} is the distance between sounds i and j for latent class t, x_{ir} is the coordinate of sound i on dimension r, R is the total number of dimensions, w_{tr} is the weight on dimension r for class t, s_i is the specificity on sound i, and v_t is the weight on the whole set of specificities for class t. The basic model doesn't have

weights or specificities and has only one class of listeners. EXCAL has specificities, but no weights. For INDSCAL, the number of latent classes is equal to the number of listeners. Finally, the CONSCAL model allows for continuous mapping functions between audio descriptors and the position of sounds along a perceptual dimension to be modeled for each listener by using spline functions, with the proviso that the position along the perceptual dimension respect the ordering along the physical dimension (Winsberg & De Soete, 1997). This technique allows one to determine the auditory transform of each physical parameter for each listener. Examples of the use of these different analysis models include Kruskal's technique by Plomp (1970), INDSCAL by Wessel (1973) and Grey (1977), EXSCAL by Krumhansl (1989), CLASCAL by McAdams et al. (1995) and CONSCAL by Caclin et al. (2005). Descriptions of how to use the CLASCAL and CONSCAL models in the context of timbre research are provided in McAdams et al. (1995) and Caclin et al. (2005), respectively.

Specificities are often found for complex acoustic and synthesized sounds. They are considered to represent the presence of a unique feature that distinguishes a sound from all others in a given context. For example, in a set of brass, woodwind, and string sounds, a harpsichord has a feature shared with no other sound: the return of the hopper, which creates a slight "thump" and quickly damps the sound at the end. Or in a set of sounds with fairly smooth spectral envelopes such as brass instruments, the jagged spectral envelope of the clarinet due to the attenuation of the even harmonics at lower harmonic ranks would be a feature specific to that instrument. Such features might appear as specificities in the EXSCAL and CLASCAL distance models (Krumhansl, 1989; McAdams et al., 1995), and the strength of each feature is represented by the square root of the specificity value in Equation 1.

Some models include individual and class differences as weighting factors on the different dimensions and the set of specificities. For example, some listeners might pay more attention to spectral properties than to temporal aspects, whereas others might have the inverse pattern. Such variability could reflect either differences in sensory processing or in listening and rating strategies. Interestingly, no study to date has demonstrated that such individual differences have anything to do with musical experience or training. For example, McAdams et al. (1995) found that similar proportions of nonmusicians, music students, and professional musicians fell into the different latent classes, suggesting that whereas listeners differ in terms of the perceptual weight accorded to the different dimensions, these interindividual differences are unrelated to musical training. It may be that because timbre perception is so closely allied with the ability to recognize sound sources in everyday life, everybody is an expert to some degree, although different people are sensitive to different features.

An example timbre space, drawn from McAdams et al. (1995), is shown in Figure 2. It is derived from the dissimilarity ratings of 84 listeners including nonmusicians, music students, and professional musicians. Listeners were presented digital simulations of instrument sounds and chimæric sounds combining features of different instruments (such as the *vibrone* with both vibraphonelike and

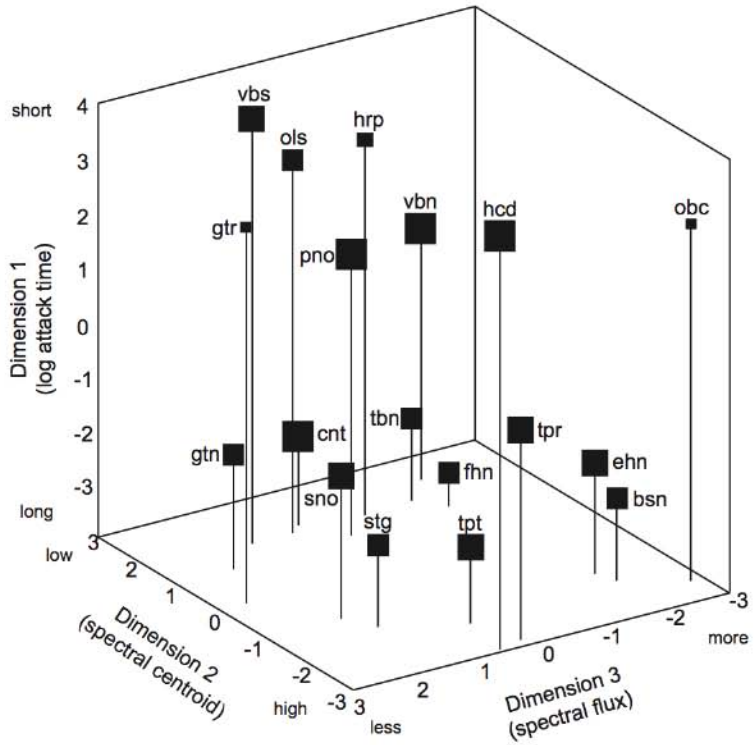

Figure 2 The timbre space found by McAdams et al. (1995) for a set of synthesized sounds. The CLASCAL solution has three dimensions with specificities (the strength of the specificity is shown by the size of the square). The acoustic correlates for each dimension are also indicated. (vbs = vibraphone, hrp = harp, ols = *obolesta* (oboe/celesta hybrid), gtr = guitar, pno = piano, vbn = *vibrone* (vibraphone/trombone hybrid), hcd = harpsichord, obc = *obochord* (oboe/harpsichord hybrid), gtn = *guitarnet* (guitar/clarinet hybrid), cnt = clarinet, sno = *striano* (bowed string/piano hybrid), tbn = trombone, fhn = French horn, stg = bowed string, tpr = *trumpar* (trumpet/guitar hybrid), ehn = English horn, bsn = bassoon, tpt = trumpet).
Modified from Figure 1, McAdams et al. (1995). ©1995 by Springer-Verlag. Adapted with permission.

trombonelike features). Wessel, Bristow, and Settel (1987) created these sounds on a Yamaha DX7 FM synthesizer. A CLASCAL analysis revealed three shared dimensions, the existence of specificities on the sounds, and five latent classes of listeners, for whom the relative weights on the shared dimensions and set of specificities differed.

The relative weights on the three dimensions and the set of specificities for the five latent classes are shown in Figure 3. Most listeners were in classes 1 and 2 and had fairly equal weights across dimensions and specificities. What distinguished these two classes was simply the use of the rating scale: Class 1 listeners used

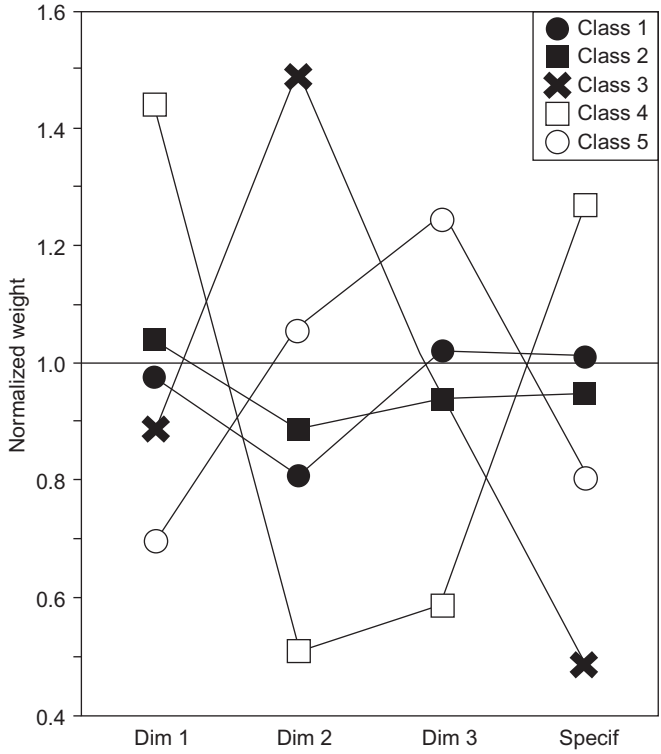

Figure 3 Normalized weights on the three shared dimensions and the set of specificities for five latent classes of listeners in the McAdams et al. (1995) study.

more of the scale than did listeners from Class 2. For the other three classes, however, some dimensions were prominent (high weights) and others were perceptually attenuated (low weights). For example, Class 3 listeners gave high weight to Dimension 2, which seems to be related to spectral characteristics of the sounds, and low weight on the specificities. Inversely, Class 4 listeners favored Dimension 1 (related to the temporal dimension of attack time) and the specificities and attenuated the spectral (Dim 2) and spectrotemporal (Dim 3) dimensions.

Timbre space models have been useful in predicting listeners' perceptions in situations beyond those specifically measured in the experiments, which suggests that they do in fact capture important aspects of timbre representation. Consistent with the predictions of a timbre model, Grey and Gordon (1978) found that by exchanging the spectral envelopes on pairs of sounds that differed primarily along one of the dimensions of their space believed to be related to spectral properties, these sounds switched positions along this dimension. Timbre space has also been useful in predicting the perception of intervals between timbres, as well as stream segregation based on timbre-related acoustic cues (see below).

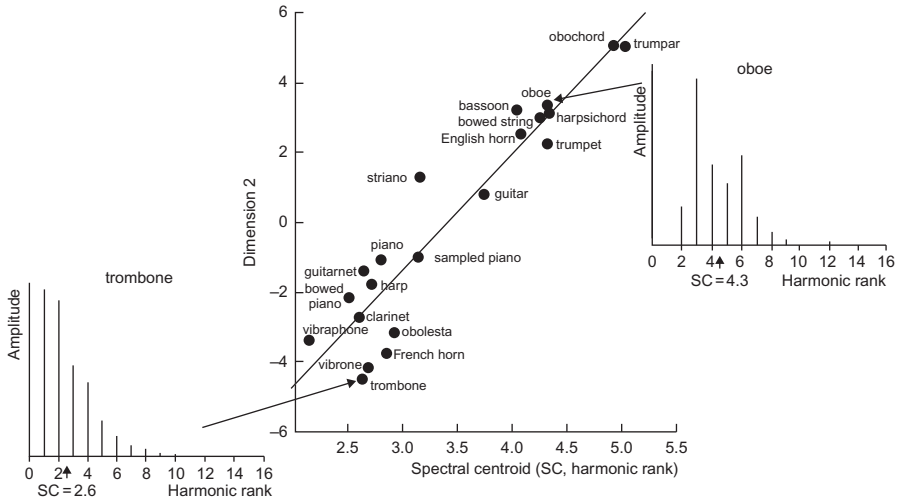

Figure 4 Spectral centroid in relation to the second dimension of Krumhansl's (1989) space using the synthesized sounds from Wessel et al. (1987). The graphs at the left and right represent the frequency spectra of two of the sounds (trombone and oboe, respectively). The arrowhead on the *x* axis indicates the location of the spectral centroid. The graph in the middle shows the regression of spectral centroid (*x* axis) onto the position along the perceptual dimension (*y* axis). Note that all the points are very close to the regression line, indicating a close association between the physical and perceptual parameters.

B. *Audio Descriptors of Timbral Dimensions*

In many studies, independent acoustic correlates have been determined for the continuous dimensions by correlating the position along the perceptual dimension with a unidimensional acoustic parameter extracted from the sounds (e.g., Grey & Gordon, 1978; Kendall et al., 1999; Krimphoff, McAdams, & Winsberg, 1994; McAdams et al., 1995). We will call such parameters "audio descriptors," although they are also referred to as audio features in the field of music information retrieval. The most ubiquitous correlates derived from musical instrument sounds include spectral centroid (representing the relative weights of high and low frequencies and corresponding to timbral brightness or nasality: an oboe has a higher spectral centroid than a French horn; see Figure 4), the logarithm of the attack time (distinguishing continuant instruments that are blown or bowed from impulsive instruments that are struck or plucked; see Figure 5), spectral flux (the degree of evolution of the spectral shape over a tone's duration which is high for brass and lower for single reeds; see Figure 6), and spectral deviation (the degree of jaggedness of the spectral shape, which is high for clarinet and vibraphone and low for trumpet; see Figure 7). Caclin et al. (2005) conducted a confirmatory study employing dissimilarity ratings on purely synthetic sounds in which the exact nature of the stimulus dimensions could be controlled. These authors confirmed the

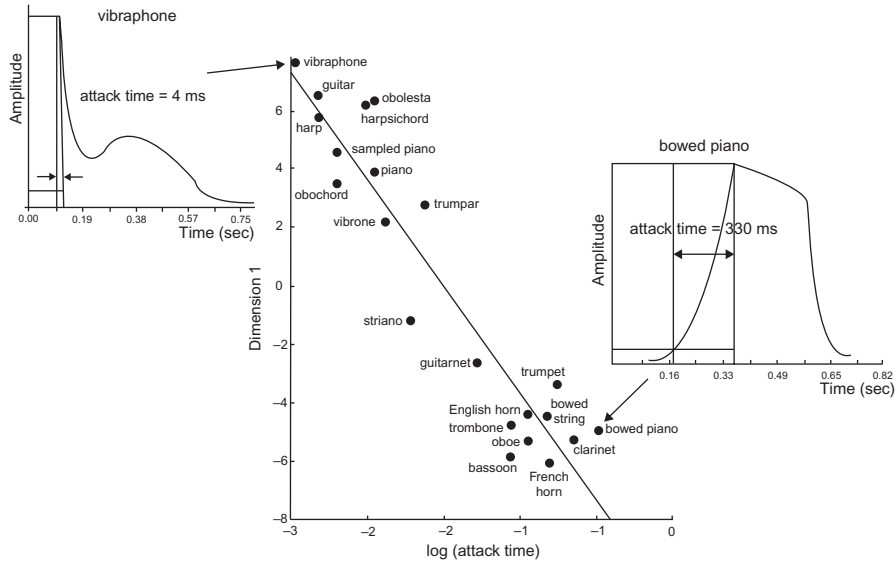

Figure 5 Log attack time in relation to the first dimension of Krumhansl's (1989) space. The graphs on the left and right sides show the amplitude envelopes of the vibraphone and bowed piano sounds. The attack time is indicated by the arrows.

perception of stimulus dimensions related to spectral centroid, log attack time, and spectral deviation but did not confirm spectral flux.

Of the studies attempting to develop audio descriptors that are correlated with the perceptual dimensions of their timbre spaces, most have focused on a small set of sounds and a small set of descriptors. Over the years, a large set of descriptors has been developed at IRCAM (Institut de Recherche et Coordination Acoustique/Musique) starting with the work of Jochen Krimphoff (Krimphoff et al., 1994). The aim was to represent a wide range of temporal, spectral, and spectrotemporal properties of the acoustic signals that could be used as metadata in content-based searches in very large sound databases. The culmination of this work has recently been published (Peeters, Giordano, Susini, Misdariis, & McAdams, 2011) and the Timbre Toolbox has been made available in the form of a Matlab toolbox[2] that contains a set of 54 descriptors based on energy envelope, short-term Fourier transform, harmonic sinusoidal components, or the gamma-tone filter-bank model of peripheral auditory processing (Patterson, Allerhand, & Giguère, 1995). These audio descriptors capture temporal, spectral, spectrotemporal, and energetic properties of acoustic events. Temporal descriptors include properties such as attack, decay, release, temporal centroid, effective duration, and the frequency and amplitude of modulation in the energy envelope. Spectral shape descriptors include

[2] http://recherche.ircam.fr/pub/timbretoolbox or http://www.cirmmt.mcgill.ca/research/tools/timbretoolbox

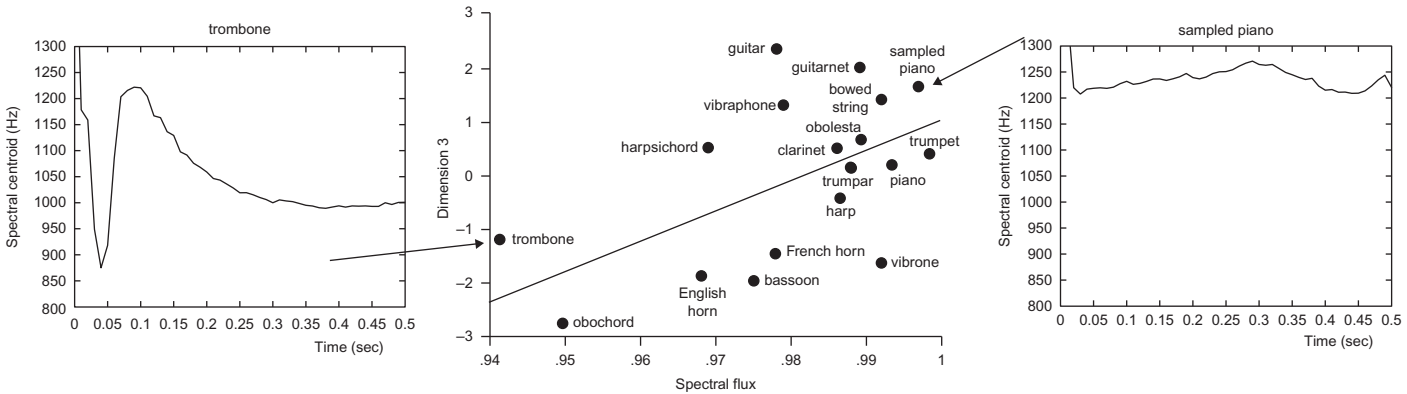

Figure 6 Spectral flux in relation to the third dimension of the space found by McAdams et al. (1995). The left and right graphs show the variation over time of the spectral centroid for the trombone and the sampled piano. Note that the points are more spread out around the regression line in the middle graph, indicating that this physical parameter explains much less of the variance in the positions of the sounds along the perceptual dimension.

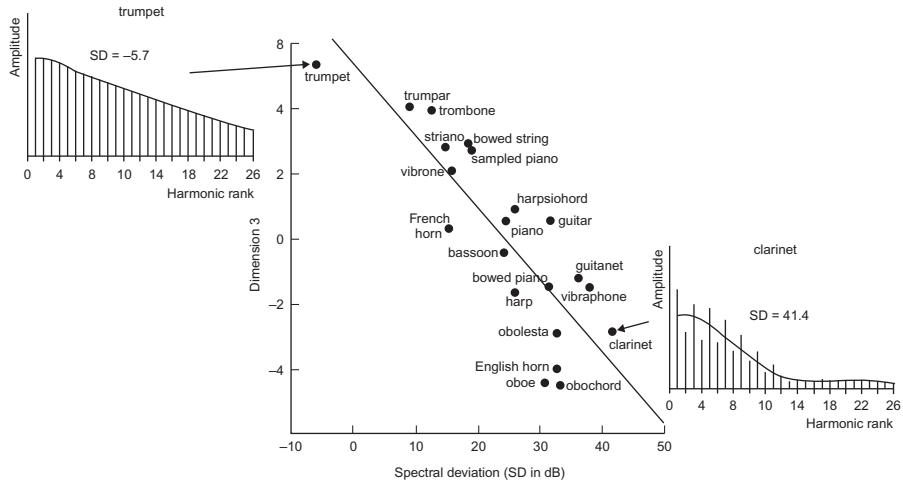

Figure 7 Spectral deviation in relation to the third dimension of the space found by Krumhansl (1989). The left and right graphs show the frequency spectra and global spectral envelopes of the trumpet and clarinet sounds. Note that the amplitudes of the frequency components are close to the global envelope for the trumpet, but deviate above and below this envelope for the clarinet.

measures of the centroid, spread, skewness, kurtosis, slope, rolloff, crest factor, and jaggedness of the spectral envelope. Spectrotemporal descriptors include spectral flux. Energetic descriptors include harmonic energy, noise energy, and statistical properties of the energy envelope. In addition, descriptors related to periodicity/harmonicity and noisiness were included. Certain of these descriptors have a single value for a sound event, such as attack time, whereas others represent time-varying quantities, such as the variation of spectral centroid over the duration of a sound event. Statistical properties of these time-varying quantities can then be used, such as measures of central tendency or variability (robust statistics of median and interquartile range were used by Peeters et al., 2011).

One problem with a large number of descriptors is that they may be correlated among themselves for a given set of sounds, particularly if they are applied to a limited sound set. Peeters et al. (2011) examined the information redundancy across the audio descriptors by performing correlational analyses between descriptors calculated on a very large set of highly heterogeneous musical sounds (more than 6000 sounds from the McGill University Master Samples, MUMS; Opolko & Wapnick, 2006). They then subjected the resulting correlation matrix to hierarchical clustering. The analysis also sought to assess whether the Timbre Toolbox could account for the dimensional richness of real musical sounds and to provide a user of the Toolbox with a set of guidelines for selecting among the numerous descriptors implemented therein. The analyses yielded roughly 10 classes of descriptors that are relatively independent. Two clusters represented spectral shape

properties, one based primarily on median values (11 descriptors) and the other uniquely on the interquartile ranges of the time-varying measures of these spectral properties (7 descriptors). Thus central tendencies and variability of spectral shape behave independently across the MUMS database. A large third cluster of 16 descriptors included most of the temporal descriptors, such as log attack time, and energetic descriptors, such as variability in noise energy and total energy over time. A fourth large cluster included 10 descriptors related to periodicity, noisiness, and jaggedness of the spectral envelope. The remaining smaller clusters had one or two descriptors each and included descriptors of spectral shape, spectral variation, and amplitude and frequency of modulations in the temporal envelope.

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 content-based search in sound databases (McAdams & Misdariis, 1999; Peeters, McAdams, & Herrera, 2000). Such representations are only useful to the extent that they are (a) generalizable beyond the set of sounds actually studied, (b) robust with respect to changes in musical context, and (c) generalizable to other kinds of listening tasks than those used to construct the model. To the degree that a representation has these properties, it may be considered as an accurate account of musical timbre, characterized by an important feature of a scientific model, the ability to predict new empirical phenomena.

C. *Interaction of Timbre with Pitch and Dynamics*

Most timbre space studies have restricted the pitch and loudness to single values for all of the instrument sounds compared in order to focus listeners' attention on timbre alone. An important question arises, however, concerning whether the timbral relations revealed for a single pitch and/or a single dynamic level hold at different pitches and dynamic levels and, more importantly for extending this work to real musical contexts, whether they hold for timbres being compared *across* pitches and dynamic levels.

It is clear that for many instruments the timbre varies as a function of pitch because the spectral, temporal, and spectrotemporal properties of the sounds covary with pitch. Marozeau, de Cheveigné, McAdams, and Winsberg (2003) have shown that timbre spaces for recorded musical instrument tones are similar at different pitches (B_3, $C\sharp_4$, $B\flat_4$). Listeners are also able to ignore pitch differences within an octave when asked to compare only the timbres of the tones. When the pitch variation is greater than an octave, interactions between the two attributes occur. Marozeau and de Cheveigné (2007) varied the brightness of a set of synthesized sounds, while also varying the pitch over a range of 18 semitones. They found that differences in pitch affected timbre relations in two ways: (1) pitch shows up in the timbre space representation as a dimension orthogonal to the timbre dimensions (indicating simply that listeners were no longer ignoring the pitch difference), and (2) pitch differences systematically affect the timbre dimension related to spectral centroid. Handel and Erickson (2004) also found that listeners had difficulty

extrapolating the timbre of a sound source across large differences in pitch. Inversely, Vurma, Raju, and Kuuda (2011) have reported that timbre differences on two tones for which the in-tuneness of the pitches was to be judged affected the pitch judgments to an extent that could potentially lead to conflicts between subjective and fundamental-frequency-based assessments of tuning. Krumhansl and Iverson (1992) found that speeded classifications of pitches and of timbres were symmetrically affected by uncorrelated variation along the other parameter. These results suggest a close relation between timbral brightness and pitch height and perhaps even more temporally fine-grained features related to the coding of periodicity in the auditory system or larger-scale timbral properties related to the energy envelope. This link would be consistent with underlying neural representations that share common attributes, such as tonotopic and periodicity organizations in the brain.

Similarly to pitch, changes in dynamics also produce changes in timbre for a given instrument, particularly, but not exclusively, as concerns spectral properties. Sounds produced with greater playing effort (e.g., fortissimo vs. pianissimo) not only have greater energy at the frequencies present in the softer sound, but the spectrum spreads toward higher frequencies, creating a higher spectral centroid, a greater spectral spread, and a lower spectral slope. No studies to date of which we are aware have examined the effect of change in dynamic level on timbre perception, but some work has looked at the role of timbre in the perception of dynamic level independently of the physical level of the signal. Fabiani and Friberg (2011) studied the effect of variations in pitch, sound level, and instrumental timbre (clarinet, flute, piano, trumpet, and violin) on the perception of the dynamics of isolated instrumental tones produced at different pitches and dynamics. They subsequently presented these sounds to listeners at different physical levels. Listeners were asked to indicate the perceived dynamics of each stimulus on a scale from pianissimo to fortissimo. The results showed that the timbral effects produced at different dynamics, as well as the physical level, had equally large effects for all five instruments, whereas pitch was relevant mostly for clarinet, flute, and piano. Thus estimates of the dynamics of musical tones are based both on loudness and timbre, and to a lesser degree on pitch as well.

II. Timbre as a Vehicle for Source Identity

The second approach to timbre concerns its role in the recognition of the identity of a musical instrument or, in general, of a sound-generating event, that is, the interaction between objects, or a moving medium (air) and an object, that sets up vibrations in the object or a cavity enclosed by the object. One reasonable hypothesis is that the sensory dimensions that compose timbre serve as indicators used in the categorization, recognition, and identification of sound events and sound sources (Handel, 1995; McAdams, 1993).

Research on musical instrument identification is relevant to this issue. Saldanha and Corso (1964) studied identification of isolated musical instrument sounds from

the Western orchestra played with and without vibrato. They were interested in the relative importance of onset and offset transients, spectral envelope of the sustain portion of the sound, and vibrato. Identification of isolated sounds is surprisingly poor for some instruments. When attacks and decays were excised, identification decreased markedly for some instruments, particularly for the attack portion in sounds without vibrato. However when vibrato was present, the effect of cutting the attack was less, identification being better. These results suggest that important information for instrument identification is present in the attack portion, but that in the absence of the normal attack, additional information is still available in the sustain portion, particularly when vibrato is present (although it is more important for some instruments than others). The vibrato may increase our ability to extract information relative to the resonance structure of the instrument (McAdams & Rodet, 1988).

Giordano and McAdams (2010) performed a meta-analysis on previously published data concerning identification rates and dissimilarity ratings of musical instrument tones. The goal of this study was to ascertain the extent to which tones generated with large differences in the mechanisms for sound production were recovered in the perceptual data. Across all identification studies, listeners frequently confused tones generated by musical instruments with a similar physical structure (e.g., clarinets and saxophones, both single-reed instruments) and seldom confused tones generated by very different physical systems (e.g., the trumpet, a lip-valve instrument, and the bassoon, a double-reed instrument). Consistently, the vast majority of previously published timbre spaces revealed that tones generated with similar resonating structures (e.g., string instruments vs. wind instruments) or with similar excitation mechanisms (e.g., impulsive excitation as in piano tones vs. sustained excitation as in flute tones) occupied the same region in the space. These results suggest that listeners can reliably identify large differences in the mechanisms of tone production, focusing on the timbre attributes used to evaluate the dissimilarities among musical sounds.

Several investigations on the perception of everyday sounds extend the concept of timbre beyond the musical context (see McAdams, 1993; Handel, 1995; Lutfi, 2008, for reviews). Among them, studies on impact sounds provide information on the timbre attributes useful to the perception of the properties of percussion instruments: bar geometry (Lakatos, McAdams & Caussé, 1997), bar material (McAdams, Chaigne, & Roussarie, 2004), plate material (Giordano & McAdams, 2006; McAdams, Roussarie, Chaigne, & Giordano, 2010), and mallet hardness (Freed, 1990; Giordano, Rocchesso, & McAdams, 2010). The timbral factors relevant to perceptual judgments vary with the task at hand. Spectral factors are primary for the perception of geometry (Lakatos et al., 1997). Spectrotemporal factors (e.g., the rate of change of spectral centroid and loudness) dominate the perception of the material of struck objects (McAdams et al., 2004; Giordano & McAdams, 2006) and of mallets (Freed, 1990). But spectral and temporal factors can also play a role in the perception of different kinds of gestures used to set an instrument into vibration, such as the angle and position of a plucking finger on a guitar string (Traube, Depalle & Wanderley, 2003).

The perception of an instrument's identity in spite of variations in pitch may be related to timbral invariance, those aspects of timbre that remain constant with change in pitch and loudness. Handel and Erickson (2001) found that musically untrained listeners are able to recognize two sounds produced at different pitches as coming from the same instrument or voice only within a pitch range of about an octave. Steele and Williams (2006) found that musically trained listeners could perform this task at about 80% correct even with pitch differences on the order of 2.5 octaves. Taken together, these results suggest that there are limits to timbral invariance across pitch, but that they depend on musical training.

Its role in source identification and categorization is perhaps the more neglected aspect of timbre and brings with it advantages and disadvantages for the use of timbre as a form-bearing dimension in music (McAdams, 1989). One of the advantages is that categorization and identification of a sound source may bring into play perceptual knowledge (acquired by listeners implicitly through experience in the everyday world and in musical situations) that helps them track a given voice or instrument in a complex musical texture. Listeners do this easily and some research has shown that timbral factors may make an important contribution in such voice tracking (Culling & Darwin, 1993; Gregory, 1994), which is particularly important in polyphonic settings.

The disadvantages may arise in situations in which the composer seeks to create melodies across instrumental timbres, e.g., the *Klangfarbenmelodien* of Schoenberg (1911/1978). Our predisposition to identify the sound source and follow it through time would impede a more relative perception in which the timbral differences were perceived as a movement through timbre space rather than as a simple change of sound source. For cases in which such timbral compositions work, the composers have often taken special precautions to create a musical situation that draws the listener more into a relative than into an absolute mode of perceiving.

III. Timbre as a Structuring Force in Music Perception

Timbre perception is at the heart of orchestration, a realm of musical practice that has received relatively little experimental study or even music-theoretic treatment for that matter. Instrumental combinations can give rise to new timbres if the sounds are perceived as blended. Timbral differences can also both create the auditory streaming of similar timbres and the segregation of dissimilar timbres, as well as induce segmentations of sequences when timbral discontinuities occur. Listeners can perceive intervals between timbres as similar when they are transposed to a different part of timbre space, even though such relations have not been used explicitly in music composition. Timbre can play a role in creating and releasing musical tension. And finally, there is some evidence that listeners can learn statistical regularities in timbre sequences, opening up the possibility of developing timbre-based grammars in music.

A. Timbral Blend

The creation of new timbres through orchestration necessarily depends on the degree to which the constituent sound sources fuse together or blend to create the newly emergent sound (Brant, 1971; Erickson, 1975). Sandell (1995) has proposed that there are three classes of perceptual goals in combining instruments: *timbral heterogeneity* in which one seeks to keep the instruments perceptually distinct, *timbral augmentation* in which one instrument embellishes another one that perceptually dominates the combination, and *timbral emergence* in which a new sound results that is identified as none of its constituents. Blend appears to depend on a number of acoustic factors such as onset synchrony of the constituent sounds and others that are more directly related to timbre, such as the similarity of the attacks, the difference in the spectral centroids, and the overall centroid of the combination. For instance, Sandell (1989) found that by submitting blend ratings taken as a measure of proximity to multidimensional scaling, a "blend space" could be obtained; the dimensions of this space were correlated with attack time and spectral centroid, suggesting that the more these parameters were similar for the two combined sounds, the greater their blend (Figure 8). A similar trend concerning the role of spectrotemporal similarity in blend was found for wind instrument combinations by Kendall and Carterette (1993). These authors also revealed an inverse relation between blend and identifiability of the constituent sounds, i.e., sounds that blend

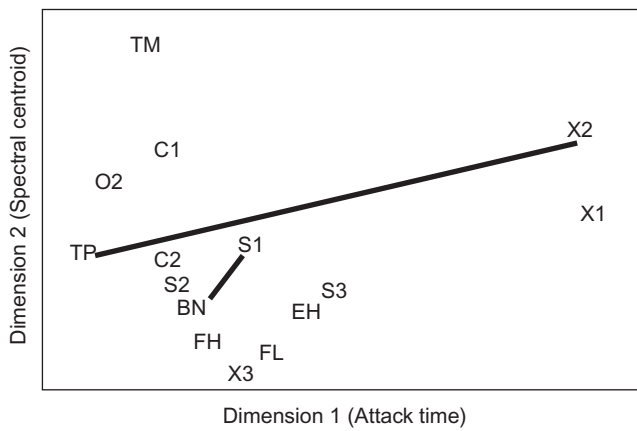

Figure 8 Multidimensional analysis of blend ratings for all pairs of sounds drawn from the timbre space of Grey (1977). If two instruments are close in the space (e.g., BN and S1), the degree of blend is rated as being strong. If they are far apart (e.g., TP and X2), the blending is weak and the sounds tend to be heard separately. The dimensions of this "blend space" are moderately correlated with the attack time (*x* axis) and strongly correlated with spectral centroid (*y* axis). (TM = muted trombone, C1-C2 = clarinets, O1-O2 = oboes, TP = trumpet, BN = bassoon, FH = French horn, FL = flute, S1-S3 = strings, X1-X3 = saxophones, EH = English horn).

better are more difficult to identify separately in the mixture. For dyads of impulsive and continuant sounds, the blend is greater for slower attacks and lower spectral centroids and the resulting emergent timbre is determined primarily by the properties of the impulsive sound (Tardieu & McAdams, in press).

B. *Timbre and Musical Grouping*

An important way in which timbre can contribute to the organization of musical structure is related to the fact that listeners tend 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 Bregman, 1990, Chapter 2; McAdams & Bregman, 1979, 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 & 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 (i.e., in their pitches and timbres) 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. Early demonstrations (see Figure 9) of auditory streaming on the basis of timbre suggest a link between the timbre-space representation and the tendency for auditory streaming on the basis of the spectral differences that are created (McAdams & Bregman, 1979; Wessel, 1979).

Hartmann and Johnson's (1991) experimental results convinced them 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. More recently the picture has changed significantly, and several studies indicate an important role for both spectral and temporal attributes of

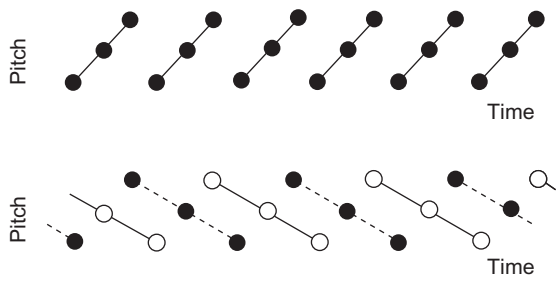

Figure 9 The two versions of a melody created by David Wessel with one instrument (top) or two alternating instruments (bottom). In the upper single-timbre melody, a single rising triplet pattern is perceived. In the lower alternating-timbre melody, if the timbral difference is sufficient, two interleaved patterns of descending triplets at half the tempo of the original sequence are heard.

timbre in auditory stream segregation (Moore & Gockel, 2002). 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. Multidimensional scaling of the segregation judgments treated as a measure of dissimilarity 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 & 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.

This result was refined in an experiment by Singh and Bregman (1997) in which amplitude envelope and spectral content were independently varied and their relative contributions to stream segregation were measured. For the parameters used, a change from two to four harmonics produced a greater effect on segregation than did a change from a 5-ms attack and a 95-ms decay to a 95-ms attack and a 5-ms decay. Combining the two gave no greater segregation than was obtained with the spectral change, suggesting a stronger contribution of this sound property to segregation.

Bey and McAdams (2003) used a melody discrimination paradigm in which a target melody interleaved with a distractor melody was presented first, followed by a test melody that was either identical to the target or differed by two notes that changed the contour (Figure 10). The timbre difference between target and distractor melodies was varied within the timbre space of McAdams et al. (1995).

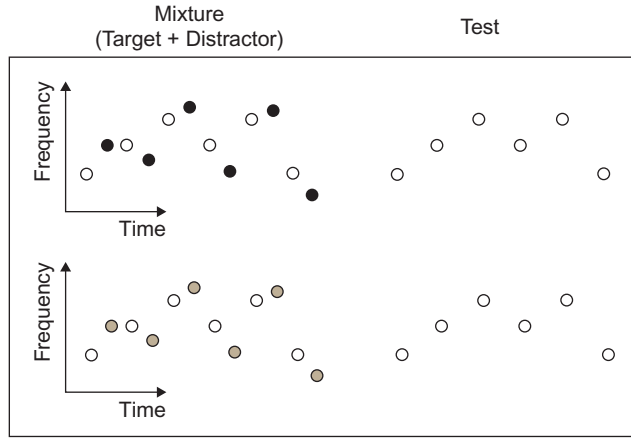

Figure 10 Sequences used for testing the role of timbre in stream segregation. The task was to determine whether the isolated test melody had been present in the mixture of the target melody (empty circles) and an interleaved distractor melody (filled circles, with the darkness indicating degree of timbre difference between distractor and target). The test and target melodies always had the same timbre.
Redrawn from Figure 2, Bey and McAdams (2003). ©2003 by The American Psychological Association, Inc. Adapted with permission.

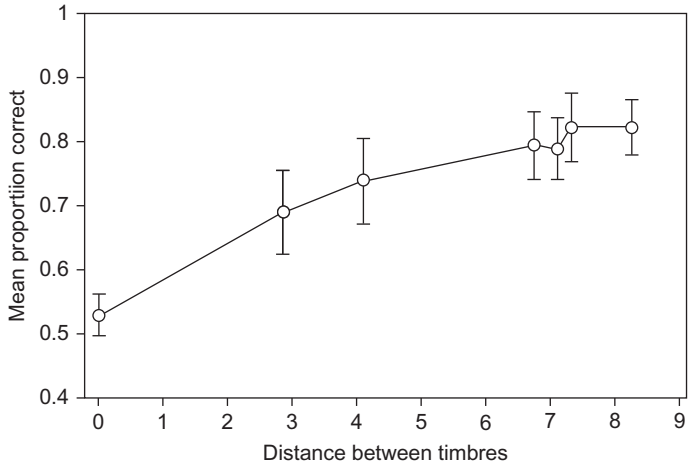

Figure 11 A monotone relation between the timbral distance and the rate of discrimination between target and test melodies shows that distance in timbre space predicts stream segregation.
Redrawn from Figure 4, Bey and McAdams (2003). ©2003 by The American Psychological Association, Inc. Adapted with permission.

In line with the previously cited results, melody discrimination increased monotonically with the distance between the target and distractor timbres, which varied along the dimensions of attack time, spectral centroid, and spectral flux (Figure 11).

All of these results are important for auditory stream segregation theory, because they show that several of a source's acoustic properties are taken into account when forming auditory streams. They are also important for music making (whether it be with electroacoustic or acoustic instruments), because they show that many aspects of timbre strongly affect the basic organization of the musical surface into streams. Different orchestrations of a given pitch sequence can completely change what is heard as melody and rhythm, as has been demonstrated by Wessel (1979). Timbre is also an important component in the perception of musical groupings, whether they are at the level of sequences of notes being set off by sudden changes in timbre (Deliège, 1987) or of larger-scale musical sections delimited by marked changes in orchestration and timbral texture (Deliège, 1989).

C. *Timbral Intervals*

Consider the timbral trajectory shown in Figure 12 through the McAdams et al. (1995) timbre space starting with the *guitarnet* (gtn) and ending with the English horn (ehn). How would one construct a melody starting from the bowed string (stg) so that it would be perceived as a transposition of this *Klangfarbenmelodie*? The notion of transposing the relation between two timbres to another point in the timbre space poses the question of whether listeners can indeed perceive timbral

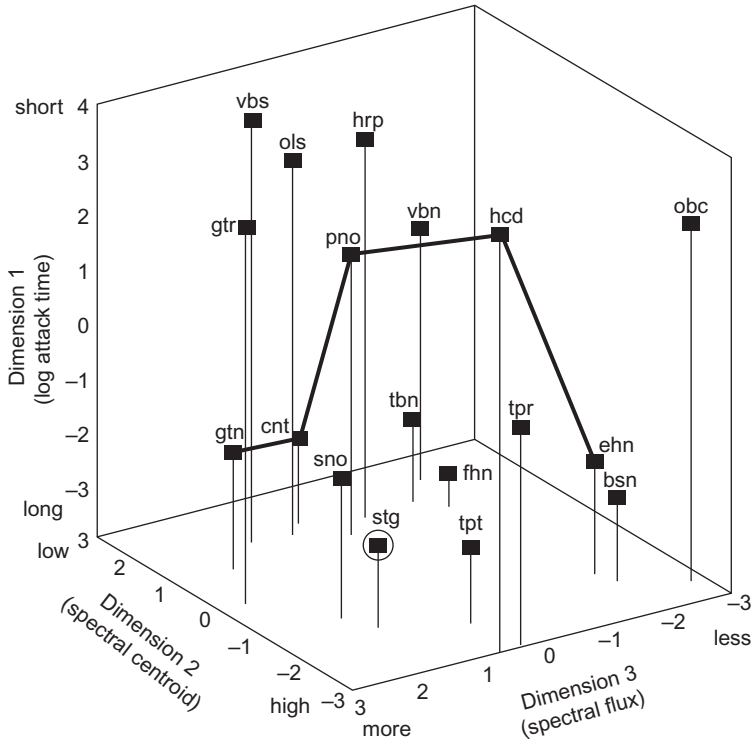

Figure 12 A trajectory of a short timbre melody through timbre space. How would one transpose the timbre melody starting on gtn to one starting on stg?

intervals. If timbral interval perception can be demonstrated, it opens the door to applying some of the operations commonly used on pitch sequences to timbre sequences (Slawson, 1985). Another interest of this exploration is that it extends the use of the timbre space as a perceptual model beyond the dissimilarity paradigm.

Ehresman and Wessel (1978) took a first step forward in this direction. Based on previous work on semantic spaces and analogical reasoning (Henley, 1969; Rumelhart & 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 timbral 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." But what does transposition mean in a multidimensional space? A timbral interval can be considered 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

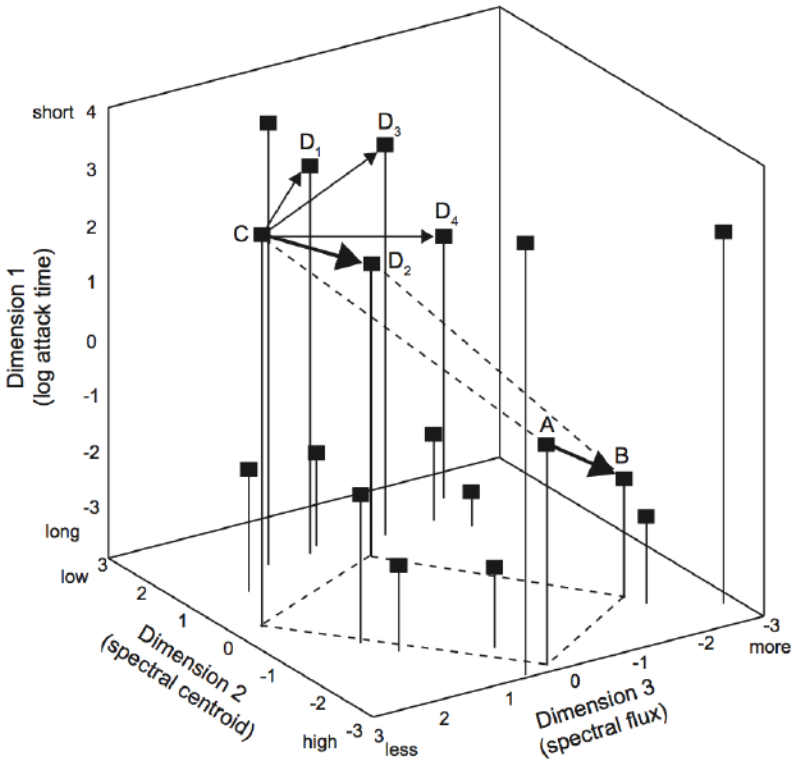

Figure 13 Examples of timbral intervals in a timbre space. The aim is to find an interval starting with C and ending on a timbre D that resembles the interval between timbres A and B. If we present timbres D_1–D_4 (in a manner similar to that of Ehresman & Wessel, 1978), the vector model would predict that listeners would prefer D_2, because the vector CD_2 is the closest in length and orientation to that of AB.

and linear from a perceptual point of view, then pairs of timbres characterized by the same vector relation should have the same perceptual relation and thus embody the same timbral 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 (1978) tested this hypothesis using a task in which listeners had to compare two timbral intervals (e.g., A-B vs. C-D) and rank various timbre D's according to how well they fulfilled the analogy: timbre A is to timbre B as timbre C is to timbre D (see Figure 13). They essentially found that the closer timbre D was to the ideal point defined by the vector model in timbre space, the higher the ranking, i.e., the ideal C-D vector was a simple translation of the A-B vector and A, B, C and D form a parallelogram (shown with dashed lines in Figure 13).

McAdams and Cunibile (1992) subsequently tested the vector model using the 3D space from Krumhansl (1989) (ignoring the specificities). 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 have had almost no experience with music that uses timbral 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 electroacoustic 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.

One confounding factor is that the specificities on some timbres in this set were ignored. These specificities would necessarily distort the vectors that were used to choose the timbres, because they are like an additional dimension for each timbre. As such, certain timbral intervals correspond well to what is predicted because specificities are absent or low in value, whereas others are seriously distorted and thus not perceived as similar to other intervals due to moderate or high specificity values. What this line of reasoning suggests is that the use of timbral intervals as an integral part of a musical discourse runs the risk of being very difficult to achieve with very complex and idiosyncratic sound sources, because they will in all probability have specificities of some kind or another. The use of timbral intervals may, in the long run, be limited to synthesized sounds or blended sounds created through the combination of several instruments.

D. Building and Releasing Musical Tension with Timbre

Timbre can also contribute 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 (1983) generative theory of tonal music. 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, & 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 (cf. Parncutt, 1989; Plomp, 1976, for reviews).

As a first step toward understanding how this operates in music, Paraskeva and McAdams (1997) measured the inflection of musical tension and relaxation due to timbral change. Listeners were asked to make judgments on a seven-point scale concerning the perceived degree of completion of the music at several points at

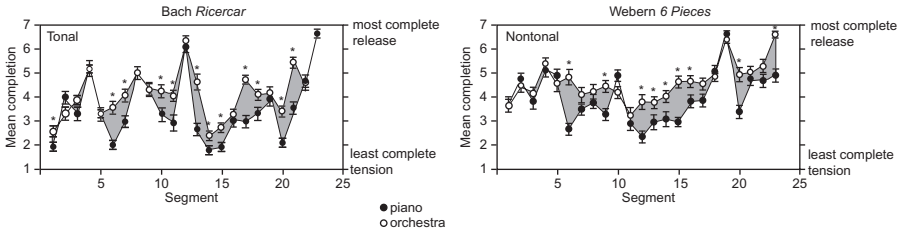

Figure 14 Rated degree of completion at different stopping points (segments) for works by Bach and Webern, averaged over musician and nonmusician groups. The filled circles correspond to the piano version and the open circles to the orchestral version. The vertical bars represent the standard deviation. The asterisks over certain segments indicate a statistical difference between the two versions for that stopping point.
Redrawn from Figure 1 in Paraskeva and McAdams (1997). ©1997 by the authors. Adapted with permission.

which the music stopped. What results is a completion profile (Figure 14), which can be used to infer musical tension by equating completion with release and lack of completion with tension. Two pieces were tested: a fragment of the *Ricercar* from the *Musical Offering* for six voices by Bach (tonal) and the first movement of the *Six Pieces for Orchestra, Op. 6* by Webern (nontonal). Each piece was played in an orchestral version (Webern's orchestration of the *Musical Offering* was used for the Bach) and in a direct transcription of this orchestral version for piano on a digital sampler. Although there were only small differences between the profiles for musicians and nonmusicians, there were significant differences between the piano and orchestral versions, indicating a significant effect of timbre change on perceived musical tension. However, when they *were* significantly different, the orchestral version was always more relaxed than the piano version.

The hypothesis advanced by Paraskeva and McAdams (1997) for this effect was that the higher relaxation of the orchestral version might have been due to processes involved in auditory stream formation and the dependence of perceived roughness on the results of such processes (Wright & Bregman, 1987). Roughness, or any other auditory attribute of a single sound event, is computed after auditory organization processes have grouped the bits of acoustic information 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 quite synchronous. Because they all start at the same time and have similar amplitude envelopes and similar timbres, they will tend to be fused together. The computed roughness will then 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, whereas others have faster attacks. There could then be greater asynchrony between the instruments in terms of perceived attack time (Gordon, 1987). In addition, because 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 in separate auditory streams. 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 those of the piano version. So once again, timbral composition can have a very tight interaction with auditory scene analysis processes.

E. Implicit Learning of Timbre-Based Grammars

In order to use timbre syntactically in music, listeners would need to be able to learn rules for ordering timbres in sequences, as for duration and pitch. This possibility was first explored by Bigand, Perruchet, and Boyer (1998), who presented artificial grammars of musical sounds for which sequencing rules were created. After being exposed to sequences constructed with the grammar, listeners heard new sequences and had to decide whether each one conformed or not to the learned grammar, without having to say why. Indeed, with the implicit learning of the structures of language and music, we can know whether a sequence corresponds to our "language" without knowing why: it just doesn't sound right. The correct response rate was above chance for these sequences, demonstrating the listeners' ability to learn a timbral grammar.

Tillmann and McAdams (2004) extended this work by studying the influence of acoustic properties on implicit learning of statistical regularities (transition probabilities between temporally adjacent events) in sequences of musical sounds differing only in timbre. These regularities formed triplets of timbres drawn from the timbre space of McAdams et al. (1995). The transition probability between the first and second and between the second and third timbres was much higher than that between the third timbre of a given triplet and the first timbre of any other triplet in the "language" used in their experiment. In the implicit learning phase, listeners heard a rhythmically regular sequence of timbres, all at the same pitch and loudness, for 33 minutes. The sequence was composed of all of the triplets in the "language" in a varied sequence. The goal was to determine whether listeners could learn the regularities that defined the triplets by simply listening to the sequences for a fairly short time.

In addition to the principle of higher transition probability between timbres within the triplets than between those in different triplets, the sequences were also constructed so that the auditory grouping on the basis of timbral similarity was either congruent with the triplet structure or not (Figure 15). To achieve this, three grammars were created. For the congruent sequence (S1), the timbres within each triplet were fairly close within the McAdams et al. (1995) timbre space, and the distance between the last timbre of one triplet and the first timbre of the succeeding triplet was large. If the timbral discontinuities created by the jumps in timbre space between triplets created a segmentation of the sequence, this segmentation would

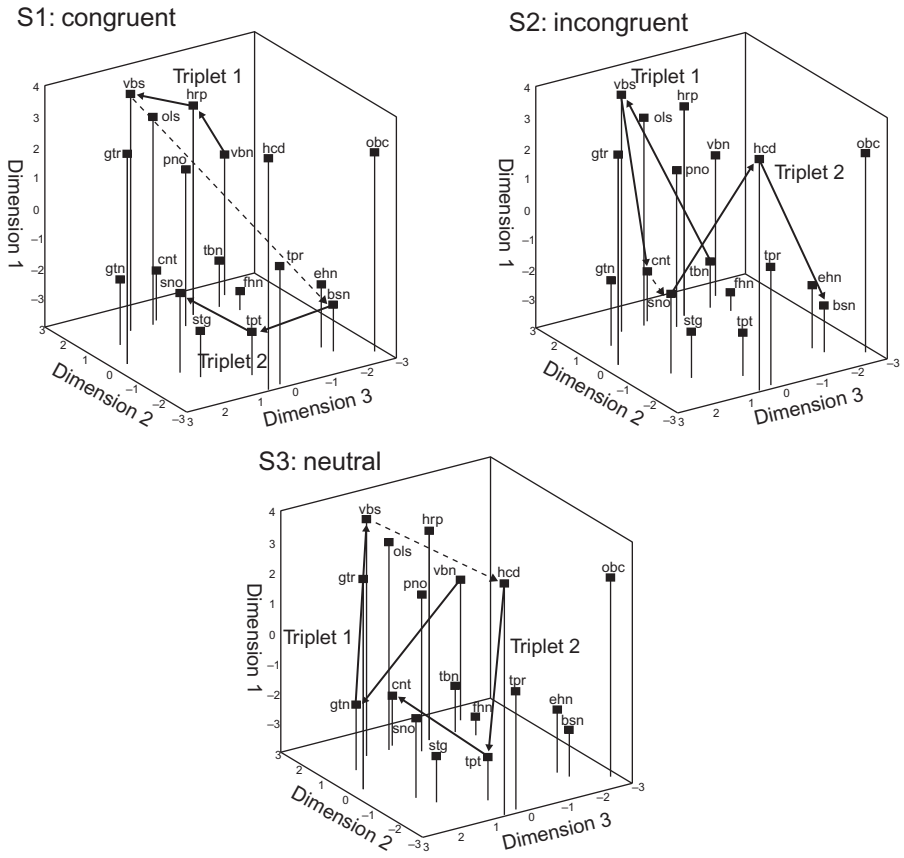

Figure 15 Examples of timbre triplets used in the three timbral grammars drawn from the McAdams et al. (1995) timbre space. In S1 (congruent), the segmentation of the sequence into groups of timbres that are close in the space corresponded to the triplets of the grammar defined in terms of transition probabilities. In S2 (incongruent), the segmentation groups the last timbre of a triplet with the first of the next triplet, isolating the middle timbre of each triplet. In S3 (neutral), all timbres are more or less equidistant, thereby not creating segmentation.

correspond to the triplets themselves. For the incongruent sequence (S2), there was a large distance between successive timbres within the triplets and a small distance from one triplet to the next. Accordingly, sequential grouping processes would create segmentations into two timbres traversing adjacent triplets and an isolated timbre in the middle of each triplet. Finally, a third sequence (S3) was composed so that all of the distances within and between triplets were uniformly medium within the McAdams et al. (1995) space, thus avoiding segmentation.

After listening to one of the three sequences for 33 minutes, two groups of three timbres were presented, and the listener had to decide which one formed a triplet that was present in the sequence just heard. Another group of listeners did not hear the 33-minute sequence beforehand and had to decide which of the two groups

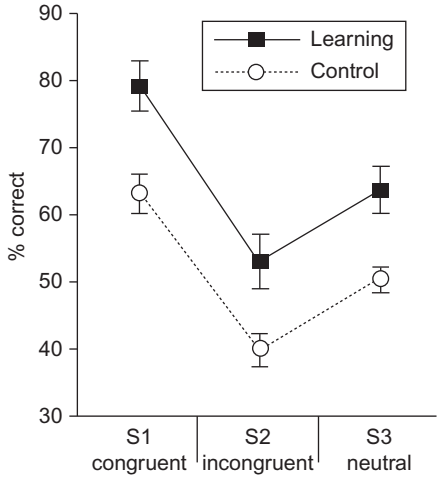

Figure 16 Percent correct choice of triplets of the constructed grammar for sequences in which the perceptual segmentation was congruent, incongruent, or neutral with respect to the triplets of the grammar. The "control" group did not hear the learning sequence before the test session. The "learning" group was exposed to the grammar for 33 minutes before the test session.
Redrawn from Figure 1, Tillmann and McAdams (2004). ©2004 by The American Psychological Association, Inc. Adapted with permission.

of three timbres best formed a unit that could be part of a longer sequence of timbres. Choices of a triplet that were part of the grammar were scored as correct.

Listeners were able to learn the grammar implicitly by simply listening to it, because the correct response rates of the learning group were higher than those of the group who were not exposed to the sequences beforehand (Figure 16). But curiously, this learning did not depend on the congruence between the grouping structure created by the acoustic discontinuities and the structure created by the statistical regularities determined by the transition probabilities between timbres within and between triplets. The same increase in correct response rate was obtained for all three sequences. This result suggests that the choice was affected by the grouping structure—listeners prefer the "well-formed" triplets—but the degree of statistical learning that occurred while listening to the sequences was the same in all conditions. The listeners thus seem to be able to learn the grammar constructed by the timbre-sequencing rule, whether the timbre sequences of the grammar are composed of similar or dissimilar timbres. Nevertheless, listeners prefer an organization in motifs compose of timbres that are close in timbre space and distant in timbre from other motifs.

IV. Concluding Remarks

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. This perceptual structure is represented in a timbre space, a powerful psychological model that allows predictions to be made about timbre perception in situations both within and beyond those used to derive the model from dissimilarity ratings. Timbral 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.

The formalization of audio descriptors to capture quantitatively the acoustic properties that give rise to many aspects of timbre perception is beginning to provide an important set of tools that benefits several domains, including the use of signal-based metadata related to timbre that can be used in automatic instrument recognition and categorization (Eronen & Klapuri, 2000; Fujinaga & MacMillan, 2000), content-based searches in very large sound and music databases (Kobayashi & Osaka, 2008), characterization of sound and music samples in standards such as MPEG (Peeters et al., 2000), and many other music information retrieval and musical machine learning applications. These descriptors, particularly the time-varying ones, are proving to be useful in computer-aided orchestration environments (Carpentier, Tardieu, Harvey, Assayag, & Saint-James, 2010; Esling, Carpentier, & Agon, 2010; Rose & Hetrick, 2007), in which the research challenge is to predict the perceptual results of instrumental combinations and sequencings to fit a goal expressed by a composer, arranger, or sound designer.

Timbre can also play a role in phrase-level variations that contribute to musical expression. Measurements of timbral variation in phrasing on the clarinet demonstrate that players control spectral and temporal properties as part of their arsenal of expressive devices. Further, mimicking instrumental variations of timbre in synthesized sound sequences increases listeners' preferences compared to sequences lacking such variation (Barthet, Kronland-Martinet & Ystad, 2007). And in the realm of computer sound synthesis, there is increasing interest in continuous control of timbral attributes to enhance musical expression (Lee & Wessel, 1992; Momeni & Wessel, 2003).

Larger-scale changes in timbre can also contribute to the expression of higher-level structural functions in music. Under conditions of high blend among instruments composing a vertical sonority, timbral roughness is a major component of musical tension. However, it strongly depends on the way auditory grouping processes have parsed the incoming acoustic information into events and streams. Orchestration can play a major role in addition to pitch and rhythmic patterns in the structuring of musical tension and relaxation schemas that are an important component of the æsthetic response to musical form. In the realm of electroacoustic music and in some orchestral music, timbre plays a primary grammatical role. This is particularly true in cases in which orchestration is an integral part of the compositional process, what the composer John Rea calls *prima facie orchestration*, rather than being a level of expression that is added after the primary structuring forces of pitch and duration have been determined, what Rea calls *normative orchestration*. In such cases, the structuring and sculpting of timbral changes and relations among complex auditory events provide a universe of possibilities that composers have been exploring for decades (cf. Risset, 2004), but which musicologists have only

recently begun to address (Nattiez, 2007; Roy, 2003) and psychologists have yet to tackle with any scope or in any depth.

Nattiez (2007) in particular has taken Meyer's (1989) distinction between primary and secondary musical parameters and questioned his relegating of timbre to secondary status. In Meyer's conception, primary parameters such as pitch and duration[3] are able to carry syntax. Syntactic relations for Meyer are based on expectations that are resolved in closure, that is, on implications and realizations. Secondary parameters, on the other hand, are not organized in discrete units or clearly recognizable categories. According to Snyder (2000), we hear secondary parameters (among which he also includes timbre) simply in terms of their relative amounts, which are useful more for musical expression and nuance than for building grammatical structures. However, Nattiez (2007) notes that, according to his own analyses of instrumental music and those of Roy (2003) in electroacoustic music, timbre can be used to create syntactic relations that depend on expectations leading to a perception of closure. As such, the main limit of Meyer's conclusion concerning timbre was that he confined his analyses to works composed in terms of pitch and rhythm and in which timbre was in effect allowed to play only a secondary functional role. This recalls Rea's distinction between prima facie and normative orchestration mentioned previously. It suffices to cite the music of electroacoustic composers such as Dennis Smalley, orchestral music by György Ligeti or mixed music by Trevor Wishart to understand the possibilities. But even in the orchestral music of Beethoven in the high Classical period, timbre plays a structuring role at the level of sectional segmentation induced by changes in instrumentation and at the level of distinguishing individual voices or orchestral layers composed of similar timbres.

As a factor responsible for structuring tension and release, timbre has been used effectively by electroacoustic composers such as Francis Dhomont and Jean-Claude Risset. According to Roy's (2003) analyses, Dhomont's music, for example, uses timbre to build expectancies and deceptions in a musical context that isn't "contaminated" by strong pitch structures. Underlying this last remark is the implication that in a context in which pitch is a structuring force, timbre may have a hard time imposing itself as a dominant parameter, suggesting a sort of dominance hierarchy favoring rhythm and pitch when several parameters are brought into play. Research on conditions in which the different musical parameters can act in the presence of others in the perceptual structuring of music are not legion and rarely go beyond the royal couple of pitch and rhythm (see the discussion in McAdams, 1989).[4] The terrain for exploring interactions among musical parameters, and thus situating their potential relative roles in bearing musical forms, will necessitate a joint effort involving musicological analysis and psychological experimentation, but it is potentially vast, rich, and very exciting.

[3] He probably really meant interonset intervals, because note duration itself is probably a secondary parameter related to articulation.

[4] One exception is work by Krumhansl and Iverson (1992) showing that in the perception of sequences, there is an asymmetry in the relation between pitch and timbre such that pitch seems to be perceived more in relative terms and timbre in absolute terms.

Acknowledgments

The preparation of this chapter was supported by the Natural Sciences and Engineering Research Council and the Social Sciences and Humanities Research Council of Canada and the Canada Research Chairs program.

References

Barthet, M., Kronland-Martinet, R., & Ystad, S. (2007). Improving musical expressiveness by time-varying brightness shaping. In R. Kronland-Martinet, S. Ystad, & K. Jensen (Eds.), *Computer music modeling and retrieval: Sense of sounds* (pp. 313−336). Berlin, Germany: Springer.

Bey, C., & McAdams, S. (2003). Post-recognition of interleaved melodies as an indirect measure of auditory stream formation. *Journal of Experimental Psychology: Human Perception and Performance*, *29*, 267−279.

Bigand, E., Parncutt, R., & Lerdahl, F. (1996). Perception of musical tension in short chord sequences: The influence of harmonic function, sensory dissonance, horizontal motion, and musical training. *Perception & Psychophysics*, *58*, 125−141.

Bigand, E., Perruchet, P., & Boyer, M. (1998). Implicit learning of an artificial grammar of musical timbres. *Cahiers de Psychologie Cognitive*, *17*, 577−600.

Brant, H. (1971). Orchestration. In J. Vinton (Ed.), *Dictionary of contemporary music* (pp. 538−546). New York, NY: E. P. Dutton.

Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound*. Cambridge, MA: MIT Press.

Bregman, A. S., & Campbell, J. (1971). Primary auditory stream segregation and perception of order in rapid sequences of tones. *Journal of Experimental Psychology*, *89*, 244−249.

Caclin, A., McAdams, S., Smith, B. K., & Winsberg, S. (2005). Acoustic correlates of timbre space dimensions: A confirmatory study using synthetic tones. *Journal of the Acoustical Society of America*, *118*, 471−482.

Carpentier, G., Tardieu, D., Harvey, J., Assayag, G., & Saint-James, E. (2010). Predicting timbre features of instrument sound combinations: Application to automatic orchestration. *Journal of New Music Research*, *39*, 47−61.

Carroll, D., & Chang, J. (1970). Analysis of individual differences in multidimensional scaling via an N-way generalization of Eckart-Young decomposition. *Psychometrika*, *35*, 283−319.

Culling, J. F., & Darwin, C. J. (1993). The role of timbre in the segregation of simultaneous voices with intersecting F_0 contours. *Perception & Psychophysics*, *34*, 303−309.

Deliège, I. (1987). Grouping conditions in listening to music: An approach to Lerdahl & Jackendoff's grouping preference rules. *Music Perception*, *4*, 325−360.

Deliège, I. (1989). A perceptual approach to contemporary musical forms. *Contemporary Music Review*, *4*, 213−230.

Ehresman, D., & Wessel, D. L. (1978). Perception of timbral analogies, *Rapports de l'IRCAM* (Vol. 13). Paris, France: IRCAM-Centre Pompidou.

Erickson, R. (1975). *Sound structure in music*. Berkeley, CA: University of California Press.

Eronen, A., & Klapuri, A. (2000). Musical instrument recognition using cepstral coefficients and temporal features. *Proceedings of the 2000 IEEE International Conference on Acoustics, Speech, and Signal Processing, Istanbul*, *2*, II753−II756.

Esling, P., Carpentier, G., & Agon, C. (2010). Dynamic musical orchestration using genetic algorithms and a spectrotemporal description of musical instruments. In C. Di Chio, et al. (Eds.), *Applications of evolutionary computation, LNCS 6025* (pp. 371−380). Berlin, Germany: Springer-Verlag.

Fabiani, M., & Friberg, A. (2011). Influence of pitch, loudness, and timbre on the perception of instrument dynamics. *Journal of the Acoustical Society of America*, *130*, EL193−EL199.

Freed, D. J. (1990). Auditory correlates of perceived mallet hardness for a set of recorded percussive events. *Journal of the Acoustical Society of America*, *87*, 1236−1249.

Fujinaga, I., & MacMillan, K. (2000). Realtime recognition of orchestral instruments. *Proceedings of the International Computer Music Conference, Berlin* (pp. 141−143). San Francisco, CA: International Computer Music Association.

Giordano, B. L., & McAdams, S. (2006). Material identification of real impact sounds: Effects of size variation in steel, glass, wood and plexiglass plates. *Journal of the Acoustical Society of America*, *119*, 1171−1181.

Giordano, B. L., & McAdams, S. (2010). Sound source mechanics and musical timbre perception: Evidence from previous studies. *Music Perception*, *28*, 155−168.

Giordano, B. L., Rocchesso, D., & McAdams, S. (2010). Integration of acoustical information in the perception of impacted sound sources: The role of information accuracy and exploitability. *Journal of Experimental Psychology: Human Perception and Performance*, *36*, 462−476.

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

Gregory, A. H. (1994). Timbre and auditory streaming. *Music Perception*, *12*, 161−174.

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.

Hajda, J. M., Kendall, R. A., Carterette, E. C., & Harshberger, M. L. (1997). Methodological issues in timbre research. In I. Deliège, & J. Sloboda (Eds.), *Perception and cognition of music* (pp. 253−306). Hove, U.K.: Psychology Press.

Handel, S. (1995). Timbre perception and auditory object identification. In B. C. J. Moore (Ed.), *Hearing* (pp. 425−462). San Diego, CA: Academic Press.

Handel, S., & Erickson, M. (2001). A rule of thumb: The bandwidth for timbre invariance is one octave. *Music Perception*, *19*, 121−126.

Handel, S., & Erickson, M. (2004). Sound source identification: The possible role of timbre transformations. *Music Perception*, *21*, 587−610.

Hartmann, W. M., & Johnson, D. (1991). Stream segregation and peripheral channeling. *Music Perception*, *9*, 155−184.

Helmholtz, H. L. F. von (1885). *On the sensations of tone as a physiological basis for the theory of music*. New York, NY: Dover. (A. J. Ellis, Trans. from the 4th German ed., 1877; republ. 1954).

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., & 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.

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

Kendall, R. A., Carterette, E. C., & Hajda, J. M. (1999). Perceptual and acoustical features of natural and synthetic orchestral instrument tones. *Music Perception*, *16*, 327−364.

Kobayashi, Y., & Osaka, N. (2008). Construction of an electronic timbre dictionary for environmental sounds by timbre symbol. *Proceedings of the International Computer Music Conference, Belfast*. San Francisco, CA: International Computer Music Association.

Krimphoff, J., McAdams, S., & Winsberg, S. (1994). Caractérisation du timbre des sons complexes. II: Analyses acoustiques et quantification psychophysique [Characterization of the timbre of complex sounds. II: Acoustical 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, & O. Olsson (Eds.), *Structure and perception of electroacoustic sound and music* (pp. 43−53). Amsterdam, The Netherlands: Excerpta Medica.

Krumhansl, C. L., & Iverson, P. (1992). Perceptual interactions between musical pitch and timbre. *Journal of Experimental Psychology: Human Perception and Performance*, *18*, 739−751.

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

Kruskal, J. (1964b). Nonmetric multidimensional scaling: A numerical method. *Psychometrika*, *29*, 115−129.

Lakatos, S. (2000). A common perceptual space for harmonic and percussive timbres. *Perception & Psychophysics*, *62*, 1426−1439.

Lakatos, S., McAdams, S., & Caussé, R. (1997). The representation of auditory source characteristics: Simple geometric form. *Perception & Psychophysics*, *59*, 1180−1190.

Lee, M., & Wessel, D. L. (1992). Connectionist models for real-time control of synthesis and compositional algorithms. *Proceedings of the 1992 International Computer Music Conference, San Jose* (pp. 277−280). San Francisco, CA: International Computer Music Association.

Lerdahl, F., & Jackendoff, R. (1983). *The generative theory of tonal music*. Cambridge, MA: MIT Press.

Lutfi, R. (2008). Human sound source identification. In W. Yost, A. Popper, & R. Fay (Eds.), *Auditory perception of sound sources* (pp. 13−42). New York, NY: Springer-Verlag.

Marozeau, J., de Cheveigné, A., McAdams, S., & Winsberg, S. (2003). The dependency of timbre on fundamental frequency. *Journal of the Acoustical Society of America*, *114*, 2946−2957.

Marozeau, J., & de Cheveigné, A. (2007). The effect of fundamental frequency on the brightness dimension of timbre. *Journal of the Acoustical Society of America*, *121*, 383−387.

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, & E. Bigand (Eds.), *Thinking in sound: The cognitive psychology of human audition* (pp. 146−198). Oxford, U.K.: Oxford University Press.

McAdams, S., & Bregman, A. S. (1979). Hearing musical streams. *Computer Music Journal*, *3*(4), 26−43.

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

McAdams, S., & Misdariis, N. (1999). Perceptual-based retrieval in large musical sound databases. In P. Lenca (Ed.), *Proceedings of Human Centred Processes '99, Brest* (pp. 445−450). Brest, France: ENST Bretagne.

McAdams, S., & Rodet, X. (1988). The role of FM-induced AM in dynamic spectral profile analysis. In H. Duifhuis, J. W. Horst, & H. P. Wit (Eds.), *Basic issues in hearing* (pp. 359−369). London, England: Academic Press.

McAdams, S., Chaigne, A., & Roussarie, V. (2004). The psychomechanics of simulated sound sources: Material properties of impacted bars. *Journal of the Acoustical Society of America*, *115*, 1306−1320.

McAdams, S., Depalle, P., & Clarke, E. (2004). Analyzing musical sound. In E. Clarke, & N. Cook (Eds.), *Empirical musicology: Aims, methods, prospects* (pp. 157−196). New York, NY: Oxford University Press.

McAdams, S., Roussarie, V., Chaigne, A., & Giordano, B. L. (2010). The psychomechanics of simulated sound sources: Material properties of impacted plates. *Journal of the Acoustical Society of America*, *128*, 1401−1413.

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

Meyer, L. B. (1989). *Style and music: Theory, history, and ideology*. Philadelphia, PA: University of Pennsylvania Press.

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

Momeni, A., & Wessel, D. L. (2003). Characterizing and controlling musical material intuitively with geometric models. In F. Thibault (Ed.), *Proceedings of the 2003 Conference on New Interfaces for Music Expression, Montreal* (pp. 54−62). Montreal, Canada: McGill University.

Moore, B. C. J., & Gockel, H. (2002). Factors influencing sequential stream segregation. *Acustica United with Acta Acustica*, *88*, 320−332.

Nattiez, J. -J. (2007). Le timbre est-il un paramètre secondaire? [Is timbre a secondary parameter?]. *Cahiers de la Société Québécoise de Recherche en Musique*, *9*(1−2), 13−24.

Opolko, F., & Wapnick, J. (2006). *McGill University master samples* [DVD set]. Montreal, Canada: McGill University.

Paraskeva, S., & McAdams, S. (1997). 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, Thessaloniki* (pp. 438−441). San Francisco, CA: International Computer Music Association.

Parncutt, R. (1989). *Harmony: A psychoacoustical approach*. Berlin, Germany: Springer-Verlag.

Patterson, R. D., Allerhand, M., & Giguère, C. (1995). Time-domain modeling of peripheral auditory processing: A modular architecture and a software platform. *Journal of the Acoustical Society of America*, *98*, 1890−1894.

Peeters, G., McAdams, S., & Herrera, P. (2000). Instrument sound description in the context of MPEG-7. *Proceedings of the 2000 International Computer Music Conference, Berlin* (pp. 166−169). San Francisco, CA: International Computer Music Association.

Peeters, G., Giordano, B. L., Susini, P., Misdariis, N., & McAdams, S. (2011). The Timbre Toolbox: Extracting audio descriptors from musical signals. *Journal of the Acoustical Society of America*, *130*, 2902−2916.

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, The Netherlands: Sijthoff.

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

Risset, J.-C. (2004). Timbre. In J.-J. Nattiez, M. Bent, R. Dalmonte, & M. Baroni (Eds.), *Musiques. Une encyclopédie pour le XXIe siècle. Vol. 2.: Les savoirs musicaux [Musics. An encyclopedia for the 21st century. Vol. 2: Musical knowledge]* (pp. 134−161). Paris, France: Actes Sud.

Risset, J.-C., & Wessel, D. L. (1999). Exploration of timbre by analysis and synthesis. In D. Deutsch (Ed.), *The psychology of music* (2nd ed., pp. 113−168). San Diego, CA: Academic Press.

Rose, F., & Hetrick, J. (2007). L'analyse spectrale comme aide à l'orchestration contemporaine [Spectral analysis as an aid for contemporary orchestration]. *Cahiers de la Société Québécoise de Recherche en Musique*, *9*(1−2), 63−68.

Roy, S. (2003). *L'analyse des musiques électroacoustiques: Modèles et propositions [The analysis of electroacoustic music: Models and proposals]*. Paris, France: L'Harmattan.

Rumelhart, D. E., & Abrahamson, A. A. (1973). A model for analogical reasoning. *Cognitive Psychology*, *5*, 1−28.

Saldanha, E. L., & Corso, J. F. (1964). Timbre cues and the identification of musical instruments. *Journal of the Acoustical Society of America*, *36*, 2021−2126.

Sandell, G. J. (1989). Perception of concurrent timbres and implications for orchestration. *Proceedings of the 1989 International Computer Music Conference, Columbus* (pp. 268−272). San Francisco, CA: International Computer Music Association.

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

Schoenberg, A. (1978). *Theory of harmony*. Berkeley, CA: University of California Press. (R. E. Carter, Trans. from original German edition, 1911).

Singh, P. G., & Bregman, A. S. (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.

Slawson, W. (1985). *Sound color*. Berkeley, CA: University of California Press.

Snyder, B. (2000). *Music and memory: An introduction*. Cambridge, MA: MIT Press.

Steele, K., & Williams, A. (2006). Is the bandwidth for timbre invariance only one octave? *Music Perception*, *23*, 215−220.

Tardieu, D., & McAdams, S. (in press). Perception of dyads of impulsive and sustained instrument sounds. *Music Perception*.

Tillmann, B., & McAdams, S. (2004). Implicit learning of musical timbre sequences: Statistical regularities confronted with acoustical (dis)similarities. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 1131−1142.

Traube, C., Depalle, P., & Wanderley, M. (2003). Indirect acquisition of instrumental gesture based on signal, physical and perceptual information. In F. Thibault (Ed.), *Proceedings of the 2003 Conference on New Interfaces for Musical Expression, Montreal* (pp. 42−47). Montreal, Canada: McGill University.

Vurma, A., Raju, M., & Kuuda, A. (2011). Does timbre affect pitch? Estimations by musicians and non-musicians. *Psychology of Music*, *39*, 291−306.

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

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

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

Winsberg, S., & De Soete, G. (1997). Multidimensional scaling with constrained dimensions: CONSCAL. *British Journal of Mathematical and Statistical Psychology*, *50*, 55−72.

Wright, J. K., & Bregman, A. S. (1987). Auditory stream segregation and the control of dissonance in polyphonic music. *Contemporary Music Review*, *2*(1), 63−92.