Perception and cognition of musical timbre[1]

Running title: Musical timbre

Stephen McAdams
Schulich School of Music
McGill University
555 Sherbrooke Street West
Montreal, QC
Canada H3A 1E3
smc@music.mcgill.ca

Kai Siedenburg
Department of Medical Physics and Acoustics
University of Oldenburg
Küpkersweg 74
26129 Oldenburg, Germany
kai.siedenburg@uni-oldenburg.de

Published in P. J. Rentfrow & D. J. Levitin (Eds.),
Foundations of Music Psychology: Theory and Research, pp. 71-120,
Cambridge, MA, MIT Press, 2019.

ABSTRACT
This chapter explains timbre as a perceptual property of a specific fused auditory event. Timbre refers to a complex set of auditory attributes that carry musical qualities, collectively contribute to sound source recognition and identification, and complement other auditory attributes of music. Psychophysics of timbre and covariance with other musical parameters will be discussed. This chapter will also explore absolute and relative perception of timbre, memory for timbre, and timbre as a structuring force in music perception. This chapter also includes discussion of current and future applications of timbre research for music information retrieval and musical machine learning.

KEYWORDS:
Timbre, Music perception, Music cognition, Hearing, Psychophysics, Neuroscience

---

[1] This chapter is an updated and expanded version of McAdams (2013).

Timbre refers to a complex set of auditory attributes that carry musical qualities, and that collectively contribute to sound source recognition and identification. It complements other auditory attributes of musical sounds such as pitch, loudness, duration, and spatial position. Timbral attributes arise from an event produced by a single acoustic or electroacoustic sound source or from events produced by several sound sources that are perceptually fused or blended into a single auditory image. Timbre is thus a perceptual property of a specific fused auditory event.

A major property of timbre is that it covaries with many other musical parameters in all acoustic and some electroacoustic instruments. For example, a specific oboe played with a given fingering (pitch) at a given playing effort (dynamic) with a particular articulation and embouchure configuration produces a note that has a distinct timbre. The timbre will change if any of these parameters are changed. Therefore, an instrument such as an oboe does not have "a timbre", it has a constrained universe of timbres that covary with the other musical parameters. For example, the timbres of clarinet sounds are vastly different in the lower (chalumeau) register than in the higher (clarion) register, and a trombone player can make the sound darker by playing a bit softer. There may, however, be certain acoustic invariants that are common across all of the events producible by an instrument that signal its identity. Timbre is thus a rather vague word that implies a multiplicity of perceptual qualities, some of which were addressed as early as the late 19th century by Helmholtz (1885/1954) followed by the seminal work of Stumpf (1926) and Schumann (1929). The vast majority of contemporary research has been conducted over the last 45 years or so, starting with the pioneering work of Plomp (1970) and Wessel (1973).

We now understand timbre to have two broad characteristics that contribute to the perception of music (see reviews in Hajda, Kendall, Carterette & Harshberger, 1997; Handel, 1995; McAdams, 1993; Risset, 2004):

1) It is a multitudinous set of perceptual attributes, some 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), others of which are continuously varying (e.g. attack sharpness, brightness, noisiness, richness, roughness). In this sense, timbre varies continuously as do the other auditory attributes of pitch, loudness and spatial position. Just as sounds can be higher or lower and louder or softer and to the left or right, up or down, near or far, they can also be more or less bright, rough, sharp in attack, rich, nasal, inharmonic, and a plethora of other qualities.
2) It is one of the primary perceptual vehicles for the recognition, identification, and tracking over time of a sounding object (tenor voice, violin, tubular bells), and thus is involved in the absolute categorization of a sounding object.

Understanding timbre perception thus involves a wide range of issues connecting the physics of sound sources to relevant aspects of perception and cognition: determining the properties of vibrating objects and of the acoustical 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, and understanding the role timbre can play in perceiving musical patterns and forms and in 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 perception of timbre relations, memory for timbre, 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 auditory streams and musical patterns, the role of timbre in the building and release of musical tension, and implicit learning of timbral grammars. We also include a brief survey of neuroscientific studies of timbre. We will conclude by examining a number of issues that have not been extensively studied yet—issues concerning the role of quantifying timbral characteristics 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 orchestration and electroacoustic music composition, and finally, consideration of timbre's status as a primary or secondary parameter in musical structure.

Psychophysics of timbre

One of the main approaches to timbre perception attempts to quantify the ways in which people perceive sounds to differ. The primary conception of timbre from the time of Helmholtz in the mid-1800s until the 1970s was in terms of its relation to spectral shape. 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, 2004b). Helmholtz (1885/1954) invented ingenious resonating devices for controlling spectral shape to explore these spectral aspects of timbre. He used air jets blowing out of tubes across tuned water jugs and varied the air speed stimulating each jug and the tuning of the jugs to vary the timbre. Licklider (1951) discussed various aspects of complex sounds but only concluded that "Until careful scientific work has been done on the subject, it can hardly be possible to say more about timbre than that it is a 'multidimensional' dimension" (p. 1019). Plomp (1970) explored the notion of timbre's multidimensionality, but 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. Wessel (1973) was the first to apply these to timbre perception and in particular began to emphasize the importance of time-varying aspects of sound for timbre perception. From this approach, the notion of timbre dimensions in a timbre space was developed.

Timbre space

Multidimensional scaling (MDS) makes no preconceptions about the physical or perceptual structure of the data it is being used to analyze. For timbre, listeners typically 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 spatial location 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 (or spatial map) 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 (Caclin, McAdams, Smith & Winsberg, 2005; Miller & Carterette, 1975; Plomp, 1976, chap. 6), 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 (Elliott, Hamilton & Theunissen, 2013; Iverson & Krumhansl, 1993; Lakatos, 2000; Wessel, 1973), and dyads of recorded instrument sounds (Kendall & Carterette, 1991; Tardieu & McAdams, 2012).

[Insert Figure 1 about here]

The basic MDS algorithm (Kruskal, 1964a,b), 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 algorithms, 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:

$$\partial_{ijc} = \sqrt{\sum_{d=1}^{D} w_{cd}\left(x_{id} - x_{jd}\right)^2 + v_c\left(s_i + s_j\right)}, \qquad \text{(Eq. 1)}$$

where ∂_{ijc} is the distance between sounds i and j for latent class c, x_{id} is the coordinate of sound i on dimension d, D is the total number of dimensions, w_{cd} is the weight on dimension d for class c, s_i is the specificity on sound i, and v_c is the weight on the whole set of specificities for class c. The basic algorithm doesn't model weights or specificities and only has one class of listeners. EXSCAL has specificities, but no weights. INDSCAL has no specificities, but has weights on each dimension for each listener. Finally, the CONSCAL algorithm allows for continuous mapping functions between audio descriptors and the position of sounds along a perceptual dimension to be modeled for each listener using spline functions, with the constraint 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 algorithms include: Kruskal's technique by Plomp (1976), INDSCAL by Wessel (1973), Miller and Carterette (1975), Plomp (1976), 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 algorithms in the context of timbre research are provided in McAdams et al. (1995) and Caclin et al. (2005), respectively. One of the difficulties of this approach is that the number of ratings that each listener has to make increases quadratically with the number of sounds to be compared. Elliott et al. (2013) used de Leeuw and Mair's (2009) SMACOF algorithm, which can perform multi-way constrained MDS in which multiple similarity ratings from different listeners are used for each pair of stimuli; a given listener only rates a subset of a large set of stimulus pairs.

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 yields a perceptual feature (often described as "hollowness") that is specific to that instrument. Such features might appear as specificities in the distance models derived with the EXSCAL and CLASCAL algorithms (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.

As an example, the timbre space reported by McAdams et al. (1995) is shown in Figure 2. It is based on dissimilarity ratings by 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 vibraphone-like and trombone-like 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 on the set of specificities differed.

[Insert Figure 2 about here]

One question that arises with timbre spaces is the extent to which the relations among sounds depend on the global stimulus context in which their dissimilarities are rated: if one were to add new sounds of a different nature, would the relations be distorted by the presence of the new sounds, perhaps due to making the listener focus on different sound properties? Grey and Gordon (1978) took half of the sounds from Grey (1977), paired sounds that differed primarily in terms of timbral brightness, exchanged their spectral envelopes, and resynthesized them. So, for example, the spectral envelope of the muted trombone was applied to the trumpet sound and vice versa. They reran an MDS study and found a space of similar dimensionality in which the sounds with exchanged envelopes switched positions along a dimension related to timbral brightness, as predicted. More importantly, McAdams and Giordano (2006) examined the distance relations among the sounds that were not changed and found that their relations remained constant across the two spatial models, demonstrating that the changes in the positions of the modified sounds did not affect those of the unmodified sounds. McAdams and Giordano also examined a more extreme case taken from Lakatos (2000) who had one timbre space of wind and string sounds, another with bowed and struck percussion sounds, and a third one that combined 10 sounds from each of the two spaces. A fine-grained analysis reported in McAdams (2015, chap. 2) showed that with the exception of one percussion instrument, the relations among the 10 sounds of each set maintained their dissimilarity relations in the presence of the very different new sounds from the other set. This important result suggests a certain robustness of timbre relations across different orchestration contexts. For example, in one section of a piece that uses subtle differences between different types of articulation produced by strings, listeners will hear the resulting timbral relations. Then if the composer adds brass and percussion in a later section, these perceptual relations won't necessarily be perturbed by the new orchestral context. This robustness may be related to timbre perception being strongly related to the recognition and categorization of sound sources as suggested by the results of McAdams and Giordano (2010).

Some algorithms, such as INDSCAL and CLASCAL, include individual and class differences, respectively, as weighting factors on the different dimensions for both algorithms and the set of specificities for CLASCAL. Members of each latent class have a similar weight structure in their data. For example, one group of listeners might pay more attention to spectral properties than to temporal aspects, whereas another might have the inverse pattern. The normalized weights on the dimensions and specificities for the five classes from McAdams et al. (1995) are shown in Figure 3. The majority of the 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 more of the scale than did those 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.

[Insert Figure 3 about here]

Interestingly, no study to date has provided behavioral evidence that such individual differences are related to 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 although listeners differ in terms of the perceptual weight accorded to the different dimensions, these inter-individual differences are unrelated to musical training based on biographical variables collected in questionnaires. 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. Nevertheless, there exists some neurophysiological evidence that timbre processing is affected by the instrument played by musicians (see section below on Neuroscientific studies of timbre).

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; Iverson & Krumhansl, 1993; Kendall et al., 1999; Krimphoff, McAdams, & Winsberg, 1994; McAdams et al., 1995). We call such parameters "audio descriptors", although they are also referred to as "audio features" in the field of music information retrieval (MIR). The correlates derived from musical instrument sounds that are most often found include:

1) spectral centroid (also called spectral center of gravity, 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 Fig. 4),
2) the logarithm of the attack time (distinguishing continuant instruments that are blown or bowed from impulsive instruments that are struck or plucked; see Fig. 5),
3) 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 Fig. 6), and

4) spectral deviation (the degree of jaggedness of the spectral shape, which is high for clarinet and vibraphone and low for trumpet; see Fig. 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 perception of stimulus dimensions related to spectral centroid, log attack time and spectral deviation, but did not confirm spectral flux.

[Insert Figures 4-7 about here]

Other studies have found a variety of acoustic correlates for the perceptual dimensions depending on the stimulus set. Grey and Gordon (1978) found a dimension that could be quantitatively related to spectral centroid, and qualitatively described two other dimensions as related to the synchronicity of the onsets of partial tones plus the presence of spectral fluctuations, on the one hand, and the presence of inharmonic attack transients, on the other. Iverson and Krumhansl (1993) found three two-dimensional spaces for: 1) whole tones, 2) tones with only the onsets (first 80 ms), and 3) tones with only the remainder of the sound after removing the onsets. In all cases, the spatial models yielded one dimension related to spectral centroid and another one related either to the attack time for onset-only tones or to the amplitude envelope shape for whole and remainder tones. Kendall et al. (1999) found dimensions that correlated with spectral centroid and spectral variability, as well as another that distinguished natural acoustic from synthesized sounds. McAdams (2015, chap. 2) reanalyzed data from Lakatos (2000) and found dimensions related to spectral centroid and attack time for a set of wind and string sounds, decay time and spectral density or pitch clarity/noisiness for a set of percussion sounds, and temporal envelope, spectral centroid and pitch clarity for a combined set of the two types of sounds.

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 descriptors (cf., Siedenburg, Fujinaga & McAdams, 2016). Over the years, a large set of descriptors has been developed at IRCAM in Paris starting with the work of Krimphoff et al. (1994), further developed in IRCAM's CUIDADO project (Rioux, McAdams, Susini & Peeters, 2006). 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 gammatone filterbank 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 (i.e., the time point that is the center of gravity of the amplitude envelope), effective duration and the frequency and amplitude of modulation in the energy envelope. Spectral shape descriptors include measures of the centroid, spread, skewness, kurtosis, slope, rolloff, crest factor, and jaggedness of the spectral envelope. The only spectrotemporal descriptor was 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,

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

whereas others represent time-varying quantities, such as the variation of spectral centroid over the duration of a sound event. Another popular set of audio descriptors with many related to timbre is the MIRtoolbox (Lartillot & Toiviainen, 2007).

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 set of over 6000 highly heterogeneous musical sounds from the McGill University Master Samples (Opolko & Wapnick, 2006), produced by all of the traditional orchestral instruments across their whole pitch range at various dynamics (from *pp* to *ff*) and with different playing techniques, such as flutter tonguing. 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 ten classes of descriptors that are relatively independent.

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 domains such as sound analysis and synthesis, and intelligent content-based search of sound databases (McAdams & Misdariis, 1999; Peeters, McAdams & Herrera, 2000). 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 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. For example, Siedenburg, Jones-Mollerup and McAdams (2016) have shown fairly robust generalization of timbre space representations to new sets of sounds, and even less dependence on the set of participants.

A completely different approach to timbre perception is to treat the representation of timbre as a monolithic high-dimensional structure rather than as a set of orthogonal dimensions. In *modulation spectrum* (Elliott et al., 2013) or *spectrotemporal receptive field* representations (Shamma, 2000), sound signals are characterized in terms of a higher-dimensional topography of the evolution of frequency-specific temporal-envelope profiles, in addition to the more traditional frequency (tonotopic) and amplitude variation over time. In this framework, sounds are characterized according to the dimensions of time, tonotopy, and modulation rate and scale. The latter two represent temporal modulations derived from the cochlear filter envelopes (rate dimension) and modulations present in the spectral shape derived from the spectral envelope (scale dimension), respectively. These representations have been proposed as potential models of timbre (Elliott et al., 2013; Patil et al., 2012). However, the modeling of timbre dissimilarity ratings with these representations is heavily based on dimensionality-reduction techniques performed by machine learning algorithms (e.g., projecting a 3,840D representation with 64 frequency, 10 rate and 6 scale filters into a 420D space in Patil et al., 2012). The results are difficult to interpret from a psychological standpoint. Indeed, in using high-dimensional modulation spectra as predictors of positions of timbres in low-dimensional MDS spaces, the more parsimonious acoustic descriptor approach has similar predictive power to the modulation spectrum approach (Elliott et al., 2013). One might therefore question whether timbre is indeed an emergent, high-dimensional spectrotemporal form or whether it relies on a limited bundle of orthogonal perceptual dimensions.

Interaction of timbre with pitch and dynamics

Most timbre space studies have restricted 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.

As mentioned in the introductory section, it is clear that for many instruments timbre varies as a function of pitch because many of 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 (B3, C#4, Bb4). 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 that listeners were no longer ignoring the pitch difference), and 2) pitch differences systematically affect the timbre dimension related to spectral centroid, suggesting a common sensory code for timbral brightness and pitch height. Handel and Erickson (2001) also found that nonmusician listeners had difficulty extrapolating the timbre of a sound source across large differences in pitch. However, Steele and Williams (2006) found that musician listeners can do so over a span greater than two octaves. Taken together, these results suggest that there are limits to timbral invariance across pitch, but that they depend on musical training.

Timbre can also affect pitch perception. Vurma, Raju and Kuuda (2011) have reported that timbre differences on two successive tones affected judgments of the in-tuneness of the pitches: when the second tone in a pair with identical fundamental frequencies had a brighter timbre than the first, it was judged as sharp and the inverse case, it was judged as flat. Krumhansl and Iverson (1992) found that speeded classification of pitches and of timbres were symmetrically affected by uncorrelated variation along the other parameter. Similar results were obtained by Allen and Oxenham (2014) who measured difference limens for musicians and nonmusicians using stimuli with concurrent random variations along the unattended dimension. Ensuring that the experimental units of timbre and pitch were of the same perceptual magnitude, these authors found symmetric mutual interference of pitch and timbre in the discrimination task. Musicians yielded higher discrimination overall, but there was no interaction of musicianship and auditory parameter (pitch/timbre). 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 for pitch and timbre that share common attributes, such as a tonotopic and periodicity organization 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., *ff* vs. *pp*) not only have greater energy at the frequencies present in the softer sound, but the spectrum spreads toward higher frequencies as more vibration modes of the physical system are excited, thus creating changes in several descriptors of spectral shape: 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 *pp* to *ff*. 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. For these three instruments, higher pitches received higher dynamic ratings in line with the pitch-driven dynamics principle described in many orchestration treatises, and which has been demonstrated perceptually (Nakamura, 1987). Thus estimates of the dynamics of musical tones are based both on loudness and timbre, and to a lesser degree on pitch as well.

Absolute and relative perception of timbre

Timbre as a vehicle for source identity

Another approach to timbre concerns its role in the recognition of the identity of a musical instrument or, in general, of a sound-generating event. 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). An evolutionary explanation for the development of this capacity is that timbre can aid in the identification of predators or prey through their vocalizations, and even the identification of individual conspecifics. In human and other species, infants imprint on their mother's voice and vice-versa.

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. Identification of isolated sounds was surprisingly poor for some instruments. When attacks and decays were excised, identification decreased markedly for some instruments, particularly when the attack was removed in sounds without vibrato. However, the effect of cutting the attack was less strong in the presence of vibrato, and identification was 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. 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. Along these lines, Siedenburg et al. (2016) presented listeners with recorded musical instrument sounds that were judged as highly familiar, as well as digitally transformed versions of these sounds judged to be highly unfamiliar. The dissimilarity ratings demonstrated that similarity between the source/cause mechanisms can affect perceived similarity, thereby confirming the meta-analysis results of Giordano and McAdams (2010).

Timbre processing for source recognition and categorization seems to be quite rapid. Agus, Suied, Thorpe and Pressnitzer (2012) found that listeners are very quick at making sound source category judgments (on the order of 150 ms taking detection time into account), which, they argue, reflects a highly efficient representation of familiar sound sources such as the human voice and musical instruments. Furthermore, Suied, Agus, Thorpe, Mesgarani and Pressnitzer (2014) used a gating paradigm in a similar categorization paradigm and found that timbre cues for source recognition are available at a variety of time scales down to threshold recognition for sounds as short as 8-16 ms. There was no effect of whether these short clips came from the beginning of a sound event or from a random location within the event, suggesting that the timbre information relevant for categorization is present throughout the sound's duration.

Several investigations on the perception of everyday sounds have extended the concept of timbre beyond the musical context (see Handel, 1995; Lutfi, 2008; McAdams, 1993, 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, 2004a), 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 (Giordano & McAdams, 2006; McAdams et al., 2004a) 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 role of timbre in source identification and categorization is perhaps its most neglected aspect, 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. This knowledge would help them track a given voice or instrument in a complex musical texture, much as we do when talking with someone in a crowded room. Listeners do this easily and 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. When 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. However, in some cases, it is exactly the tension created by absolute perception that is played upon by composers. Telling examples include Peter Ablinger's player piano pieces in which a piano mimics a speaking voice, and Jonathan Harvey's piece *Speakings* which attempts to make an orchestra speak. Here, the original sound sources are clearly recognizable as piano and orchestra, respectively, but their textures resemble speech sounds so closely that this may even create a situation of sound source bi-stability.

In summary, source-based distinctions, or what the electroacoustic composer Dennis Smalley (1994) calls "source bonding," can explain many aspects of data on musical timbre. The empirical evidence: 1) supports the source model of timbre perception (Hajda et al., 1997; Handel, 1995), 2) shows that musical listening is not completely source-independent (Gaver, 1993a,b), and 3) supports the presence of source attributes at the representational level and/or shows that different classes of acoustical systems generate aurally discriminable sounds. Whatever the theoretical conclusions, future research on musical timbre should focus on these perceptual distinctions, highlighting both its mechanical and acoustical correlates.

Perception of timbral relations

Whereas these models of timbre space may seem to be merely a theoretical abstraction, they have practical implications for composers and listeners. Consider the timbral trajectory shown in Figure 8 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 timbre intervals. If timbre 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.

[Insert Figure 8 about here]

Ehresman and Wessel (1978) took a first step forward in this direction. They proposed that timbre intervals may have properties similar to pitch intervals, i.e., 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 timbre interval can be considered as a vector in space connecting two timbres. It has a specific length (the distance between the timbres in an n-dimensional model) 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 pairs of timbres characterized by the same vector relation should have the same 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 (1978) tested this hypothesis using a task in which listeners compared two timbre intervals (e.g., A-B vs. C-D) and ranked 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 Fig. 9). 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 Fig. 9).

[Insert Figure 9 about here]

McAdams and Cunibile (1992) subsequently tested the vector model using the 3D space from Krumhansl (1989) (ignoring the specificities). 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 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 among electroacoustic composers than for nonmusicians.

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 timbre intervals correspond well to what is predicted because specificities are absent or low in value, whereas others are seriously distorted and are 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 timbre 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 timbre intervals may, in the long run, be limited to synthesized sounds or blended sounds created through the combination of several instruments.

Recent work has been conducted at McGill University in which participants were asked to reproduce a presented timbre interval along either the attack time or spectral centroid dimension starting from a different timbre. Wood (2015) found that listeners were not very good at reproducing the precise interval, although they rarely chose the wrong direction. This result suggests that timbre *contour* may be perceivable, but that we lack something essential in the way timbre is encoded that would allow us to perceive precise relations. Said another way, the timbre dimensions related to attack time and spectral centroid may exist on ordinal rather than on interval scales. Indeed, McDermott, Lehr and Oxenham (2008) argue that relative representations in the form of contours for pitch, timbre (brightness) and loudness may be a general feature of the auditory system, and may even have a common central locus.

Memory for Timbre

The various memory processes that subserve music listening are often taken for granted—perhaps by virtue of their ubiquity. In fact, the seemingly basic mnemonic operation of discriminating two musical sequences presented one after the other requires a complex mnemonic architecture that keeps track of sound identities and their serial order, and concurrently manages sensory processing, information storage, and matching of representations. Starting to shed light on mental processes such as these, the last decade has witnessed an upsurge of interest in memory for timbre, of which this section provides a brief synopsis.

Working memory for timbre

A guiding question of several studies in this domain is whether the retention and recognition of timbre is determined by similar principles compared to that of sequences of pitches or words. The classic conjecture by Miller (1956) states that people can only retain a limited number of 7 ± 2 independent chunks of information (as opposed to items per se) in immediate memory.

Providing empirical evidence across a wide range of domains such as verbal, visual, and auditory memory, Cowan (2001) outlined a contemporary revision of this hypothesis and argued that the capacity limit of short-term memory (STM) is only at around 4 ± 1 chunks if the involvement of other cognitive faculties such as long-term memory and active rehearsal processes is limited. Golubock and Janata (2013) first set out to measure the capacity of working memory for timbre. They used digitally synthesized sounds that varied along the dimensions of spectral centroid, attack time, and spectral flux, the discriminability of which was ensured via separate just-noticeble-difference measurements. They presented sequences of 2–6 tones in length that differed in timbre but were of constant pitch and loudness. Each sequence was followed by a delay of 1–6 s, and then a probe tone was presented for which participants had to judge whether it was part of the sequence or not (i.e., item recognition). The authors observed memory capacities at around 1.5 items, which notably undershoots the above-mentioned capacity estimates from other domains (Cowan, 2001). A second experiment used a more diverse set of sounds from commercial synthesizers and measured a significantly greater capacity (around 1.7 items), which suggests that an increase of perceptual variability (or timbral dissimilarity) within the stimulus set may enhance recognition memory. However, an additional factor could also have been an increase in participants' familiarity with stimuli from commercial synthesizers used in their second experiment.

In order to further disentangle the two factors of dissimilarity and familiarity, Siedenburg and McAdams (2016) compared the recognition of recorded tones from familiar acoustic instruments with that of unfamiliar synthesized tones that do not readily evoke sound-source categories. Three steps were taken in order to simultaneously control dissimilarity and familiarity:

1) The spectrotemporal signal envelopes and temporal fine structures of recorded sounds were purposely mismatched in order to generate a large set of novel and unfamiliar sounds; this approach has been demonstrated to yield "chimæric" perceptual properties in speech synthesis (cf., Smith, Delgutte, & Oxenham, 2002).
2) A group of musicians rated the familiarity of these transformed sounds, such that the 14 least familiar transformations could be used subsequently along with the 14 original recordings (rated as significantly more familiar than the transformations).

3) Using an item recognition task with sequences of three items in the main experiment, the mean timbral dissimilarity from the sounds in the sequence to the probe was equalized across recordings and transformations, using previously obtained pairwise dissimilarity ratings.

The first experiment contained list-probe delays of two and six seconds and only tested musician participants. The data revealed greater sensitivities for timbres of familiar recorded sounds compared to unfamiliar transformations, as well as better performance at shorter delays, but no interaction between the two factors. These results point towards a generally more robust form of encoding of timbral properties of familiar acoustic instruments, for which prior knowledge of instrument categories may act as representational anchors in the recognition process. Furthermore, there were significant correlations between correct rejection rates and the mean perceptual dissimilarity from the probe to the tones in the sequence for both recordings and transformations, that is, dissimilar lures were better recognized as such, highlighting the importance of perceptual similarity relations in STM.

In light of these effects of familiarity and similarity, it is suggestive to come back to the initial question on the relation of memory for timbre with other domains and to draw an analogy with verbal STM. Here the *lexicality effect* denotes the finding that words (part of the lexicon, familiar, and featuring meaningful associations) are better recognized than meaningless but phonologically coherent *pseudowords* (Thorn, Frankish & Gathercole, 2008). Furthermore, the hallmark effect of *phonological similarity* refers to the principle that lists of phonologically similar words are recalled less accurately than are dissimilar ones (Baddeley, 2012). In that sense, although the perceptual representations of words and the timbres of tones may be completely different, the retention of their memory traces may be governed by similar cognitive processes or principles (e.g., long-term memory, distinctiveness).

Another important property of verbal memory is its active nature—people tend to covertly vocalize in order to retain novel verbal information (think of phone numbers, street names, or vocabulary in a foreign language class). It has been argued that working memory for pitch sequences is supported by similar sensorimotor codes, which afford active subvocal rehearsal (e.g., Schulze & Koelsch, 2012). In the realm of timbre, the topic still is under debate: Comparing working memory for timbre, pitch, and words in a backwards sequence recognition task (listeners had to match a reversed comparison sequence), Schulze and Tillmann (2013) observed evidence for active retention of words and pitches, but not for timbre. To the contrary, the item recognition task of the second experiment of Siedenburg and McAdams (2016) showed a detrimental effect of a concurrent articulatory suppression task (participants had to count aloud during the delay interval), as well as of an attention-demanding visual dual task for both familiar recorded sounds and unfamiliar transformations. These results may indicate that participants keep timbral memory traces in an active state by actively attending to them, which has been called *attentional refreshing* in the verbal domain (Camos, Lagner & Barrouillet, 2009). That is, one portion of the active face of working memory for timbre may consist of listeners "replaying" timbral memory traces in their minds.

Timbre in memory for melodies and musical textures

Although timbre does not affect a melody's identity from a formal standpoint, a number of studies have highlighted the role of timbre as a salient auditory feature in memory for melodies. In the experiments of Radvansky, Fleming and Simmons (1995), participants identified which of two test melodies, a target and a distractor, was heard in the experiment's exposure phase. The accuracy of both musicians' and nonmusicians' recognition judgments was higher when the timbre of the test melody equaled the timbre of the exposure melody, that is, a change in instrumentation clearly impaired melody recognition. Poulin-Charronnat et al. (2004) studied recognition memory for both tonal (Liszt) and nontonal music (Reynolds). A timbre change from piano to orchestra or vice versa impaired recognition of tonal excerpts in both musicians and nonmusicians compared to conditions in which the timbre was held constant.

Most recently, Schellenberg and Habashi (2015) explored the temporal dynamics of musical memory by testing melody recognition with delays between exposure and test spanning ten minutes, one day, and one week. Surprisingly, recognition accuracies were similar for all three retention intervals, and there even seemed to be a trend for consolidation as reflected by a small but significant increase in accuracy for a delay of one week compared to ten minutes. Pitch transpositions of six semitones or a tempo shift of 64 beats per minute impaired recognition after ten minutes and one day, but not after one week. Notably, a change of instrument (from piano to saxophone) impaired melody recognition as strongly as the aforementioned changes in pitch or tempo, but unlike these parameters, the effect of timbre change did not reduce over time. Timbre change also perturbs melody recognition in infants (Trainor, Wu & Tsang, 2004). Overall these studies suggest that memory for melodies does not draw solely from an abstract lexicon of melodies to which perceptual tokens are matched for recognition, but that melody recognition relies on rich exemplars or auditory images that integrate various features including timbre.

Weiss and colleagues have added an interesting perspective to this field by advocating that not all timbres behave equally in memory: vocal melodies appear to be better recognized than melodies played by musical instruments (Weiss, Trehub & Schellenberg, 2012). Because the timbre of the voice is highly familiar and of great biological significance to humans, it may attract greater attention and thus be more robustly encoded in memory with more detail. Further research is required to isolate in greater detail the acoustic features that contribute to effects of voice processing superiority (cf., Bigand, Delbé, Gérard & Tillmann, 2011).

The central role of timbre in musical memory is further illustrated by studies that have explored the recognition of excerpts from popular music recordings. From "spinning the radio dial," we know that we can quickly classify musical genres and identify songs. But the results by Schellenberg, Iverson and McKinnon (1999) have shown that such snippets can be surprisingly short: even 100-ms excerpts from popular recordings could be matched to song title and artists with above-chance accuracy. Time-varying information in the high frequencies seemed to play an important role, because clips high-pass filtered at 1000 kHz were recognized with similar accuracy as unmodified clips. In comparison, matching low-pass filtered versions and clips being played backwards yielded lower accuracies.

Krumhansl (2010) used a more extensive set of clips taken from popular music hits of the last five decades in conjunction with a response format that provided less prior information to participants compared to Schellenberg et al. (2010). Based on listening to 400-ms clips, young college students freely identified artist and title in more than 25% of trials. Judgments of the release decade of songs were highly accurate ($r = .89$), and ratings of the clips' musical style as well as the emotion conveyed were consistent with ratings of 15-s excerpts from the same recordings. Because high-level information related to song identity, emotion or release date can be extracted from such short clips, the critical features of the auditory memory trace must be based on auditory images of the textural configuration of vocal and instrumental sounds, which is what Alluri and Toiviainen (2009) have called "polyphonic timbre."

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 or digital sound synthesis and processing of instrument sounds can give rise to new timbres if the sounds are perceived as blended. Timbral differences can also both segregate sounds into auditory streams of similar timbres and induce the segmentation of sequences when timbral discontinuities occur. 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.

Timbral blend

The creation of new timbres through orchestral combinations or signal processing necessarily depends on the degree to which the constituent sound sources fuse together perceptually or blend to create the newly emergent sound (Brant, 1971; Erickson, 1975; Reuter, 1996). Sandell (1995) has proposed 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 or highlights 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, the overall centroid of the combination, and the relation between peaks in the spectral envelopes of the constituent sounds. 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. 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 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 continuant sound (Tardieu & McAdams, 2012). Finally, Lembke and McAdams (2015) examined the relative contributions of global and local spectral envelope characteristics to perceived blend. They characterized wind-instrument spectra in terms of pitch-generalized spectral envelope descriptions that exhibit a formant-like structure with prominent spectral peaks. Two experiments employing blend-production and blend-rating tasks studied the perceptual relevance of these formants to the blending of dyads of a recorded instrument sound and a parametrically varied synthesized sound. Frequency relationships between formants influence blend critically, as does the degree of formant prominence. These results demonstrate the importance of spectral overlap in the perception of blend.

Timbre-based grouping in music

An important way in which timbre can contribute to the formation of auditory streams is that successive events with relatively similar 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 (see Bregman, 1990, chap. 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). Early demonstrations of auditory streaming on the basis of timbre (see Fig. 10) 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).

[Insert Figure 10 about here]

Hartmann and Johnson (1991) have argued that spectral aspects of timbre (such as spectral centroid) are primarily responsible for auditory streaming and that temporal aspects (such as attack time) have little effect. However, several subsequent studies have indicated an important role for both spectral and temporal attributes of timbre in auditory stream segregation (see Moore & Gockel, 2002, for a review). Iverson (1995), for example, used sequences alternating between two recorded instrument tones with the same pitch and loudness and asked listeners to rate the degree of segregation. The segregation ratings were treated as a measure of dissimilarity, and multidimensional scaling was performed to determine a "segregation space" in relation to which the acoustic attributes that contributed to the impression of auditory stream segregation were determined. 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 (Fig. 11). The timbre difference between target and distractor melodies was varied within the timbre space of McAdams et al. (1995). 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.

[Insert Figure 11 about here]

All of these results are important for auditory stream segregation theory, because they show that several acoustic properties of a sound source 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 was demonstrated in Figure 10. 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).

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. 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 do 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 7-point scale concerning the perceived degree of completion of the music at several points at which the music stopped. What resulted was a completion profile (Fig. 12), which could 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 six-voice fugue in the *Ricercar* from the *Musical Offering* by J. S. 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 an 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.

[Insert Figure 12 about here]

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 to the dependence of perceived roughness on the results of such processes (Wright & Bregman, 1987). Timbre, or any other auditory attribute of a unified sound event, is computed after auditory organization processes have grouped the bits of acoustic information together (Bregman & Pinker, 1978; McAdams & Bregman, 1979). It may be that the same is true of sensory dissonance or roughness, if we consider it to be property of a fused sound event. 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 have a greater tendency to be fused together. The computed roughness may then result from the interactions of all the frequency components of all the notes, although specific test of the effect of concurrent grouping on roughness perception has not been explicitly tested to our knowledge.

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) or the attack time difference could reduce the perceptual fusion (Tardieu & McAdams, 2012). In addition, because the timbres of these instruments are often quite different, several voices with different timbres may arrive momentarily at a given vertical sonority, but the verticality may not be perceived as such because the listener may continue to track individual instruments horizontally in separate auditory streams. So the attack asynchrony and the decomposition of simultaneities into timbrally similar sequentialities 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 harmonic complex. 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.

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 probabilities between timbres within the triplets 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 training 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.

In addition to the principle of higher transition probability between successive 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 (Fig. 13). To achieve this, three grammars were created. For the congruent sequence (S1), the timbres within each triplet were fairly close within the timbre space, and the distance between the last timbre of one triplet and the first timbre of the succeeding triplet was large. 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. Finally, a third sequence (S3) was composed so that all of the distances within and between triplets were uniformly medium within the space, thus avoiding segmentation.

[Insert Figure 13 about here]

After listening to one of the training sequences, listeners had to decide which of two timbre triplets presented in succession was present in the sequence just heard, with one being from the sequence and another not. Another group of listeners did not hear the training sequence beforehand and had to decide which of the two groups 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 artificial grammar (to which this group of listeners was never exposed) were scored as correct.

Listeners were best at recognizing the congruent triplets, less good for the neutral triplets, and worst for the incongruent triplets. The trained listeners were able to learn the grammar 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. But curiously, the amount of learning did not depend on the congruence between the grouping structure created by the acoustic discontinuities. 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.

Neuroscientific studies of timbre

Of the relatively little work that has started to investigate the neural underpinnings of timbre perception, several studies have addressed basic questions regarding the brain infrastructure for timbre perception, such as whether timbre processing is lateralized towards one specific hemisphere of cortex. On the other hand, research has demonstrated that neural circuitry in cortical and subcortical areas adapts in response to training and long-term experience with playing specific musical instruments.

Brain infrastructure for timbre processing

Results from early neuropsychological studies suggest that timbre processing depends on the right hemisphere of neocortex. More specifically, Samson and Zatorre (1994) found that patients with unilateral excisions in the right temporal lobe showed significant deficits in the discrimination of rise times and spectral envelopes of synthesized tones compared to left temporal lobe patients and normal controls. Timbre dissimilarity ratings obtained from right temporal lobe patients exhibited distorted timbre spaces compared to normal controls. In particular the temporal aspect of rise time information was poorly represented in these patients' ratings (Samson, Zatorre & Ramsay, 2002).

In comparison to patient studies, however, conclusions about the localization of timbre processing that have been drawn from functional magnetic resonance imaging (fMRI) have been less clear-cut. For instance, Menon et al. (2002) recorded fMRI while participants listened to short melodies. Each melody was composed by one out of two digitally synthesized harmonic sounds with different values of attack time, spectral centroid, and spectral flux. Data analysis focused on the contrast of the hemodynamic response between blocks of melodies that differed in timbre. Although the observed extent of activation was not significantly different between left and right hemispheres, the imaging data indicated that the activation of the temporal lobe was significantly more posterior on the left side. Therefore, this finding suggests a functional asymmetry of the left and right hemispheres in their respective contributions to timbre processing. Halpern and Zatorre (2004) recorded fMRI during a timbre perception and imagery task that had listeners judge the dissimilarity between perceived or imagined timbres of musical instrument sounds. The perception condition used recorded instrumental sounds. As a control condition, the authors used a visual imagery task which presented instrument names to participants and let them judge the similarity of "the shape represented by each word". Similar to the auditory perception and imagery tasks, this task also involved keeping two items in working memory and comparing them, and thus was used as a contrast in order to factor out the brain activity of interest associated with auditory perception and imagery. Their results showed a bilateral activation of primary and secondary auditory cortex with more activity on the right side of cortex. Not surprisingly, however, activity in the auditory cortex was stronger in the perception condition compared to the auditory imagery condition.

Other studies have sought neurophysiological markers of the processing of separate acoustic dimensions of timbre. Caclin et al. (2006) used electroencephalography (EEG) in order to investigate whether there exist independent neural signatures that are associated with the processing of independent acoustic dimensions. The authors used the carefully controlled stimuli from Caclin et al. (2005) in order to test for effects of changes along single dimensions (rise time, spectral center of gravity, even-harmonic attenuation) and combinations of these on the mismatch negativity (MMN). The MMN is a well-known event-related potential that indexes the discrepancy between the current sound and traces in auditory sensory memory. The authors tested whether the MMNs obtained from stimuli deviating along a single timbre dimension added up in the case of changes along multiple dimensions. If confirmed, this would suggest independent neural processing of the separate timbral dimensions. Their results indicated that the two-dimensional combinations of rise time and even-harmonic attenuation, as well as spectral center of gravity and even-harmonic attenuation showed roughly additive behavior, suggesting independent processing of these dimensions (Fig. 14). However, the pair of rise time and spectral center of gravity yielded sub-additive MMNs, that is, the sum of MMNs in response to single dimensional deviants was considerably larger than the MMNs recorded for bi-dimensional deviants. These results are intriguing because one could have expected spectral and temporal acoustic information to be processed fairly independently. Nonetheless, this study posits an example of how research may begin to tease apart neural responses to separate timbral dimensions.

[Insert Figure 14 about here]

Timbre and neural plasticity

Does playing a musical instrument alter the neural processing of basic auditory attributes such as timbre? Growing neuroimaging data suggests an affirmative answer to this question. Using magnetoencephalography, Pantev, Roberts, Schulz, Engelien and Ross (2001) observed that professional trumpet players and violinists exhibited stronger N1 event-related potential components to sounds from their own instrument. The N1 response indexes pre-attentive neural processing related to stimulus detection. Using fMRI, Margulis, Mlsna, Uppunda, Parish and Wong (2009) recorded hemodynamic activity of professional flutists and violinists while listening to instrumental sounds. They observed that flutists' left superior temporal gyrus exhibited a stronger hemodynamic response to flute sounds compared to violin sounds, contrary to violinists, who responded more strongly to violin sounds. This interaction effect between participant group and the stimulus type provides further evidence that expertise on a specific musical instrument alters cortical processing of timbre. Recent findings have even suggested that training and experience not only affect cortical activity, but modulate even "low-level processing": Strait, Chan, Ashley and Kraus (2012) demonstrated that brainstem recordings of pianists more closely correlated with the amplitude envelopes of the original piano sounds, compared to recordings of musicians who did not have extensive experience with the piano. At the same time, there was no difference between groups of participants for sounds from the tuba and bassoon. By and large, these studies yield converging evidence for the idea that basic neural circuitry adapts in order to facilitate the processing of familiar timbres. A challenge for future research will be to tie investigations of neural processing more closely to their perceptual implications—to the best of our knowledge, no published data exist yet that demonstrate analogous effects of instrument-specific expertise in behavioral timbre perception tasks.

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. Timbre intervals, for example, can be conceived as vectors within the space of common dimensions. 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 meta-data related to timbre that can be used in automatic instrument recognition and categorization (Joder, Essid & Richard, 2009), 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 (Esling, Carpentier & Agon, 2010), 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 (Momeni & Wessel, 2003; Schwarz & O'Leary, 2015).

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 and brightness are major components of musical tension. However, they strongly depend 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 can play 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 when it represents 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 which psychologists have yet to tackle with any scope or in any depth.

Nattiez (2007) in particular has examined 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, i.e., 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 only allowed to play 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

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

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. 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.

Acknowledgments

The preparation of this chapter was supported by grants to SMc from the Natural Sciences and Engineering Research Council, the Social Sciences and Humanities Research Council of Canada, the Fonds de recherche du Québec—Société et culture, and the Canada Research Chairs program.

References

Agus, T. R., Suied, C., Thorpe, S. J., & Pressnitzer, D. (2012). Fast recognition of musical sounds based on timbre. *Journal of the Acoustical Society of America*, *131*(5), 4124–4133.

Allen, E. J., & Oxenham, A. J. (2014). Symmetric interactions and interference between pitch and timbre. *Journal of the Acoustical Soceity of America, 135*(3), 1371-1379.

Alluri, V., & Toiviainen, P. (2009). Exploring perceptual and acoustical correlates of polyphonic timbre. *Music Perception, 27*(3), 223-241.

Baddeley, A. D. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology,* 63:1–29.

Barthet, M., Kronland-Martinet, R. & Ystad, S. (2007). Improving musical expressiveness by time-varying brightness shaping. *Computer Music Modeling and Retrieval. Sense of Sounds*. LNCS (pp. 313-336). Berlin: 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., Delbé, C., Gérard, Y. & Tillmann, B. (2011). Categorization of extremely brief auditory stimuli: Domain-specific or domain-general processes? *PloS ONE*, 6(10):e27024.

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: 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.
Bregman, A. S. & Pinker, S. (1978). Auditory streaming and the building of timbre. *Canadian Journal of Psychology*, *32*(1), 19-31.
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.
Caclin, A., Brattico, E., Ternaviemi, M., Näätänen, R., Morlet, D., Giard, M.-E. & McAdams, S. (2006). Separate neural processing of timbre dimensions in auditory sensory memory. *Journal of Cognitive Neuroscience*, *18*(12), 1959-1972.
Camos, V., Lagner, P. & Barrouillet, P. (2009). Two maintenance mechanisms of verbal information in working memory. *Journal of Memory and Language*, 61(3):457–469.
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.
Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, 24(1):87–114.
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.
de Leeuw, J. & Mair, P. (2009). Multidimensional scaling using majorization: SMACOF in R. *Journal of Statistical Software*, *31*(3), 1–30.
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: IRCAM-Centre Pompidou.
Elliott, T. M., Hamilton, L. S., & Theunissen, F. E. (2013). Acoustic structure of the five perceptual dimensions of timbre in orchestral instrument tones. *Journal of the Acoustical Soceity of America, 133*(1), 389-404.
Erickson, R. (1975). *Sound Structure in Music*. Berkeley, CA: University of California Press.
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: 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.
Gaver, W. W. (1993a). What in the world do we hear?: An ecological approach to auditory event perception. *Ecological Psychology, 5*(1), 1-29.
Gaver, W. W. (1993b). How do we hear in the world?: Exploration in ecological acoustics. *Ecological Psychology, 5*(4), 285-313.
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.
Goldinger, S. D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory and Cognition, 22*, 1166-1183.
Golubock, J. L., & Janata, P. (2013). Keeping timbre in mind: Working memory for complex sounds that can't be verbalized. *Journal of Experimental Psychology: Human Perception and Performance, 39*(2), 399-412.
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: Psychology Press.
Halpern, A. R., Zatorre, R. J., Bouffard, M., & Johnson, J. A. (2004). Behavioral and neural correlates of perceived and imagined musical timbre. *Neuropsychologia, 42*, 1281-1292.
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.
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*. Trans. by A. J. Ellis from the 4th German ed., 1877; republ. 1954. New York: Dover.
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.
Joder, C., Essid, S., and Richard, G. (2009). Temporal integration for audio classification with application to musical instrument classification. *IEEE Transactions on Audio, Speech, and Language Processing*, *17*(1), 174–186.
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: Excerpta Medica.
Krumhansl, C. L. (2010). Plink: "thin slices" of music. *Music Perception*, 27(5):337–354.
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.
Lartillot, O. & Toiviainen, P. (2007). A Matlab toolbox for musical feature extraction from audio. In *Proceedings of the 10th International Conference on Digital Audio Effects (DAFx),* Sep 10 – 15, Bordeaux, France (pp. 237–244).
Lembke, S., & McAdams, S. (2015). The role of spectral-envelope characteristics in perceptual blending of wind-instrument sounds. *Acta Acustica united with Acustica, 101*, 1039 – 1051.
Licklider, J. C. R. (1951). Basic correlates of the auditory stimulus. In S. S. Stevens (Ed.), *Handbook of experimental psychology* (pp. 985-1039). New York, NY: John Wiley and Sons.
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.

Margulis, E. H., Mlsna, L. M., Uppunda, A. K., Parrish, T. B., & Wong, P. (2009). Selective Neurophysiologic Responses to Music in Instrumentalists with Different Listening Biographies. *Human Brain Mapping, 30*, 267-275.
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: Oxford University Press.
McAdams, S. (2013). Musical timbre perception. In D. Deutsch (Ed.), *The psychology of music* (3rd ed., pp. 35-67). San Diego, CA: Elsevier.
McAdams, S. (2015). *Perception et cognition de la musique* [*Perception and cognition of music*]. Paris, France: J. Vrin.
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. & Giordano, B. L. (2006). Generalizing timbre space data across stimulus contexts: The meta-analytic approach. *Journal of the Acoustical Society of America*, *119*, 3395 (A).
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: 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: Academic Press.
McAdams, S., Chaigne, A. & Roussarie, V. (2004a). 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. (2004b). Analyzing musical sound. In E. Clarke & N. Cook (Eds.), *Empirical Musicology: Aims, Methods, Prospects* (pp. 157-196). New York: 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.
McDermott, J.H., Lehr, A.J. & Oxenham, A.J. (2008). Is relative pitch specific to pitch? *Psychological Science*, *19*(12), 1263-1271.
Menon, V., Levitin, D. J., Smith, B. K., Lembke, A., Krasnow, B. D., Glazer, D., . . . McAdams, S. (2002). Neural correlates of timbre change in harmonic sounds. *Neuroimage, 17*(4), 1742-1754.
Meyer, L. B. (1989). *Style and Music. Theory, History, and Ideology*. Philadelphia: University of Pennsylvania Press.
Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63(2):81–97.
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: McGill University.
Moore, B. C. J. & Gockel, H. (2002). Factors influencing sequential stream segregation. *Acta Acustica united with Acustica*, *88*, 320-332.
Nakamura, T. (1987). The communication of dynamics between musicians and listeners through musical performance. *Perception & Psychophysics*, *41*(6), 525-533.
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: McGill University.
Pantev, C., Roberts, L. E., Schulz, M., Engelien, A., & Ross, B. (2001). Timbre-specific enhancement of auditory cortical representations in musicians. *Cognitive Neuroscience and Neuropsychology, 12*(1), 169-174.
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: International Computer Music Association.

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

Patil, K., Pressnitzer, D., Shamma, S., & Elhilali, M. (2012). Music in Our Ears: The Biological Bases of Musical Timbre Perception. *PLOS Computational Biology, 8*(11), 1-16.

Patterson, R. D., Allerhand, M., and 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: 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: Sijthoff.

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

Poulin-Charronnat, B., Bigand, E., Lalitte, P., Madurell, F., Vieillard, S., & McAdams, S. (2004). Effects of a change in instrumentation on the recognition of musical materials. *Music Perception, 22*(2), 239-263.

Radvansky, G. A., Fleming, K. J., & Simmons, J. A. (1995). Timbre reliance in nonmusicians' and musicians' memory for melodies. *Music Perception: An Interdisciplinary Journal, 13*(2), 127-140.

Reuter, C. (1996). *Die auditive Diskrimination von Orchesterinstrumenten – Verschmelzung und Heraushörbarkeit von Instrumentalklangfarben im Ensemblespiel* [*Auditory discrimination of orchestral instruments—Blending and detection of instrumental timbres in ensemble playing*]. Frankfurt am Main: P. Lang.

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: Actes Sud.

Rioux, V., McAdams, S., Susini, P. & Peeters, G. (2006). Psycho-acoustic timbre descriptors. IRCAM CUIDADO Technical Report WP2.1.5.

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

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.

Samson, S., & Zatorre, R. (1994). Contribution of the right temporal lobe to musical timbre discrimination. *Neuropsychologia, 32*(2), 231-240.

Samson, S., Zatorre, R., & Ramsay, J. O. (2002). Deficits of musical timbre perception after unilateral temporal-lobe lesion revealed with multidimensional scaling. *Brain, 125*, 511-523.

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.

Schellenberg, E. G., & Habashi, P. (2015). Remembering the melody and timbre, forgetting the key and tempo. *Mem Cognit, 43*, 1021-1031. doi:10.3758/s13421-015-0519-1

Schellenberg, E. G., Iverson, P., & McKinnon, M. C. (1999). Name that tune: Identifying popular recordings from brief excerpts. *Psychonomic Bulletin & Review*, 6(4):641–646.

Schellenberg, E. G. & Habashi, P. (2015). Remembering the melody and timbre, forgetting the key and tempo. *Memory & Cognition*, 43(7):1021–1031.

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

Schulze, K. & Koelsch, S. (2012). Working memory for speech and music. *Annals of the New York Academy of Sciences*, 1252(1):229–236.Schulze, K., & Tillmann, B. (2012). Working memory for pitch, timbre, and words. *Memory, 21*(3), 377-395.

Schumann, K. E. (1929). *Physik der Klangfarben [The physics of sound color]*. Professorial dissertation, Universität Berlin.

Schwarz, D. and O'Leary, S. (2015). Smooth granular sound texture synthesis by control of timbral similarity. In Timoney, J. and Lysaght, T. (Eds.), *Proceedings of the 12th Sound and Music Computing Conference*, Maynooth, Ireland, Jul 26 – Aug 1, 2015 (pp. 471–476).

Shamma, S. (2000). Physiological Basis of Timbre Perception. In M. Gazzzaniga (Ed.), *The New Cognitive Neurosciences*. Cambridge, MA: MIT Press.

Siedenburg, K. & McAdams, S. (2016). The role of long-term familiarity and attentional maintenance in short-term memory for timbre. *Memory*. (published online 17 June 2016) doi:10.1080/09658211.2016.1197945

Siedenburg, K., Fujinaga, I. & McAdams, S. (2016). A comparison of approaches to timbre descriptors in music information retrieval and music psychology. *Journal of New Music Research*, *45*(1), 27-41.

Siedenburg, K., Jones-Mollerup, K. & McAdams, S. (2016). Acoustic and categorical dissimilarity of musical timbre: Evidence from asymmetries between acoustic and chimeric sounds. *Frontiers in Psychology*, *6*, art. 1977.

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.

Smalley, D. (1994). Defining Timbre- Refining Timbre. *Contemporary Music Review, 10*(2), 35-48.

Smith, Z. M., Delgutte, B., & Oxenham, A. J. (2002). Chimaeric sounds reveal dichotomies in auditory perception. *Nature, 416*, 87-90.

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.

Strait, D. L., Chan, K., Ashley, R. & Kraus, N. (2012). Specialization among the specialized: Auditory brainstem function is tuned in to timbre. *Cortex*, 48(3):360–362.

Stumpf, C. (1926). *Die Sprachlaute: experimentell-phonetische Untersuchungen nebst einem Anhang über Instrumentalklänge [Speech sounds: An experimental phonetic investigation with an appendix on instrument sounds]*. Berlin, Germany: Springer.

Suied, C., Agus, T. R., Thorpe, S. J., Mesgarani, N., & Pressnitzer, D. (2014). Auditory gist: Recognition of very short sounds from timbre cues. *Journal of the Acoustical Society of America*, *135*(3), 1380-1391.

Tardieu, D. & McAdams, S. (2012). Perception of dyads of percussive and sustained instruments. *Music Perception, 30*, 117-128.

Thorn, A. S., Frankish, C. R. & Gathercole, S. E. (2008). The influence of long-term knowledge on short-term memory: Evidence for multiple mechanisms. In Thorn, A. S. and Page, M. (Eds.), *Interactions between short-term and long-term memory in the verbal domain* (pp. 198–219). New York, NY: Psychology Press.

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.

Trainor, L. J., Wu, L., & Tsang, C. D. (2004). Long-term memory for music: Infants remember tempo and timbre. *Developmental Science*, *7*(3), 289-296.

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: 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.

Weiss, M. W., Trehub, S. E., & Schellenberg, E. G. (2012). Something in the way she sings: Enhanced memory for vocal melodies. *Psychological Science, 23*(10), 1074-1078.

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.
Wood, C. (2015). *Perception of timbre intervals*. Master's thesis, McGill University, Montreal, Canada.
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.

Figure captions

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

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. (bsn=bassoon, cnt=clarinet, ehn=English horn, fhn=French horn, gtn=*guitarnet* (guitar/clarinet hybrid), gtr=guitar, hcd=harpsichord, hrp=harp, obc=*obochord* (oboe/harpsichord hybrid), obl=*obolesta* (oboe/celesta hybrid), pno=piano, sno=*striano* (bowed string/piano hybrid), stg=bowed string, tbn=trombone, tpr=*trumpar* (trumpet/guitar hybrid), tpt=trumpet, vbn=*vibrone* (vibraphone/trombone hybrid) , vbs=vibraphone) [Modified from Fig. 1, McAdams et al. (1995). ©1995 by Springer-Verlag. Adapted with permission.]

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.

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 arrow head on the x axis indicates the location of the spectral centroid (expressed in harmonic rank). 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.

Figure 5. Log attack time in relation to the first dimension of Krumhansl's (1989) space. The left and right graphs show the amplitude envelope of the vibraphone and bowed piano sounds. The attack time is shown with the arrowheads.

Figure 6. Spectral flux in relation to the third dimension of McAdams et al.'s (1995) space, which used 18 of the 21 sounds in the Krumhansl (1989) study. The upper and lower graphs show the variation over time of the spectral centroid for the trombone and the sampled piano, respectively. 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 Krumhansl's (1989) space. The left and right graphs show the frequency spectra and global spectral envelopes of the trumpet and clarinet sounds, respectively. 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.

Figure 8. 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?

Figure 9. Examples of timbre 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.

Figure 10. 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.

Figure 11. 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 Fig. 2, Bey & McAdams (2003). ©2003 by The American Psychological Association, Inc. Adapted with permission.]

Figure 12. 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 & McAdams (1997). ©1997 by the authors. Adapted with permission.]

Figure 13. Examples of timbre triplets used by Tillmann and McAdams (2004) 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.

Figure 14. MMN response to stimulus change along two or three dimensions compared to the sume of unidimensional reponses as measured at the Fz electrode. ATT=attack time, SCG=spectral center of gravity, EHA=even-harmonic attenuation. [Adapted from Fig. 4, Caclin et al. (2006). ©2006 by the Massachussetts Institute of Technology.]

Figure 1

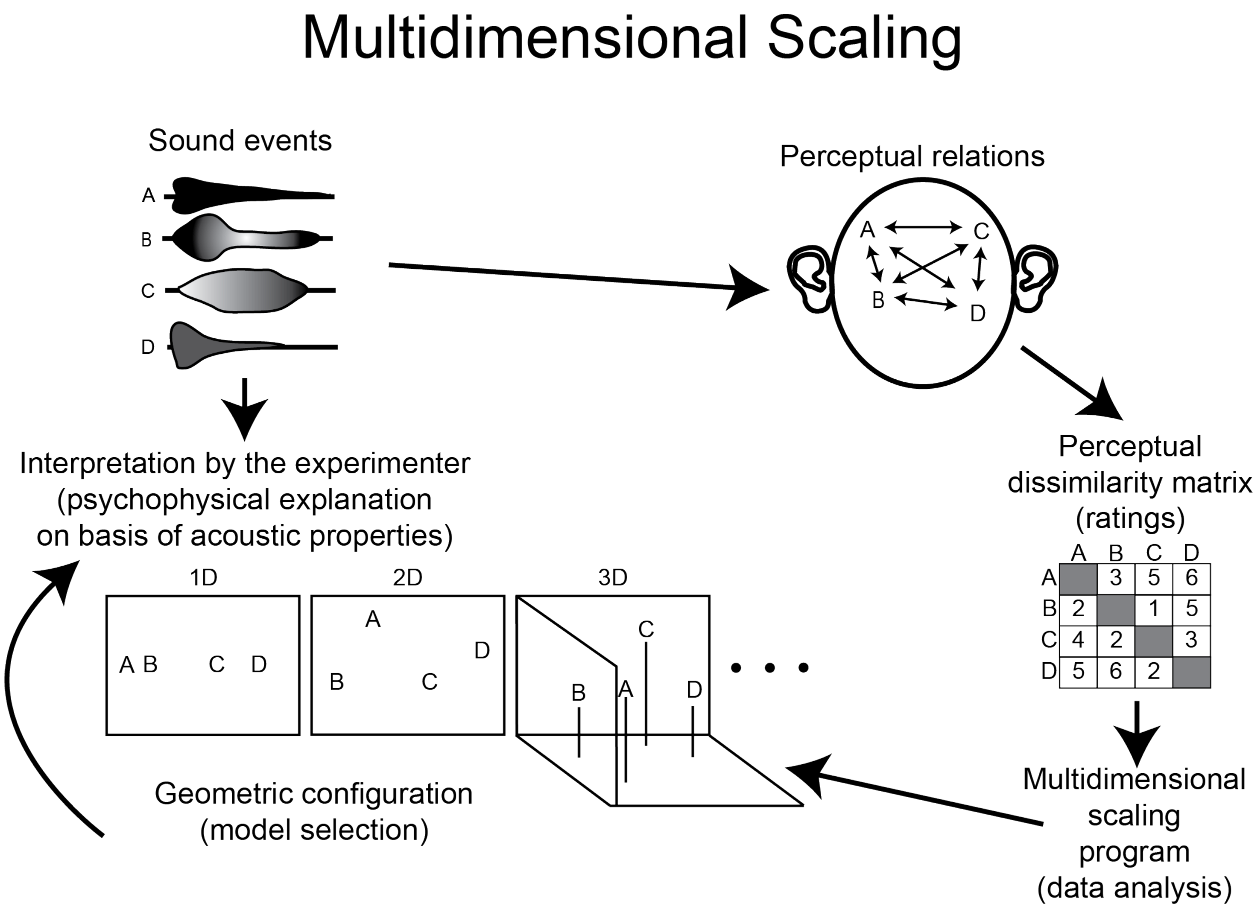

Multidimensional Scaling
Sound events
A
B
C
D
Perceptual relations
A
C
B
D
Perceptual
dissimilarity matrix
(ratings)
A B C D
A 3 5 6
B 2 1 5
C 4 2 3
D 5 6 2
Multidimensional
scaling
program
(data analysis)
Interpretation by the experimenter
(psychophysical explanation
on basis of acoustic properties)
1D
A B C D
2D
A
D
B C
3D
C
B A D
Geometric configuration
(model selection)

Figure 2

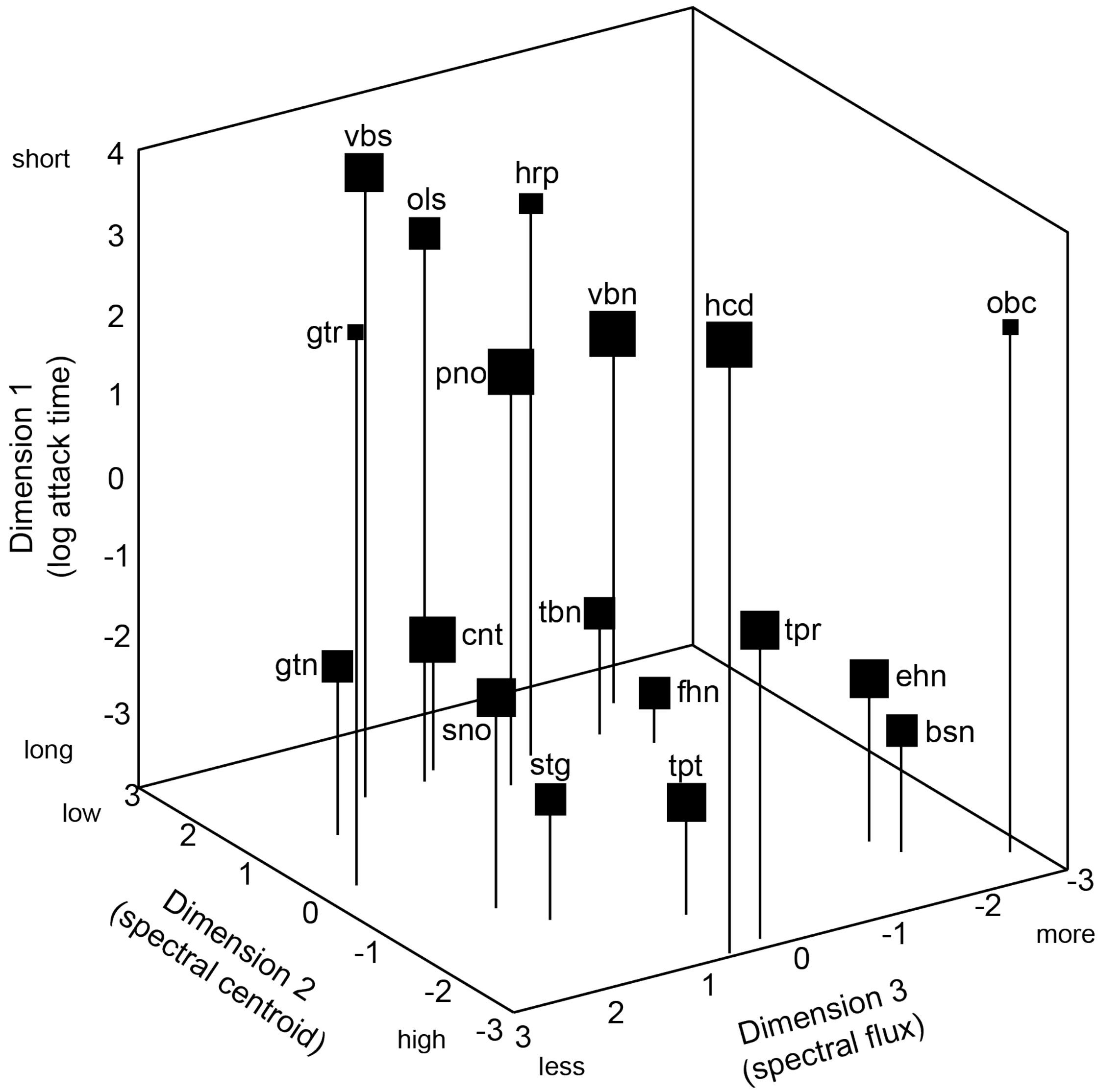
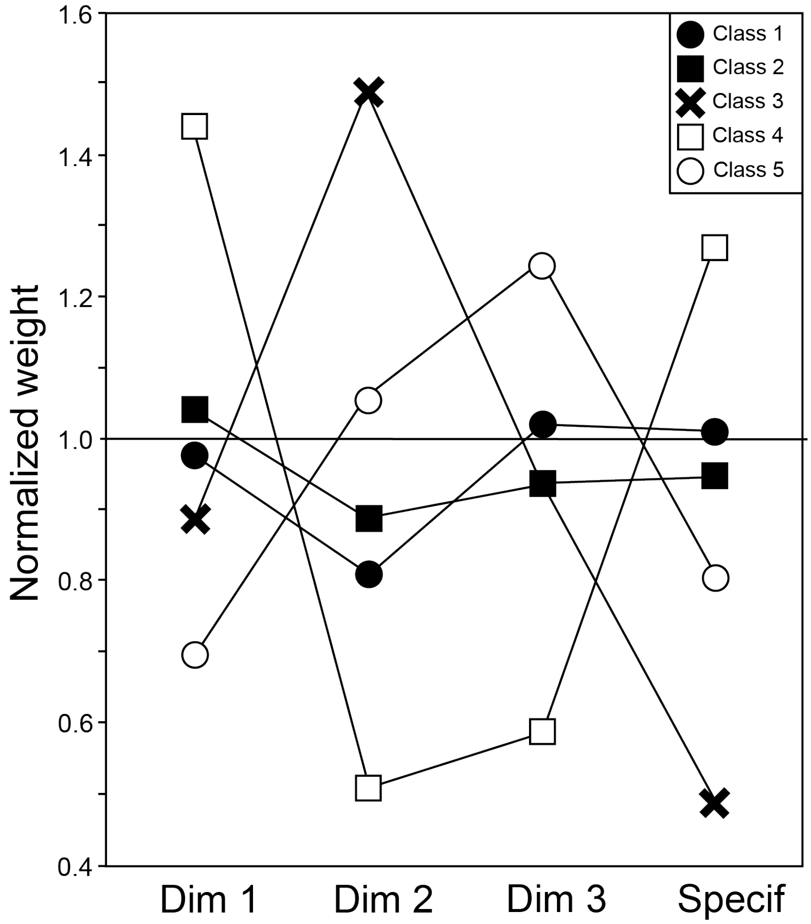

short
4
3
2
1
0
-1
-2
-3
long
Dimension 1
(log attack time)
vbs
ols
hrp
gtr
pno
vbn
hcd
obc
tbn
cnt
tpr
gtn
fhn
ehn
sno
bsn
stg
tpt
low
3
2
1
0
-1
-2
-3
high
Dimension 2
(spectral centroid)
3
2
1
0
-1
-2
-3
less
more
Dimension 3
(spectral flux)

Figure 3



Normalized weight
1.6
1.4
1.2
1.0
0.8
0.6
0.4
Class 1
Class 2
Class 3
Class 4
Class 5
Dim 1
Dim 2
Dim 3
Specif

Figure 4

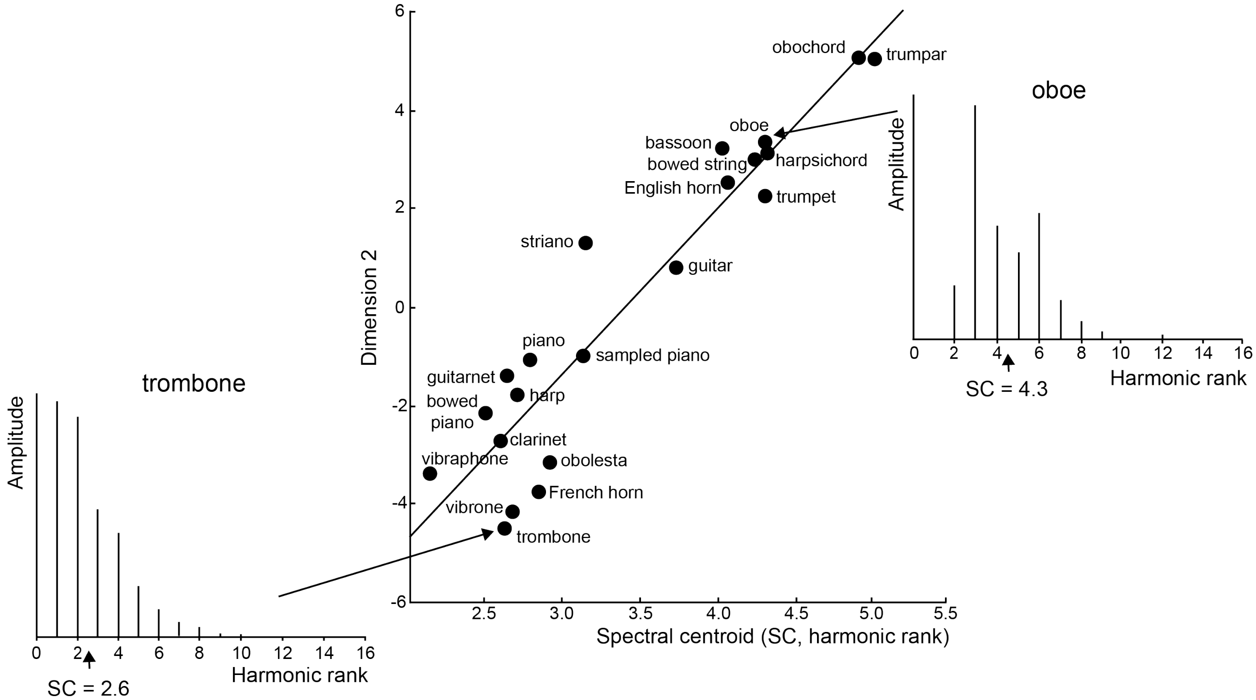

trombone
Amplitude
Harmonic rank
SC = 2.6
Dimension 2
Spectral centroid (SC, harmonic rank)
obochord
trumpar
oboe
bassoon
bowed string
harpsichord
English horn
trumpet
striano
guitar
piano
sampled piano
guitarnet
harp
bowed piano
clarinet
vibraphone
obolesta
French horn
vibrone
trombone
oboe
Amplitude
Harmonic rank
SC = 4.3

Figure 5

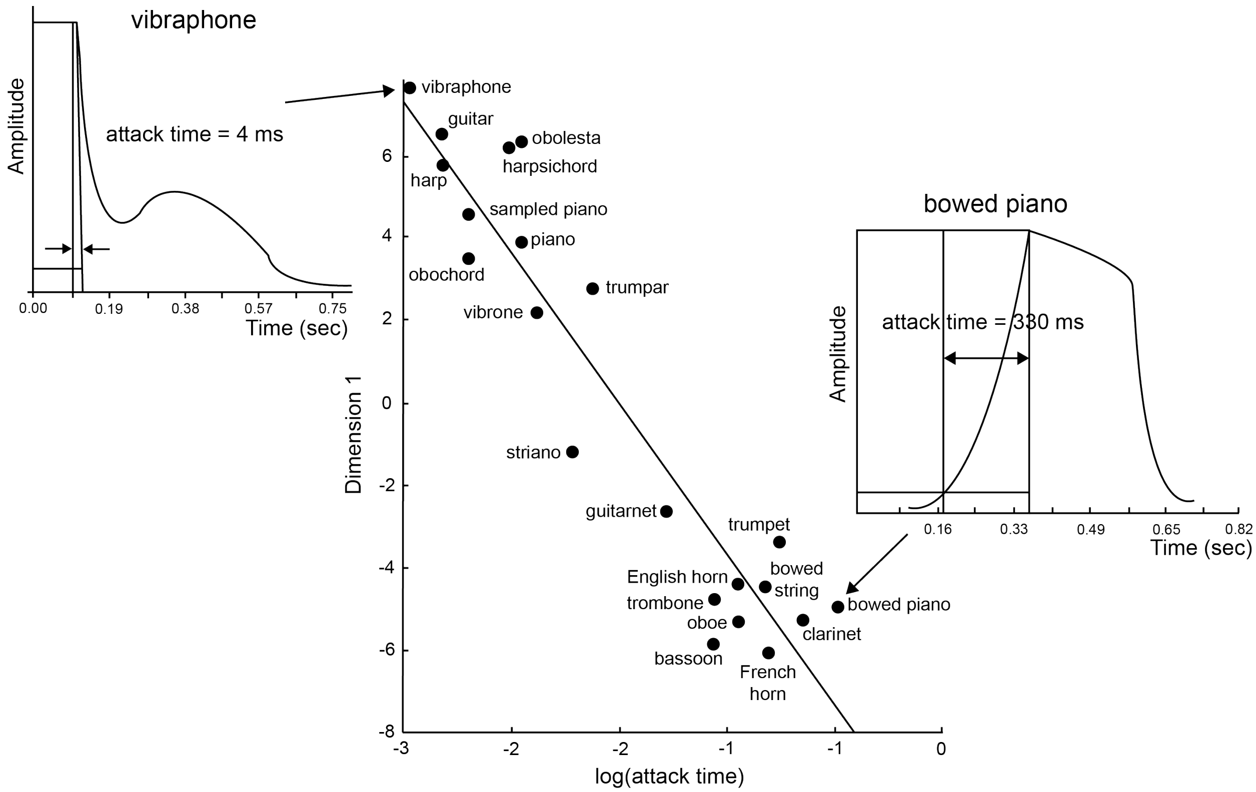

vibraphone
Amplitude
attack time = 4 ms
0.00
0.19
0.38
0.57
0.75
Time (sec)
vibraphone
guitar
obolesta
harpsichord
harp
sampled piano
piano
obochord
trumpar
vibrone
striano
guitarnet
trumpet
English horn
bowed
string
bowed piano
trombone
oboe
clarinet
bassoon
French
horn
Dimension 1
log(attack time)
bowed piano
Amplitude
attack time = 330 ms
0.16
0.33
0.49
0.65
0.82
Time (sec)

Figure 6

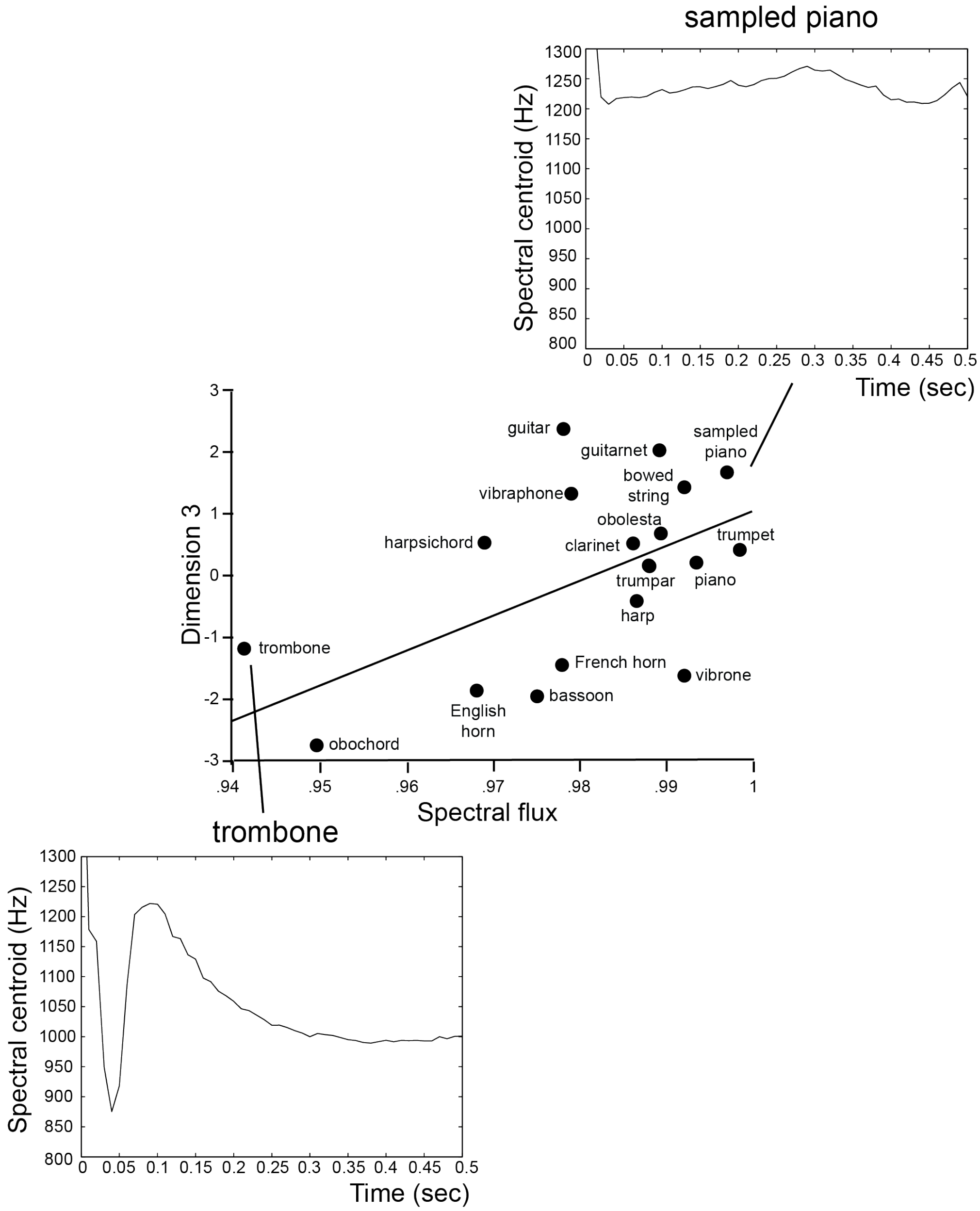

sampled piano
Spectral centroid (Hz)
Time (sec)
Dimension 3
Spectral flux
guitar
guitarnet
sampled piano
bowed string
vibraphone
obolesta
harpsichord
clarinet
trumpet
trumpar
piano
harp
trombone
French horn
vibrone
bassoon
English horn
obochord
trombone
Spectral centroid (Hz)
Time (sec)

Figure 7

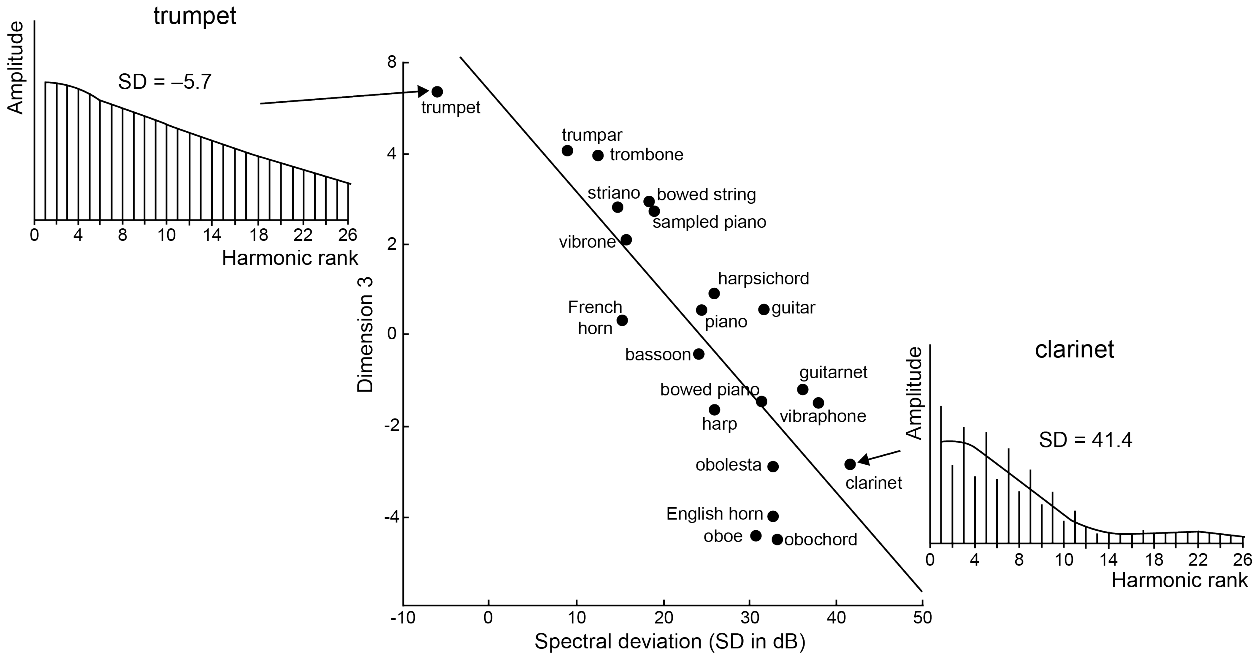
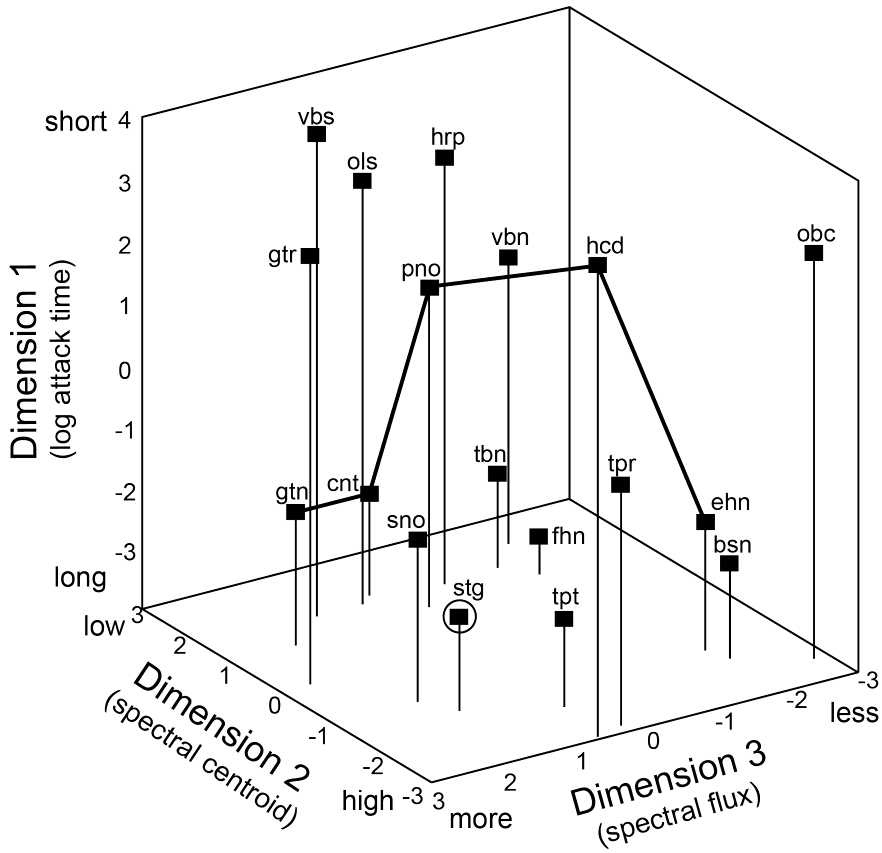

trumpet
Amplitude
SD = –5.7
0 4 8 10 14 18 22 26
Harmonic rank
trumpet
trumpar
trombone
striano
bowed string
sampled piano
vibrone
harpsichord
French horn
piano
guitar
bassoon
guitarnet
bowed piano
vibraphone
harp
obolesta
clarinet
English horn
oboe
obochord
Dimension 3
Spectral deviation (SD in dB)
clarinet
Amplitude
SD = 41.4
Harmonic rank

Figure 8



short
long
Dimension 1
(log attack time)
4
3
2
1
0
-1
-2
-3
low
high
Dimension 2
(spectral centroid)
more
less
Dimension 3
(spectral flux)
vbs
ols
hrp
gtr
pno
vbn
hcd
obc
gtn
cnt
tbn
tpr
ehn
sno
fhn
bsn
stg
tpt

Figure 9

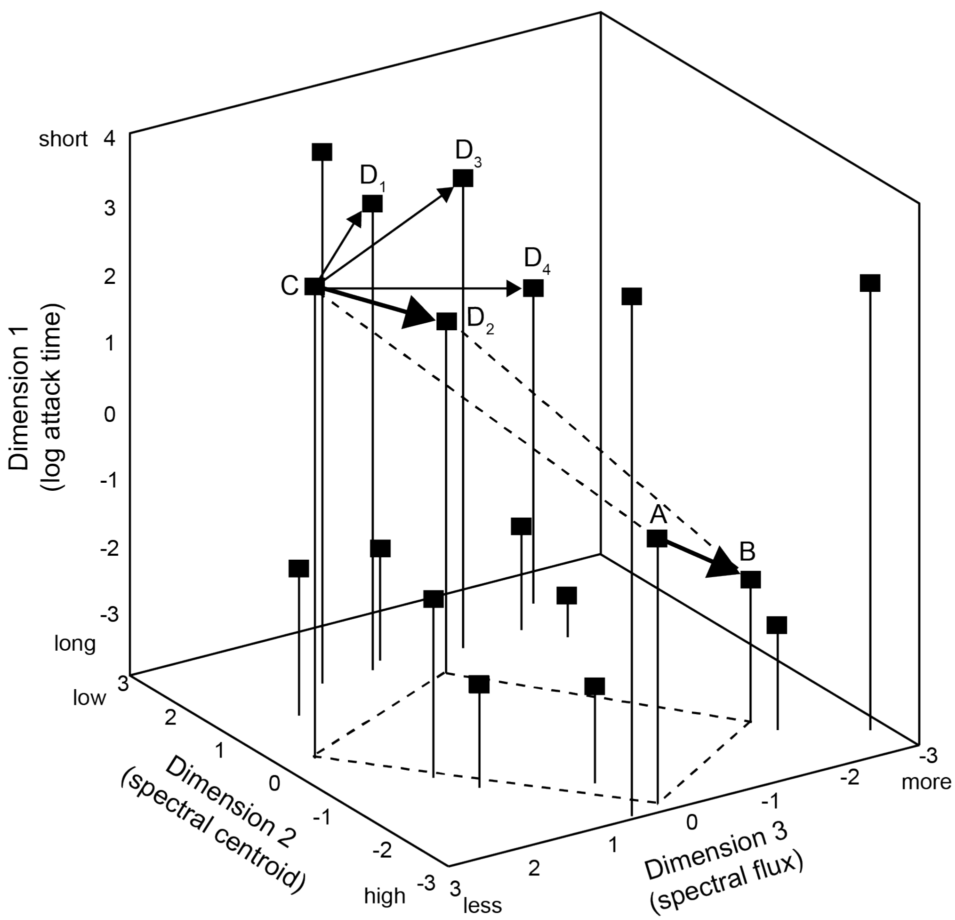

short
4
3
2
1
0
-1
-2
-3
long
Dimension 1
(log attack time)
D1
D3
D4
C
D2
A
B
3
low
2
1
0
-1
-2
-3
high
Dimension 2
(spectral centroid)
3
less
2
1
0
-1
-2
-3
more
Dimension 3
(spectral flux)

Figure 10

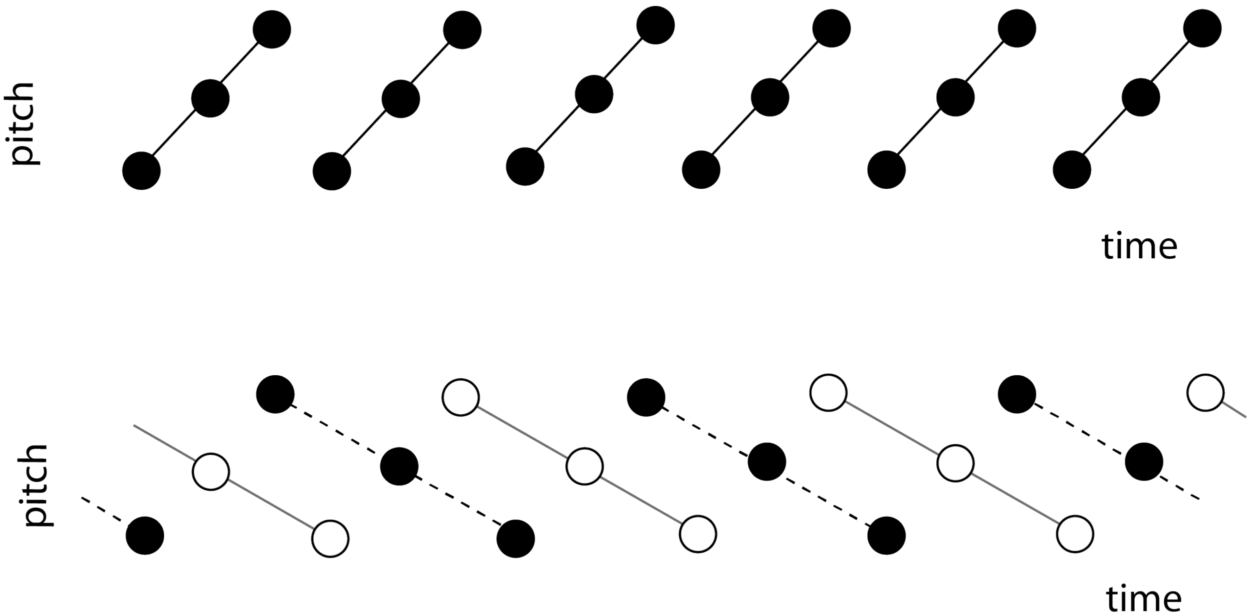


Figure 11

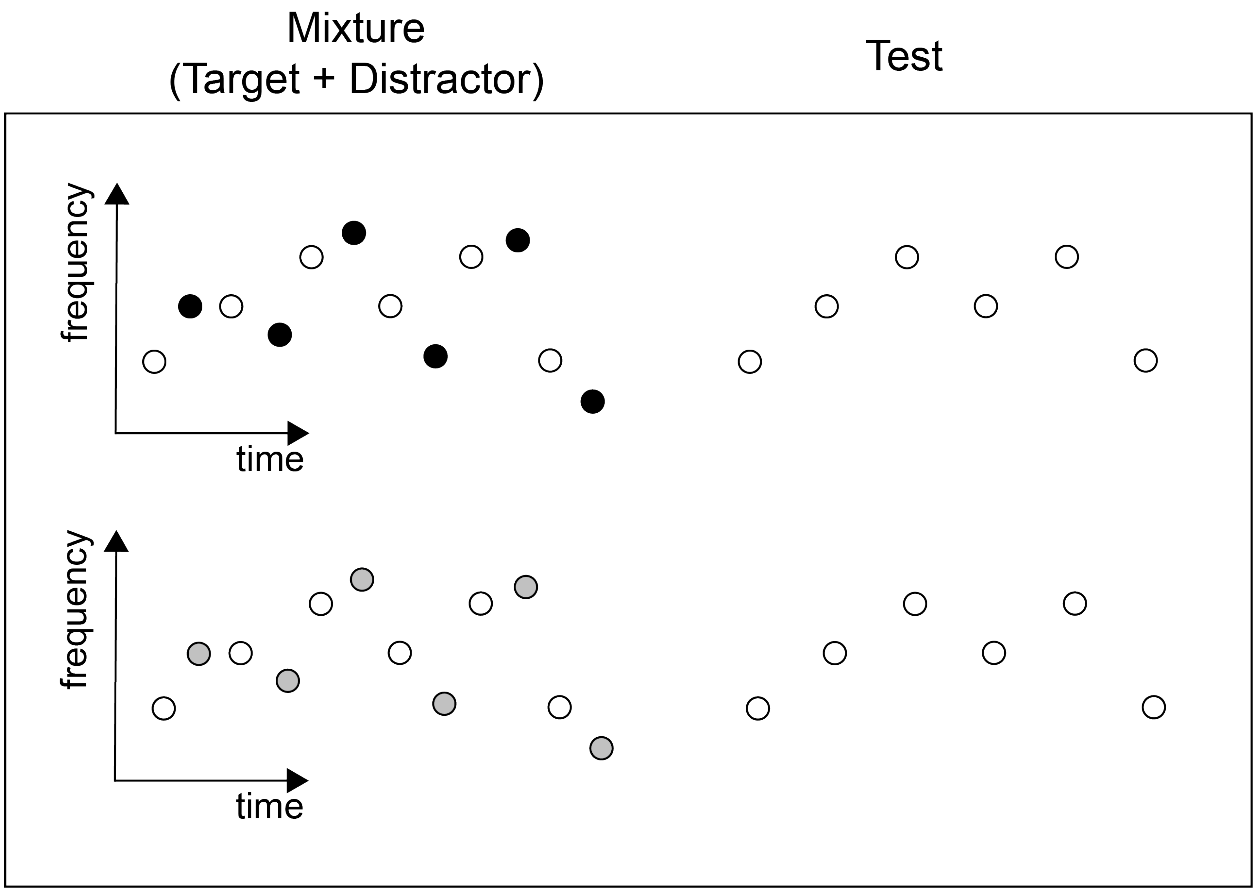

Mixture
(Target + Distractor)
Test
frequency
time
frequency
time

Figure 12

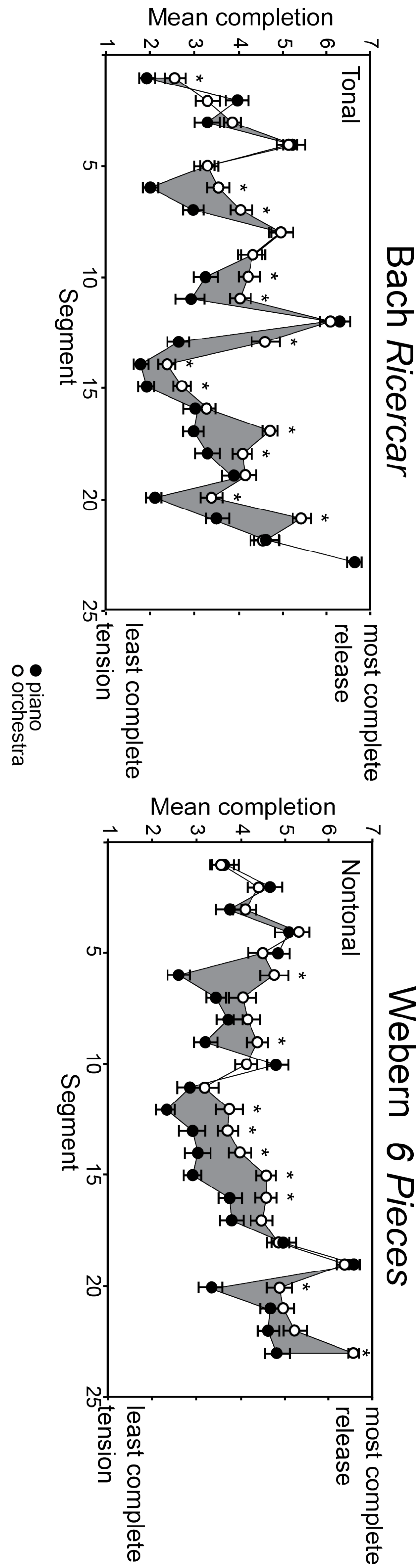

Bach Ricercar
Tonal
Mean completion
Segment
most complete release
least complete tension
● piano
o orchestra
Webern 6 Pieces
Nontonal
Mean completion
Segment
most complete release
least complete tension

Figure 13

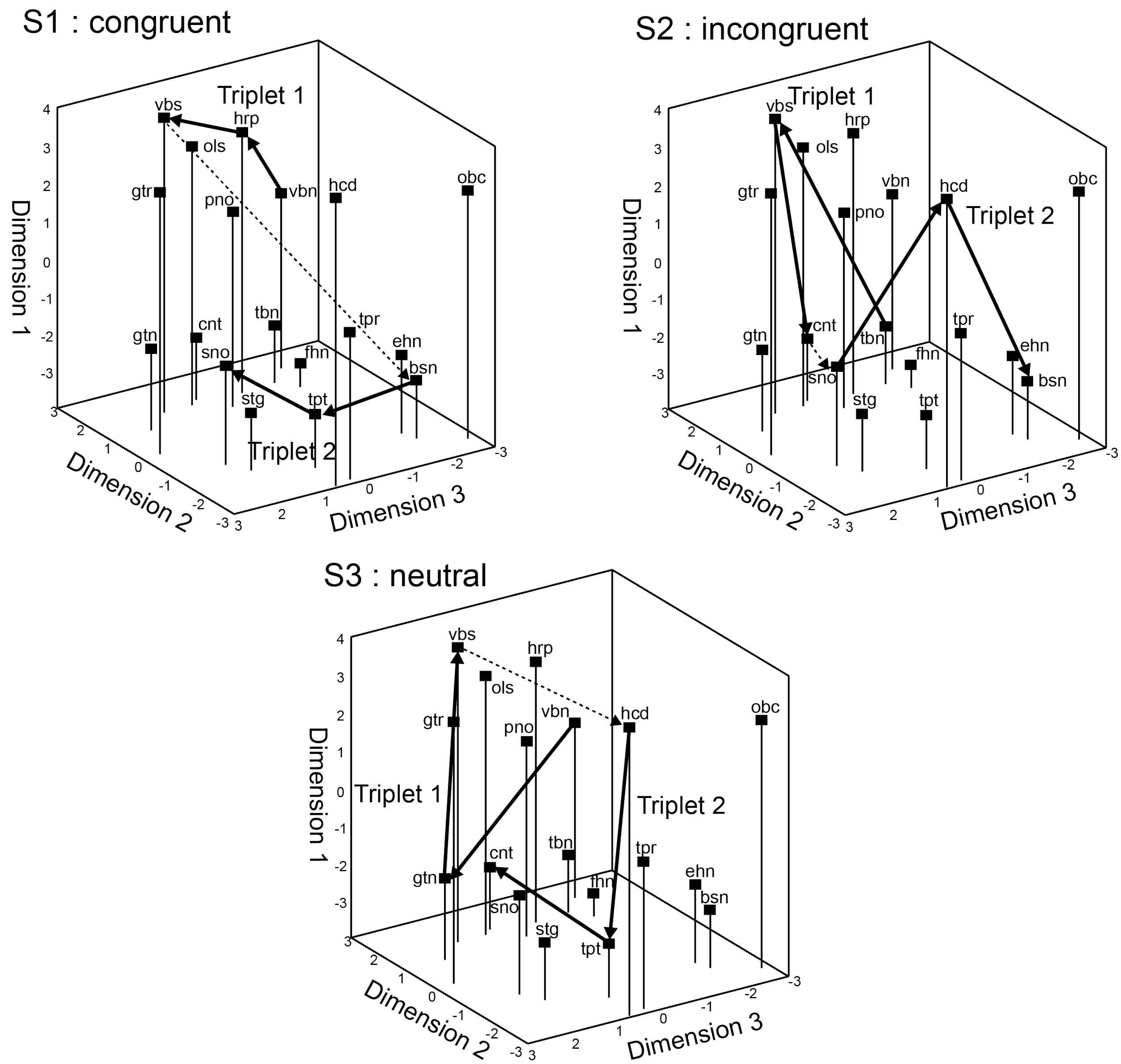
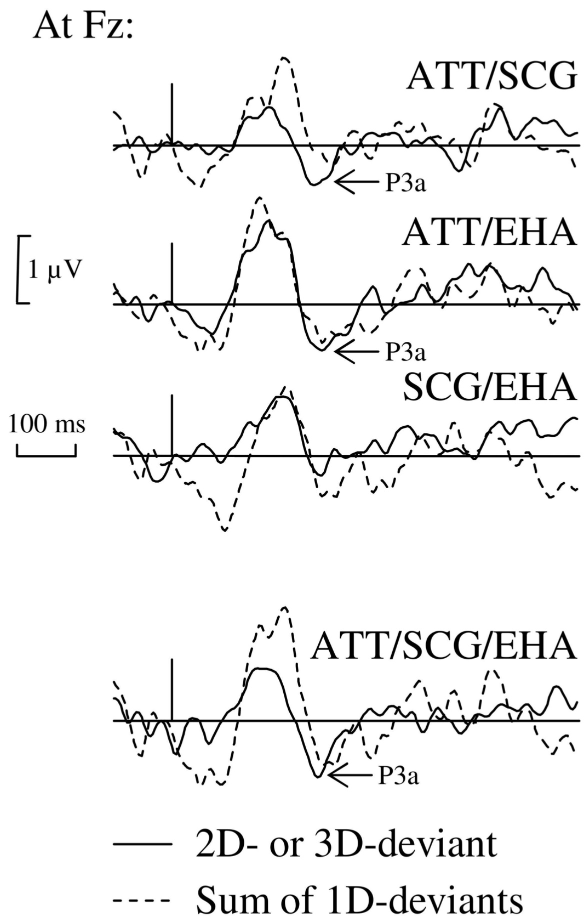

S1 : congruent
Triplet 1
Triplet 2
vbs
hrp
ols
vbn
hcd
obc
gtr
pno
gtn
cnt
tbn
tpr
ehn
fhn
bsn
sno
stg
tpt
Dimension 1
Dimension 2
Dimension 3
S2 : incongruent
S3 : neutral

Figure 14


At Fz:
ATT/SCG
P3a
ATT/EHA
1 µV
P3a
SCG/EHA
100 ms
ATT/SCG/EHA
P3a
2D- or 3D-deviant
Sum of 1D-deviants