CHAPTER 7

The perception of musical timbre[1]

Stephen McAdams and Bruno L. Giordano

TIMBRE is a misleadingly simple and vague word encompassing a very complex set of auditory attributes, as well as a plethora of psychological and musical issues. It covers many parameters of perception that are not accounted for by pitch, loudness, spatial position, duration, and various environmental characteristics such as room reverberation. This leaves a wealth of possibilities that have been explored over the last 40 years or so. We now understand timbre to have two broad characteristics that contribute to the perception of music: (1) it is a multifarious set of abstract sensory attributes, some of which are continuously varying (e.g. attack sharpness, brightness, nasality, richness), others of which are discrete or categorical (e.g., the 'blatt' at the beginning of a sforzando trombone sound or the pinched offset of a harpsichord sound), and (2) it is one of the primary perceptual vehicles for the recognition, identification, and tracking over time of a sound source (singer's voice, clarinet, set of carillon bells), and thus involves the absolute categorization of a sound (Hajda *et al.* 1997; McAdams 1993; Risset and Wessel 1999). The psychological approach to timbre has also included work on the musical implications of timbre as a set of form-bearing dimensions in music (McAdams 1989).

Timbre as a multidimensional set of auditory attributes

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

Timbre space

Multidimensional scaling makes no preconceptions about the physical or perceptual structure of timbre. Listeners simply rate on a scale varying from very similar to very dissimilar all pairs from a given set of sounds. The sounds are usually equalized in terms of pitch, loudness, and duration so that only the timbre varies in order to focus listeners' attention on this attribute. The dissimilarity ratings are then fit to a distance model in which sounds with similar timbres are close together and those with dissimilar

[1] This chapter is an updated version of two previous summaries of musical timbre research (McAdams, 1999, 2003).

timbres are far apart. The graphic representation is called a 'timbre space'. The basic model is expressed in terms of continuous dimensions that are shared among the timbres, the underlying assumption being that all listeners use the same perceptual dimensions to compare the timbres. More complex models also include dimensions or features that are specific to individual timbres (called 'specificities') and different perceptual weights accorded to the dimensions and specificities by individual listeners or classes of listeners (Grey 1977; McAdams *et al.* 1995). Such techniques have been applied to synthetic sounds (Miller and Carterette 1975; Plomp 1970), resynthesized or simulated instrument sounds (Grey 1977; Krumhansl 1989; McAdams *et al.* 1995; Wessel 1979), recorded instrument sounds (Iverson and Krumhansl 1993; Lakatos 2000), and even dyads of recorded instrument sounds (Kendall and Carterette 1991).

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. This might appear as a strong specificity in the distance model (Krumhansl 1989; McAdams *et al.* 1995).

The models integrate individual and class differences as weighting factors on the different dimensions and the set of specificities. Some listeners pay more attention to spectral properties and ignore temporal aspects, whereas others have the inverse pattern. Such variability may reflect either differences in sensory processing or in listening and rating strategies. Interestingly, no study to date has demonstrated that such individual differences have anything to do with musical experience or training (McAdams *et al.* 1995). 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.

Acoustic correlates 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 and Gordon 1978; Krimphoff *et al.* 1994). The most ubiquitous correlates derived from musical instrument sounds include spectral centroid (representing the relative weights of high and low frequencies and corresponding to timbral brightness: an oboe has a higher spectral centroid than a French horn), the logarithm of the attack time (distinguishing continuant instruments that are blown or bowed from impulsive instruments that are struck or plucked), 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), and spectral irregularity (the degree of jaggedness of the spectral shape, which is high for clarinet and vibraphone and low for trumpet). A confirmatory study employing dissimilarity ratings on purely synthetic sounds in which the exact nature of the stimulus dimensions could be controlled was performed by Caclin *et al.* (2005). These authors confirmed the perception of stimulus dimensions related to spectral centroid, attack time and spectral irregularity, but did not confirm spectral flux.

The combination of a quantitative model of perceptual relations among timbres and the psychophysical explanation of the parameters of the model is an important step in gaining predictive control of timbre in several domains such as sound analysis and synthesis and intelligent search in sound databases (Peeters *et al.* 2000). Such representations are only useful to the extent that they are: (a) generalizable beyond the set of sounds actually studied, (b) robust with respect to changes in musical context, and (c) generalizable to other kinds of listening tasks than those used to construct the model. To the degree that a representation has these properties, it may be considered as an accurate account of musical timbre, characterized by an important feature of a scientific model, the ability to predict new empirical phenomena.

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

Effects of pitch change on timbre relations

Marozeau and colleagues (2003) have shown that timbre spaces for recorded musical instrument tones are similar at different pitches (B3, C#4, B♭4). Listeners are also able to ignore pitch differences within an octave when asked to compare only the timbres of the tones. When the pitch variation is greater than an octave, interactions between the two attributes occur. Marozeau and de Cheveigné (2007) varied the brightness of a set of synthesized sounds, while also varying the pitch over a range of 18 semitones. They found that differences in pitch affected timbre relations in two ways:

1 Pitch shows up in the timbre space representation as a dimension orthogonal to the timbre dimensions (indicating simply that listeners were no longer ignoring the pitch difference), and
2 Pitch differences systematically affect the timbre dimension related to spectral centroid.

These results suggest a close relation between timbral brightness and pitch height. This link would be consistent with underlying neural representations that share common attributes, such as a tonotopic organization.

Timbre as a vehicle for source identity

The second approach to timbre concerns its role in the recognition of the identity of a musical instrument or, in general, of a sound-generating event. 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 (McAdams 1993; Handel 1995).

Research on musical instrument identification is relevant to this issue. Saldanha and Corso (1964) studied identification of isolated musical instrument sounds from the Western orchestra played with and without vibrato. They were interested in the relative importance of onset and offset transients, spectral envelope of the sustain portion of the sound, and vibrato. Identification of isolated sounds is surprisingly poor for some instruments. When attacks and decays were excised, identification decreased markedly for some instruments, particularly for the attack portion in sounds without vibrato. However, when vibrato was present, the effect of cutting the attack was less, identification being better. These results suggest that important information for instrument identification is present in the attack portion, but that in the absence of this information, additional information is still available in the sustain portion (although it is more important for some instruments than others), particularly when vibrato is present. The vibrato may increase our ability to extract information relative to the resonance structure of the instrument (McAdams and Rodet 1988).

Giordano (2005) analysed previously published data on the identification and dissimilarity ratings of musical 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-reed instrument, and the bassoon, a double-reed instrument). Consistently, the vast majority of previously published timbre spaces revealed that tones generated with similar resonating structures (e.g., string instruments vs wind instruments) or with similar excitation mechanisms (e.g., impulsive excitation as in piano tones vs sustained excitation as in flute tones) occupied the same region in the space. These results suggest that listeners can reliably identify large differences in the mechanisms of tone production, focusing on the timbre attributes used to evaluate the dissimilarity of musical sounds.

Several investigations on the perception of everyday sounds extend the concept of timbre beyond the musical context (see McAdams 1993; Handel 1995; Lutfi 2008 for reviews). Among them, studies on impact sounds provide information on the timbre attributes useful to the perception of the properties of percussion instruments: bar geometry (Lakatos *et al.* 1997), bar material (McAdams *et al.* 2004a), plate material (Giordano and McAdams 2006), and mallet hardness (Freed 1990). The timbral factors relevant to perceptual judgements vary with the task at hand. Spectral factors are primary for the perception of the geometry (Lakatos *et al.* 1997). Spectrotemporal factors (e.g., the rate of temporal change of the spectral centroid and of loudness) dominate the perception of the material of struck objects (McAdams *et al.* 2004a, Giordano and McAdams 2006) and of mallets (Freed 1990).

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

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

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

Timbral intervals

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

Ehresman and Wessel (1978; Wessel 1979) took a first step forward in this direction, developing a task in which listeners were asked to make judgements on the similarity of intervals formed between pairs of timbres. The basic idea was that timbre intervals may have properties similar to pitch intervals; that is, a pitch interval is a relation along a well-ordered dimension that retains a degree of invariance under certain kinds of transformation, such as translation along the dimension, or what musicians call 'transposition'. 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) 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 relative perceptual relation and thus embody the same timbre interval. Transposition thus consists of translating the vector anywhere else in the space as long as its length and orientation are preserved.

Ehresman and Wessel tested this hypothesis using a task in which listeners had to compare two timbral intervals (e.g. A–B vs C–D) and rank various timbre D's according to how well they fulfilled the analogy: timbre A is to timbre B as timbre C is to timbre D. They essentially found that the closer timbre D was to the ideal point defined by the vector model in timbre space (i.e. the ideal C–D vector was a simple translation of the A–B vector), the higher the ranking.

McAdams and Cunibile (1992) subsequently tested the vector model using the 3D space from Krumhansl (1989) (ignoring the specificities). Five sets of timbres at different places in timbre space were chosen for each comparison to test for the generality of the results. Both electroacoustic composers and non-musicians 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 globally the data support the vector model, although this support was much stronger for composers than for non-musicians. However, when one examines in detail the five different versions of each comparison type, it is clear that not all timbre comparisons go in the direction of the model predictions.

One confounding factor is that the specificities on some timbres in this set were ignored. These, quite to the contrary, would necessarily distort the vectors that were used to choose the timbres, because the specificities 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 would be seriously distorted and thus not perceived as similar to other intervals due to moderate or high specificity values. What this line of reasoning suggests is that the use of 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.

Timbre and musical grouping

An important way in which timbre can contribute to the organization of musical structure is related to the fact that listeners tend to connect perceptually sound events that arise from the same sound source. In general, a given source will produce sounds that are relatively similar in pitch, loudness, timbre and spatial position from one event to the next (cf. Bregman 1990; McAdams and 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 and Campbell 1971). One guiding principle that seems to operate in the formation of auditory streams is the following: successive events that are relatively similar in their spectrotemporal properties (i.e. in their timbres) may have arisen from the same source and should be grouped together; individual sources do not tend to change their acoustic properties suddenly and repeatedly from one event to the next. Early demonstrations of auditory streaming on the basis of timbre (Wessel 1979) suggest a link between the timbre–space representation and the tendency for auditory streaming on the basis of the spectral differences that were created (McAdams and Bregman 1979). Early researchers were convinced that it was primarily the spectral aspects of timbre (such as spectral centroid) that were responsible for auditory streaming and that temporal aspects (such as attack time) had little effect (Hartmann and Johnson 1991).

Recently the picture has changed significantly and several studies indicate an important role for both spectral and temporal attributes of timbre in auditory stream segregation (Moore and Gockel 2002). Iverson (1995) used sequences alternating between two recorded instrument tones with the same pitch and loudness and asked listeners to judge the degree of segregation. Multidimensional scaling of the segregation judgments treated as a measure of dissimilarity was performed to determine which acoustic attributes contributed to the impression of auditory stream segregation. A comparison with previous timbre–space work using the same sounds (Iverson and Krumhansl 1993) showed that both static acoustic cues (such as spectral centroid) and dynamic acoustic cues (such as attack time and spectral flux) were implicated in segregation. Other results consistent with this study have also been reported (Bey and McAdams 2003; Singh and Bregman 1997).

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

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. Instrumental combinations can give rise to new timbres if the sounds are perceived as blended, and timbre can play a role in creating and releasing musical tension.

Timbral blend

The creation of new timbres through orchestration necessarily depends on the degree to which the constituent sound sources fuse together or blend to create the newly emerged sound (Brant 1971; Erickson 1975). Sandell (1995) has proposed that there are three classes of perceptual goals in combining instruments:

1. timbral heterogeneity in which one seeks to keep the instruments perceptually distinct,
2. timbral augmentation in which one instrument embellishes another one that perceptually dominates the combination, and
3. timbral emergence in which a new sound results that is identified as none of its constituents.

Blend appears to depend on a number of acoustic factors such as onset synchrony of the constituent sounds and others that are more directly related to timbre, such as the similarity of the attacks, the difference in the spectral centroids, and the overall centroid of the combination.

Role of timbre in building and release of musical tension

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

As a first step toward understanding how this operates in music, Paraskeva and McAdams (1997) measured the inflection of musical tension and relaxation due to timbral change. Listeners were asked to make judgments on a 7-point scale concerning the perceived degree of completion of the music at several points at which the music stopped. What results is a completion profile, which can be used to infer musical tension by equating completion with release and lack of completion with tension. Two pieces were tested: a fragment from the Ricercar from the *Musical Offering* for six voices by Bach (tonal) and the first movement of the *Six Pieces for Orchestra* by Webern (non-tonal). Each piece was played in an orchestral version (the Webern instrumentation was used for the Bach), and a direct transcription of this orchestral version for piano on a digital sampler. There were significant differences between the piano and orchestral versions, indicating a significant effect of timbre change on perceived musical tension. However, when they were significantly different, the orchestral version was always more relaxed than the piano version.

The hypothesis advanced by Paraskeva and McAdams (1997) for this effect was that the higher relaxation of the orchestral version might have been due to processes involved in auditory stream formation and the dependence of perceived roughness on the results of such processes (Wright and Bregman 1987). Roughness, or any other auditory attribute of a single sound event, is computed after auditory organization processes have grouped the bits of acoustic information together. Piano sounds have a rather sharp attack. If several notes occur at the same time in the score and are played with a piano sound, they will be quite synchronous. Because they all start at the same time and have similar amplitude envelopes, they will tend to be fused together and the computed roughness will result from the interactions of all the frequency components of all the notes.

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

Conclusion

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. The timbre–space representation is a powerful psychological model that allows predictions to be made about timbre perception in situations beyond those used to derive the model in the first place. 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.

Timbre can play a role in larger-scale movements of tension and relaxation and thus contribute to the expression inherent in musical form. Under conditions of high blend among instruments composing a vertical sonority, timbral roughness is a major component of musical tension. However, it strongly depends on the way auditory grouping processes have parsed the incoming acoustic information into events and streams.

References

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

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

Brant H (1971). Orchestration. In J Vinton, ed., *Dictionary of contemporary music*, pp. 538–546 EP Dutton, New York.

Bregman AS (1990). *Auditory scene analysis: the perceptual organization of sound*. MIT Press, Cambridge, MA.

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

Caclin A, McAdams S, Smith BK and 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.

Culling JF and Darwin CJ (1993). The role of timbre in the segregation of simultaneous voices with intersecting Fo contours. *Perception and Psychophysics*, **34**, 303–309.

Deliège I (1987). Grouping conditions in listening to music: an approach to Lerdahl and 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 and Wessel DL (1978). Perception of timbral analogies. *Rapports de l'IRCAM*, vol. 13 IRCAM, Paris.

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

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

Giordano BL (2005). Sound source perception in impact sounds. Unpublished doctoral dissertation. University of Padova, Italy.

Giordano BL and 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.

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

Gregory AH (1994). Timbre and auditory streaming. *Music Perception*, **12**, 161–174.

Grey JM (1977). Multidimensional perceptual scaling of musical timbres. *Journal of the Acoustical Society of America*, **61**, 1270–1277.

Grey JM and Gordon JW (1978). Perceptual effects of spectral modifications on musical timbres. *Journal of the Acoustical Society of America*, **63**, 1493–1500.

Hajda JM, Kendall RA, Carterette EC and Harshberger ML (1997). Methodological issues in timbre research. In I Deliège and J Sloboda, eds, *Perception and cognition of music*, pp. 253–306. Psychology Press, Hove.

Handel S (1995). Timbre perception and auditory object identification. In B Moore, ed., *Hearing*, pp. 425–462. Academic Press, San Diego, CA.

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

Hartmann WM and Johnson D (1991). Stream segregation and peripheral channeling. *Music Perception*, **9**, 155–184.

Helmholtz HLF von (1885/1954). *On the sensations of tone as a physiological basis for the theory of music*. New York, from 1877 trans. by AJ Ellis of 4th German edn, republished 1954 by Dover, New York.

Iverson P (1995). Auditory stream segregation by musical timbre: effects of static and dynamic acoustic attributes. *Journal of Experimental Psychology: Human Perception and Performance*, **21**, 751–763.

Iverson P and Krumhansl CL (1993). Isolating the dynamic attributes of musical timbre. *Journal of the Acoustical Society of America*, **94**, 2595–2603.

Kendall RA and Carterette EC (1991). Perceptual scaling of simultaneous wind instrument timbres. *Music Perception*, **8**, 369–404.

Krimphoff J, McAdams S and Winsberg S (1994). Caractérisation du timbre des sons complexes. II: Analyses acoustiques et quantification psychophysique [Characterization of the timbre of complex sounds. II: Acoustic analyses and psychophysical quantification]. *Journal de Physique*, **4**(C5), 625–628.

Krumhansl CL (1989). Why is musical timbre so hard to understand? In S Nielzén and O Olsson, eds, *Structure and perception of electroacoustic sound and music*, pp. 43–53. Excerpta Medica, Amsterdam.

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

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

Lerdahl F and Jackendoff R (1983). T*he generative theory of tonal music*. MIT Press, Cambridge, MA.

Lutfi R (2008). Human sound source identification. In W Yost, A Popper and R Fay, eds, *Auditory perception of sound sources*, pp. 13–42. Springer, New York.

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

Marozeau J and 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 and E Bigand, eds, *Thinking in sound: the cognitive psychology of human audition*, pp. 146–198. Oxford University Press, Oxford.

McAdams S (1999). Perspectives on the contribution of timbre to musical structure. *Computer Music Journal*, **23**(2), 96–113.

McAdams S (2003). Perception of musical timbre. *Bulletin of Psychology and the Arts*, **4**, 39–42.

McAdams S and Bregman AS (1979). Hearing musical streams. *Computer Music Journal*, **3**(4), 26–43.

McAdams S and Cunibile JC (1992). Perception of timbral analogies. *Philosophical Transactions of the Royal Society, London*, series B, **336**, 383–389.

McAdams S and Rodet X (1988) The role of FM-induced AM in dynamic spectral profile analysis. In H Duifhuis, JW Horst, HP Wit, eds, *Basic issues in hearing*, pp. 359–369. Academic Press, London.
McAdams S, Chaigne A and 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 and Clarke E (2004b). Analyzing musical sound. In E Clarke and N Cook, eds, *Empirical musicology: aims, methods, prospects*, pp. 157–196. Oxford University Press, New York.
McAdams S, Winsberg S, Donnadieu S, De Soete G and Krimphoff J (1995). Perceptual scaling of synthesized musical timbres: common dimensions, specificities, and latent subject classes. *Psychological Research*, **58**, 177–192.
Miller JR and Carterette EC (1975). Perceptual space for musical structures. *Journal of the Acoustical Society of America*, **58**, 711–720.
Moore BCJ and Gockel H (2002). Factors influencing sequential stream segregation. *Acustica united with Acta Acustica*, **88**, 320–332.
Paraskeva S and 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. International Computer Music Association, San Francisco, CA.
Parncutt R (1989). *Harmony: a psychoacoustical approach*. Springer-Verlag, Berlin.
Peeters G, McAdams S and Herrera P (2000) Instrument sound description in the context of MPEG-7. *Proceedings of the 2000 International Computer Music Conference, Berlin*, pp. 166–169. International Computer Music Association, San Francisco, CA.
Plomp R (1970). Timbre as a multidimensional attribute of complex tones. In R Plomp and GF Smoorenburg, eds, *Frequency analysis and periodicity detection in hearing*, pp. 397–414. Sijthoff, Leiden.
Plomp R (1976). *Aspects of tone sensation: a psychophysical study*. Academic Press, London.
Risset J-C and Wessel DL (1999). Exploration of timbre by analysis and synthesis. In D Deutsch, ed., *The psychology of music*, 2nd edn, pp. 113–168. Academic Press, San Diego, CA.
Saldanha EL and Corso JF (1964). Timbre cues and the identification of musical instruments. *Journal of the Acoustical Society of America*, **36**, 2021–2126.
Sandell GJ (1995). Roles for spectral centroid and other factors in determining ‘blended’ instrument pairings in orchestration. *Music Perception*, **13**, 209–246.
Schoenberg A (1911/1978). *Theory of harmony*. Original German publication, 1911 edn. University of California Press, Berkeley, CA.
Singh PG and Bregman AS (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*. University of California Press, Berkeley, CA.
Steele K and Williams A (2006). Is the bandwidth for timbre invariance only one octave? *Music Perception*, **23**, 215–220.
Wessel DL (1973). Psychoacoustics and music: a report from Michigan State University. *PACE: Bulletin of the Computer Arts Society*, **30**, 1–2.
Wessel DL (1979). Timbre space as a musical control structure. *Computer Music Journal*, **3**(2), 45–52.
Wright JK and Bregman AS (1987). Auditory stream segregation and the control of dissonance in polyphonic music. *Contemporary Music Review*, **2**(1), 63–92.