In R. Ashley & R. Timmers (Eds.), T*he Routledge Companion to Music Cognition*, Routledge, New York, NY (2017).

11
MUSICAL STRUCTURE
Sound and Timbre

Stephen McAdams and Meghan Goodchild

Introduction

We define timbre as a set of auditory attributes—in addition to those of pitch, loudness, duration, and spatial position—that both carry musical qualities, and collectively contribute to sound source recognition and identification. Timbral properties can arise from an event produced by a single sound source, whether acoustic or electroacoustic, but they can also arise 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.

It is important at the outset to pinpoint a major misuse of this word, that is, referring to the "timbre" of a given instrument: the timbre of the clarinet, for instance. This formulation confuses source identification with the kinds of perceptual information that give rise to that identification. Indeed, a specific clarinet 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. Change any of these parameters and the timbre will change. Therefore, in our conception of timbre, an instrument such as a clarinet does not have "a timbre," but rather it has a constrained universe of timbres that co-vary with the musical parameters listed above to a greater or lesser extent depending on the instrument and the parameter(s) being varied. For example, a French horn player can make the sound darker by playing a bit softer, and the timbre of clarinet sounds is vastly different in the lower chalumeau register than in the higher clarion register. That being said, as we will see below, there may 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. It is associated with a plethora of psychological and musical issues concerning its role as a form-bearing element in music (McAdams, 1989). The issues that will be addressed in this chapter include the perceptual and acoustic characterization of timbre, its role in the identification of sound sources and events, the perception of sequential timbral relations, timbre's dependence on concurrent grouping, its role in sequential and segmental grouping, and its contribution to musical structuring.

Acoustic and Psychophysical Characterization of Musical Sounds

Timbre perception depends on acoustic properties of sounds and how these properties are represented in the auditory system. These acoustic properties, in turn, depend on the mechanical properties of vibrating objects in the case of acoustic instruments, or the properties of electronic circuits, digital algorithms, amplifiers, and sound reproduction systems in the case of electroacoustic sounds. It thus behooves us to understand the perceptual structure of timbre and its acoustic and mechanical underpinnings.

The notion of timbre encompasses many properties such as auditory brightness, roughness, attack quality, richness, hollowness, inharmonicity and so on. One primary approach to revealing and modeling the complex perceptual representation of timbre is through multidimensional scaling (MDS) analyses of subjective ratings of how dissimilar sounds are from one another. All pairs of sounds in a set are compared, giving rise to a matrix of dissimilarities, which are then analyzed by an algorithm that fits the dissimilarities to a distance model with a certain number of dimensions (see McAdams, 1993, for more details). Studies with synthesized sounds (Grey, 1977; Krumhansl, 1989; McAdams et al., 1995; Wessel, 1979) and recorded sounds (Iverson & Krumhansl, 1993; Lakatos, 2000) generally find "timbre spaces" with two or three dimensions. This low dimensionality, compared with all of the ways the timbre could vary between sounds in a given set, suggests limits either in listeners' abilities to form ratings based on a large number of dimensions, or in the algorithms' abilities to reliably distinguish high-dimensional structures given the inter-individual variability in the data. Some algorithms also allow for specific dimensions or discrete features on individual sounds (called "specificities") and for different weights on the various dimensions and specificities for individual listeners or groups of listeners (McAdams et al., 1995). Specificities capture unique qualities of a sound that are not shared with other sounds and that make it dissimilar with respect to them. The weights on the dimensions reflect listeners' differing sensitivities to the individual dimensions and specificities. The perceptual dimensions correlate most often with acoustic descriptors that are temporal (e.g., attack qualities), spectral (e.g., timbral brightness) or spectrotemporal (e.g., timbral variation over the duration of a tone); but, the acoustic nature of the perceptual dimensions of a given space depends on the stimulus set: different properties emerge for a set of wind and string tones than for percussive sounds, for example (Lakatos, 2000).

The interpretive challenge of timbre spaces is to determine whether they have an underlying acoustic basis. This approach presumes that individual perceptual dimensions would have independent (orthogonal) acoustic correlates. A profusion of quantitative acoustic descriptors derived directly from the acoustic signal or from models of its processing by peripheral auditory mechanisms has been developed and integrated into MATLAB toolboxes such as the MIR Toolbox (Lartillot & Toiviainen, 2007) and the Timbre Toolbox (Peeters et al., 2011). Authors often pick and choose the descriptors that seem most relevant, such as spectral centroid (related to timbral brightness or nasality), attack time of the energy envelope, spectral flux (degree of variation of the spectral envelope over time), and spectral deviation (jaggedness of the spectral fine structure). However, Peeters et al. (2011) computed measures of central tendency and variability over time of the acoustic descriptors in the Timbre Toolbox on a set of over 6,000 musical instrument sounds with different pitches, dynamics, articulations, and playing techniques. They found that many of the descriptors co-vary quite strongly within even such a varied sound set and concluded that there were only about ten classes of independent descriptors. This can make the choice among similar

descriptors somewhat arbitrary. Techniques such as partial least-squares regression allow for an agnostic approach, reducing similar descriptors to a single variable (principal component) that represents the common variation among them.

What has yet to be established is whether these descriptors actually have status as perceptual dimensions: are they organized along ordinal, interval or ratio scales? Another question is whether the descriptors can actually be perceived independently as is suggested by the MDS approach. It is also unclear whether combinations of descriptors may actually form perceptual dimensions through the long-term auditory experience of them as strongly co-varying parameters. One confirmatory study on synthetic sounds has shown that spectral centroid, attack time and spectral deviation do maintain perceptual independence, but that spectral flux collapses in the presence of variation in spectral centroid and attack time (Caclin et al., 2005).

A very different approach is to treat the neural representation of timbre as a monolithic high-dimensional structure rather than as a set of orthogonal dimensions. In a new class of *modulation representations*, sound signals are described not only according to their frequency (tonotopic, arranged by frequency or placement in a collection of neurons) and amplitude variation over time, but include a higher-dimensional topography of the evolution of frequency-specific temporal-envelope profiles. This approach uses modulation power spectra (Elliott et al., 2013) or simulations of cortical spectrotemporal receptive fields (STRF; Shamma, 2000). Sounds are thus described 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 possible models for timbre (Elliott et al., 2013; Patil et al., 2012). However, the predictions of timbre dissimilarity ratings have relied heavily on dimensionality reduction techniques driven 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), essentially yielding difficult-to-interpret black-box approaches, at least 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). This leads to the question of whether timbre is indeed an emergent, high-dimensional spectrotemporal "footprint" or whether it relies on a limited bundle of orthogonal perceptual dimensions.

Most studies have equalized stimuli as much as is feasible in terms of pitch, duration, and loudness in order to focus listeners on timbral differences. However, some studies have included sounds from different instruments at several pitches. They have found that relations among the instruments in the timbre spaces are similar at pitches differing by as much as a major seventh, but that interactions between pitch and timbre appear with sounds differing by more than an octave (Marozeau & de Cheveigné, 2007). Therefore, pitch appears as an orthogonal dimension independent of timbre, and pitch differences systematically affect the timbre dimension that is related to the spectral centroid (auditory brightness)—the center of mass of the frequency spectrum. This result suggests that changes in both pitch height and timbral brightness shift the spectral distribution along the tonotopic axis in the auditory nervous system. That being said, Demany and Semal (1993) found with three highly trained listeners that detection of a change in pitch and timbral brightness was independent, even

when these parameters co-varied, suggesting that the degree of independence may depend on stimulus, task, and training.

Timbre's Contribution to the Identity of Sound Sources and Events

The sensory dimensions making up timbre constitute indicators that collectively contribute to the categorization, recognition, and identification of sound events and sound sources (McAdams, 1993). Studies on musical instrument identification show that important information is present in the attack portion of the sound, but also in the sustain portion, particularly when vibrato is present (Saldanha & Corso, 1964); in fact, the vibrato may better define the resonant structure of the instrument (McAdams & Rodet, 1988). In a meta-analysis on published data derived from instrument identification and dissimilarity rating studies, Giordano and McAdams (2010) found that listeners more often confuse tones generated by musical instruments with a similar mechanical structure (confusing oboe with English horn, both with double reeds) than with sounds generated by very different structures. In a like manner for dissimilarity ratings, sounds resulting from similar resonating structures and/or from similar excitation mechanisms (two struck strings—guitar and harp) occupied the same region in timbre space, whereas those with large differences (a struck bar and a sustained air jet—xylophone and flute) occupied distinct regions. Listeners thus seem able to identify differences in the mechanisms of tone production by using the timbral properties that reliably carry that information. Furthermore, dissimilarity ratings on recorded acoustic instrument sounds and their digital transformations, which maintain a similar acoustic complexity but reduce familiarity, are affected by long-term memory for the familiar acoustic sounds (Siedenburg et al., 2016). A model that includes both acoustic factors and categorical factors such as instrument family best explains the results.

A fascinating problem that has been little studied is how one builds up a model of a sound source whose timbral properties vary significantly with dynamics and pitch. Some evidence indicates that musically untrained listeners can recognize sounds at different pitches as coming from the same instrument only within the range of about an octave (Handel & Erickson, 2001), although musically trained listeners can perform this task fairly well even at differences of about 2.5 octaves (Steele & Williams, 2006). Instrument identification across pitches thus depends on musical training, and it seems that the mental model that represents the timbral covariation with pitch needs to be acquired through experience. An important question concerns how recognition and identification are achieved by information accumulation strategies with such acoustically variable sources. The STRF representation is claimed both to capture source properties of musical instruments that are invariant over pitch and sound level (Shamma, 2000) and to provide a sound source signature that allows very rapid and robust musical source categorization (Agus, et al., 2012). One proposal that emerges from this view is that extracting and attending separately to the individual features or dimensions may require additional processing. These possibilities need further study and could elucidate the confusion in the field between timbre as a vehicle for source identity and timbre as an abstract musical quality.

Perception of Timbral Relations

Timbre space provides a model for relations among timbres. Based on this representation, we can consider theoretically the extension of certain properties of pitch relations and many of

the operations traditionally used on pitch sequences to the realm of timbre. A timbre interval can be considered as a vector in timbre space, and transposing that interval maintains the same amount of change along each perceptual dimension of timbre. One question concerns whether listeners can perceive timbral intervals and recognize transpositions of those intervals to other points in the timbre space as one can perceive pitch intervals and their transpositions in pitch space. McAdams and Cunibile (1992) selected reference pairs of timbres from the space used by Krumhansl (1989), as well as comparison pairs that either respected the interval relation or violated it in terms of the orientation or length of the vector. Listeners were generally better than chance at choosing the correct interval, although electroacoustic composers outperformed nonmusicians. Not all pairs of timbre intervals were as successfully perceived as related, suggesting that factors such as specificities of individual timbres may have distorted the intervals. It may well be difficult to use timbre intervals as an element of musical discourse in a general way in instrumental music given that timbre spaces of acoustic instruments tend both to be unevenly distributed and to possess specificities, unlike the equal spacing of pitches in equal-temperament. However, one should not rule out the possibility in the case of synthesized sounds or blended sounds created through the combination of several instruments. At any rate, whether or not specific intervals are precisely perceived and memorized, work in progress shows that perception of the *direction of change* along the various dimensions is fairly robust, allowing for the perception of similar contours in trajectories through timbre space.

Timbre and Auditory Grouping

Timbre emerges from the perceptual fusion of acoustic components into a single auditory event, including the blending of sounds produced by separate instruments. According to Sandell (1995), the possible perceptual results of instrument combinations include timbral heterogeneity (sounds are segregated and identified), augmentation (subservient sounds are blended into a dominant, identifiable sound), and emergence (all sounds are blended and unidentifiable). There is an inverse relation between degree of blend and identification of the constituent sounds (Kendall & Carterette, 1993). Fusion depends on concurrent grouping cues, such as onset synchrony and harmonicity (McAdams & Bregman, 1979); that is, instruments that play with synchronous onsets and in consonant harmonic relations are more likely to blend. However, the degree of fusion also depends on spectrotemporal relations among the concurrent sounds: Some instrument pairs can still be distinguished in dyads with identical pitches and synchronous onsets because their spectra do not overlap significantly. Generally, sounds blend better when they have similar attacks and spectral centroids, as well as when their composite spectral centroid is lower (Sandell, 1995). When impulsive and sustained sounds are combined, blend is greater for lower spectral centroids and slower attacks, and the timbre resulting from the blend is primarily determined by the attack of the impulsive sound and the spectral envelope of the sustained sound (Tardieu & McAdams, 2012). More work is needed on how to predict blend from the underlying perceptual representation, on the resulting timbral qualia of blended sounds, and on which timbres will remain identifiable in a blend.

Timbre also plays a strong role in determining whether successive sounds are integrated into an auditory stream or segregated into separate streams on the basis of timbral differences that potentially signal the presence of multiple sound sources (McAdams & Bregman, 1979). Larger differences in timbre create stream segregation, and thus timbre strongly affects

what is heard as melody and rhythm, because these perceptual properties of sequences are computed within auditory streams. In sequences with two alternating timbres, the more the timbres are dissimilar, the greater is the resulting degree of segregation into two streams (Iverson, 1995; Bey & McAdams, 2003). The exact representation underlying this sequential organization principle and how it interacts with attentional processes is not yet understood, but it seems to include both spectral and temporal factors that contribute to timbre. Timbral difference is also an important cue for following a voice that crosses other voices in pitch or for hearing out a given voice in a polyphonic texture (McAdams & Bregman, 1979). If a composer seeks to create melodies that change in instrumental timbre from note to note (called *Klangfarbenmelodien* or sound-color melodies by Schoenberg, 1911/1978), timbre-based streaming may prevent the listener from integrating the separate sound sources into a single melody if the changes are too drastic. We have a predisposition to identify a sound source and follow it through time on the basis of continuity in pitch, timbre, loudness, and spatial position. Cases in which such timbral compositions work have successfully used smaller changes in timbre from instrument to instrument, unless pointillistic fragmentation is the desired aim, in which case significant timbre change is effective in inducing perceptual discontinuity.

We propose two other kinds of grouping that are often mentioned in orchestration treatises: textural integration and stratification or layering. Textural integration occurs when two or more instruments featuring contrasting rhythmic figures and pitch material coalesce into a single textural layer. This is perceived as being more than a single instrument, but less than two or more clearly segregated melodic lines. Stratification creates two or more different layers of orchestral material, separated into more and less prominent strands (foreground and background), with one or more instruments in each layer. Integrated textures often occupy an orchestral layer in a middleground or background position. Future research will need to test the hypothesis that it is the timbral similarity within layers, and timbral differences between layers, that allow for the separation of layers and also whether timbral characteristics determine the prominence of a given layer (i.e., more salient timbres or timbral combinations occur more frequently in foreground layers).

Role of Timbre in Musical Structuring

In addition to timbre's involvement in concurrent and sequential grouping processes, timbral discontinuities also promote segmental grouping, a process by which listeners "chunk" musical streams into units such as phrases and themes. Specific evaluation of the role that timbre plays in segmental structuring in real pieces of music is limited in the literature. Repeating timbral patterns and transition probabilities that are learned over sufficient periods of time can also create segmentation of sequences into smaller-scale timbral patterns (Tillmann & McAdams, 2004). Discontinuities in timbre (contrasting instrument changes) can provoke segmentation of longer sequences of notes into smaller groups or of larger-scale sections delimited by significant changes in instrumentation and texture (Deliège, 1989).

We are developing a taxonomy of timbral contrasts that occur frequently in the orchestral repertoire. These contrast types include: 1) antiphonal alternation of instrumental groups in call-and-response phrase structure, 2) timbral echoing in which a repeated musical phrase or idea appears with different orchestrations, with one seeming more distant than the other due to the change in timbre and dynamics, and 3) timbral shifts in which musical materials are reiterated with varying orchestrations, being passed around the orchestra and often

accompanied by motivic elaboration or fragmentation. The perceptual strengths of these different contrasts, as well as the call-response or echo-like relations, depend on the timbral changes used.

Formal functions (e.g., exposition, recapitulation), processes (e.g., repetition, fragmentation) and types (e.g., motives, ideas, sentences, periods, sonata, rondo) have been theorized in Classical music, and there has been some discussion of how they are articulated through orchestration. Cannon (2015) demonstrates that contrasts in dynamics and orchestration (instrument density) are key determinants that influence whether the onset of a recapitulation serves as a resolution, climax, or arrival, on the one hand, or as a new beginning or relaunch, on the other. Dolan (2013) examines Haydn's structural and dramatic use of orchestration, including the process of developing variation. Future work should address how orchestral variations are used to reinforce, vary or even contradict these pitch- and rhythm-based structures and their resulting effect on the listening experience.

There is very little music-theoretical or perceptual research on the topic of large-scale orchestral shaping. However, these orchestral gestures, such as the sudden contrast between the orchestra and a soloist, have been shown to contribute to peak emotional experiences in orchestral music (Guhn et al., 2007). While some orchestration treatises mention certain of these gestures, a major concern is the lack of a clear taxonomy of techniques and of a conceptual framework related to their musical function. Goodchild (2016) developed a typology of orchestral gestures defined by changes in instrumentation based on the time course (gradual or sudden) and direction (additive or reductive) of change. She hypothesized that extended patterns of textural and timbral evolution create orchestral gestures, which possess a certain cohesiveness as auditory images and have a goal-directed sense of motion. These gestures often give rise to strong emotional experiences due to a confluence of change along many timbral dimensions, but also in loudness, tempo, and registral extent, giving them expressive agency. Listeners' continuous ratings of emotional intensity while listening to orchestral excerpts reveal different response profiles for each gestural type, in particular a lingering effect of high emotional intensity for the reductive gestures (Fig. 11.1). Using re-orchestration and digital orchestral rendering as tools for testing hypotheses concerning the role of timbral brightness in emotional valence, Goodchild showed with psychophysiological measures that the brightness of the orchestration (measured as spectral centroid) leading up to an expressive event dramatically shapes the resulting experience.

Another way that timbre contributes to larger-scale musical form by way of orchestration is through the sense of movement between tension and relaxation. Timbre may affect the perception of harmonic tension by influencing the perception of voice leading through sequential integration of notes with similar timbres and segregation of those with different timbres. The competition between fusion and sequential streaming has been argued to affect the perception of dissonance (Wright & Bregman, 1987). Paraskeva & McAdams (1997) asked listeners to make ratings of perceived degree of completion (the inverse of tension) at several points in piano and orchestral versions for both tonal and nontonal works (Bach's *Ricercar* from the *Musical Offering* orchestrated by Webern and the first of Webern's *Six Pieces for Orchestra*, Op. 6). When differences were found in the completion profiles, the orchestral version was consistently less tense than the piano version. This effect may have been due to the processes involved in auditory stream formation, especially the perception of timbral roughness. Roughness is a timbral attribute that results from the beating of proximal frequency components within auditory filters. Roughness largely determines sensory dissonance. As a timbral attribute, it depends on what gets grouped concurrently: If several notes

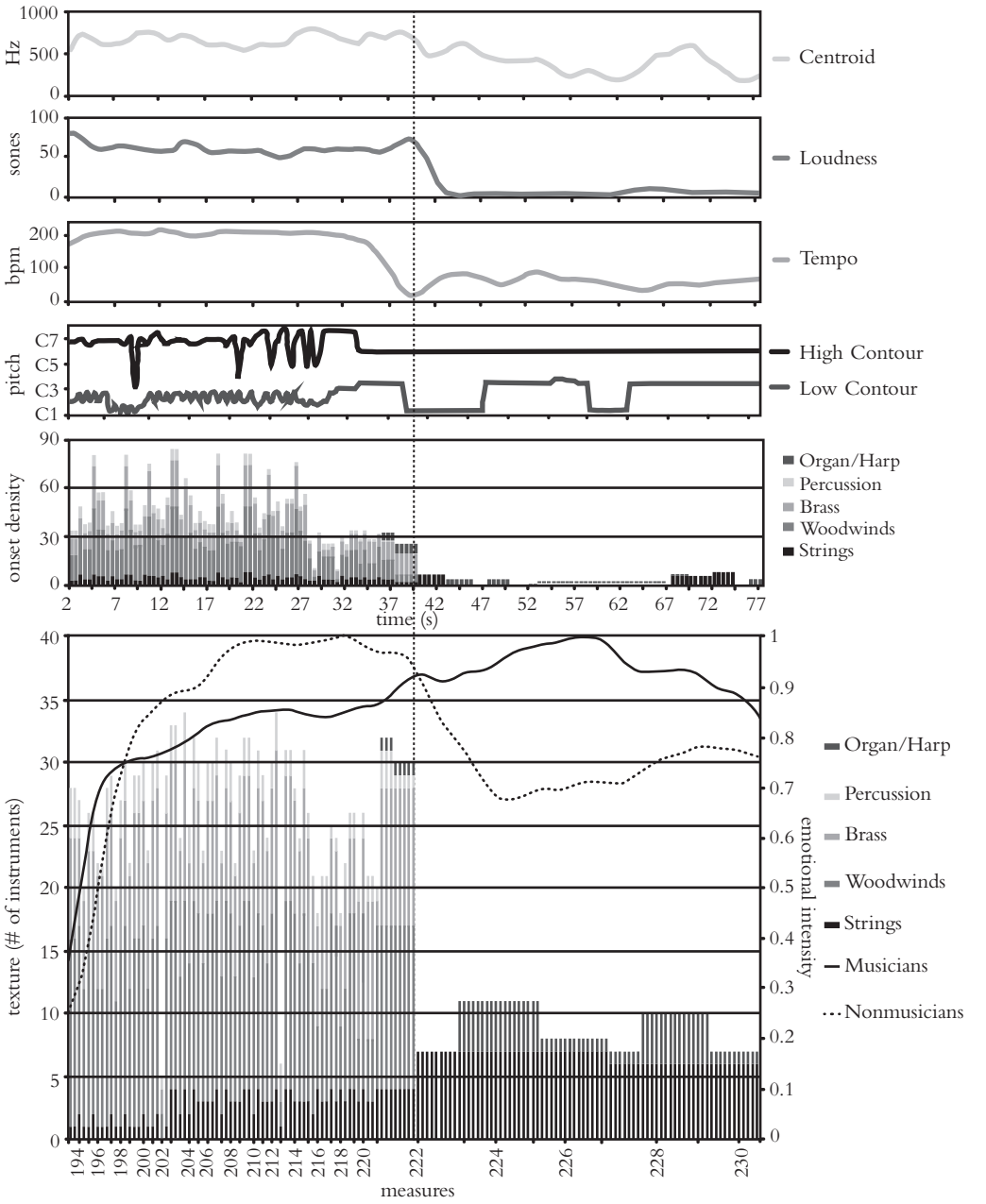

Figure 11.1 Visualization of Holst, *The Planets*, "Uranus," mm. 193–236, with score-based features (instrumental texture, onset density, and melodic contour), performance-based features (loudness, spectral centroid, and tempo), and average emotional intensity ratings for musician and non-musician listeners. See insert for color figure.

From Goodchild (2016), Fig. A.10.

with the impulsive attack of a piano sound occur simultaneously in a vertical sonority with dissonant intervals, they would be synchronous and of similar timbre, leading to greater concurrent grouping and resulting sensory dissonance. Due to timbral differentiation and attack differences in the orchestral version, individual voices played by different instruments would have a greater tendency to segregate, thus decreasing the fusion of vertical sonorities and the dissonance that derives from that fusion, thereby reducing the perception of musical tension.

Conclusion

Timbre depends on concurrent auditory grouping processes. Its qualities are based on emergent acoustic properties that arise from perceptual fusion. Timbre can distinguish voices in polyphonic textures and among orchestral layers. It can underscore contrastive structures and define sectional structure. It also contributes to the building of large-scale orchestral gestures. Future possibilities for timbre research in music theory, orchestration theory, and music psychology include: determining how to predict blend from the underlying perceptual representation, the resulting timbral qualia of blended sounds, and which timbres will remain identifiable in a blend; the way timbre affects the interaction of concurrent and sequential grouping processes in the perception of dissonance and harmonic tension; and the contributions of timbre to the perception and cognition of formal processes and harmonic schemas.

Acknowledgments

Portions of the research reported in this chapter were supported by funding from the Fonds de recherche Québec—Société et culture (2014-SE-171434) and the Social Sciences and Humanities Research Council of Canada (890–2014–0008).

Core Reading

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

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.

McAdams, S. (1989). Psychological constraints on form-bearing dimensions in music. *Contemporary Music Review*, *4*(1), 18–98.

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

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.

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

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

Further References

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

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.

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.

Cannon, S. (2015). Arrival or relaunch? Dynamics, orchestration, and the function of recapitulation in the nineteenth-century symphony. Poster presented at the 2015 meeting of the Society for Music Theory, Saint Louis, MO.

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

Demany, L., & Semal, C. (1993). Pitch versus brightness of timbre: Detecting combined shifts in fundamental and formant frequency. *Music Perception*, *11*, 1–14.

Dolan, E. (2013). *The orchestral revolution: Haydn and the technologies of timbre.* Cambridge: Cambridge University Press.

Elliott, T., Hamilton, L., & Theunissen, F. (2013). Acoustic structure of the five perceptual dimensions of timbre in orchestral instrument tones. *Journal of the Acoustical Society of America*, *133*, 389–404.

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

Goodchild, M. (2016). *Orchestral gestures: Music-theoretical perspectives and emotional responses*. PhD Dissertation, McGill University, Montreal, Canada.

Guhn, M., Hamm, A., & Zentner, M. (2007). Physiological and music-acoustic correlates of the chill response, *Music Perception*, *24*, 473–483.

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

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

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

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

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

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-07)*, Bordeaux, France.

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

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.

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

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).

Patil, K., Pressnitzer, D., Shamma, S., & Elhilali, M. (2012). Music in our ears: The biological basis of musical timbre perception. *PLoS Computational Biology, 8,* e1002759. doi:10.1371/journal.pcbi.1002759

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.

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.

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

Shamma, S. (2000). The physiological basis of timbre perception. In M. Gazzaniga (Ed.), *The new cognitive neurosciences* (pp. 411–423), Cambridge, MA: MIT Press.

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*, 1977. doi: 10.3389/fpsyg.2015.01977

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

Tardieu, D., & McAdams, S. (2012). Perception of dyads of impulsive and sustained instrument sounds. *Music Perception, 30*, 17–128.

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.

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.