Chapter 8
Timbre as a Structuring Force in Music

Stephen McAdams

Abstract The study of timbre by music researchers is seriously underdeveloped in both the humanities and human sciences. As applied to instrumental music, theories to explain instrumental combinations and timbral shaping through instrumentation and orchestration are rare. Analyses of orchestration treatises and musical scores reveal an implicit understanding of auditory grouping principles by which many orchestral techniques and their perceptual effects function. This chapter, with a primary focus on classical Western orchestral and electroacoustic music, discusses connections between orchestration practice and perceptual principles based on research in auditory scene analysis and timbre perception. The chapter explores: (1) listeners' abilities to perceive relations among timbres; (2) how concurrent grouping cues result in blended or heterogeneous combinations of instruments; (3) how sequential groupings into segregated melodic streams and stratified foreground and background layers are influenced by timbral similarities and differences; and (4) how segmental grouping cues based on changes in instrument timbre and instrumental textures create musical units, formal boundaries, and expressive shaping of timbre melodies and larger-scale orchestral gestures.

Keywords Auditory grouping · Auditory scene analysis · Auditory stream segregation · Concurrent grouping · Orchestral gesture · Perceptual fusion · Segmental grouping · Sequential grouping · Timbre blend · Timbre contrast · Timbre interval

S. McAdams (✉)
Schulich School of Music, McGill University, Montreal, QC, Canada
e-mail: stephen.mcadams@mcgill.ca

K. Siedenburg et al. (eds.), *Timbre: Acoustics, Perception, and Cognition*,
Springer Handbook of Auditory Research 69,
https://doi.org/10.1007/978-3-030-14832-4_8

8.1 Introduction

Timbre perception is at the heart of orchestration practice, that is, the choice, combination, and juxtaposition of instruments to create a specific musical effect. Examples include picking a particular instrument for the emotional tone it can convey, such as the melancholy English horn in the third act of Wagner's opera "Tristan und Isolde", or bouncing musical patterns between contrasting instrument families as in the second movement of Beethoven's ninth Symphony (where a repeating call and response pattern alternates between woodwinds plus brass and strings plus brass). For practicing composers and conductors, the potential of timbre to structure musical forms and to sculpt music's emotional impact is evident; however, until relatively recently (see Thoret et al. 2018), these roles of timbre have been addressed only rarely in music research in both the humanities (music theory, musicology, ethnomusicology) and the behavioral sciences (experimental psychology). Most researchers and theorists focus on the musical parameters of pitch and duration that give rise to melody and harmony on the one hand and rhythm and meter on the other (obviously in concert with other so-called "secondary" parameters such as loudness or musical dynamics and timbre).

An examination of writings on orchestration practice from the middle of the nineteenth century to present times (e.g., Berlioz and Strauss 1948; Adler 2002) reveals that the communication of knowledge about orchestration is primarily based on a multitude of examples of various techniques. From these, students must memorize all the cases or somehow implicitly derive theory by studying scores and listening carefully to recordings over a span of many years. They also learn by practicing orchestration techniques, with the added difficulty of not always being able to hear the musical result of what they might write on the page of a score because they rarely have access to an orchestra.

An alternative approach would be to start with the assumption that orchestration conventions have some implicit basis in auditory grouping principles, given that composers are most likely grounding what they do in their own auditory experience (Goodchild and McAdams 2018). As detailed in Sect. 8.3, auditory grouping includes the perceptual fusion of concurrent acoustic information into auditory events, the perceptual connection through time of similar events into auditory streams (melodic lines) or foreground and background layers, and the segmentation of streams or layers into "chunks" that can be processed in short-term memory. Timbre arises from perceptual fusion as a property of an event. It can then influence the way successive events form auditory streams because listeners tend to connect events coming from the same sound source and because, generally speaking, a given source varies relatively little in timbre compared to the differences between distinct sources. Timbral contrasts can provoke segmentation in which successions of events with similar timbres form units separated from preceding or succeeding material with different timbres. From this perspective, the role of timbre as a structuring force in music can be addressed through the following set of questions:

- Can relations among timbres in sequences be perceived, stored in memory, and subsequently recognized as intervals or contours analogous to the perception of pitch and duration relations?

- What is the relation between auditory fusion and the perception of timbre?
- In what way do auditory-scene-analysis principles and acoustic properties contribute to determining whether separate sound events will blend together?
- How do timbral continuity and discontinuity contribute to the formation of auditory streams and the formation of foreground and background orchestral layers?
- How do timbral changes and the learning of timbral patterns affect perceptual segmentation of sequences, and what role do these play in music?
- How do gradual and sudden timbral changes contribute to larger-scale musical form?

8.2 Perception of Timbral Relations

One of the properties of musical pitch that endows it with its psychological capacity to serve as a vehicle for musical patterns and forms is that *relations* among pitches (contours or intervals) can be perceived as musical qualities per se. Musical patterns can be constructed with these qualities, and operations on those patterns that maintain the structural relations, such as transposition, also maintain a strong degree of perceptual similarity between the original and transformed materials. For example, someone can hum the tune to *Happy Birthday* starting on any pitch and, if the intervals are correct, the melody is still recognized. In order to extend these form-bearing possibilities of pitch into the realm of timbre, it would be necessary to determine the kinds of structuring of timbral relations that can be perceived by listeners and that still provide a certain richness to be reasoned with by composers. For the psychologist, several interesting questions arise concerning a listener's ability to perceive and remember timbral relations in tone sequences and to build up hierarchical mental representations based on those relations (McAdams 1989).

Timbre space provides a model for relations among timbres. A *timbre space* is derived from dissimilarity ratings on all pairs of a set of sounds (usually equalized for pitch, duration, and loudness) to which a multidimensional scaling algorithm is applied to model the dissimilarities as distances in a Euclidean space (for more detail, see McAdams, Chap. 2). Sounds with similar timbres are close in the space and different ones are farther apart. The dimensions are presumed to be perceptual. A timbre interval can be considered as a vector connecting two timbres in such a space, and transposing that interval maintains the same amount of change along each perceptual dimension of timbre. One might ask 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. Consider the timbral trajectory shown in Fig. 8.1 through the McAdams et al. (1995) timbre space starting with the *guitarnet* (a synthetic hybrid of guitar and clarinet) and ending with the English horn imitation. How would one construct a timbre sequence starting from the bowed string so that it would be perceived as a transposition of this *Klangfarbenmelodie* (the German term for tone color melody

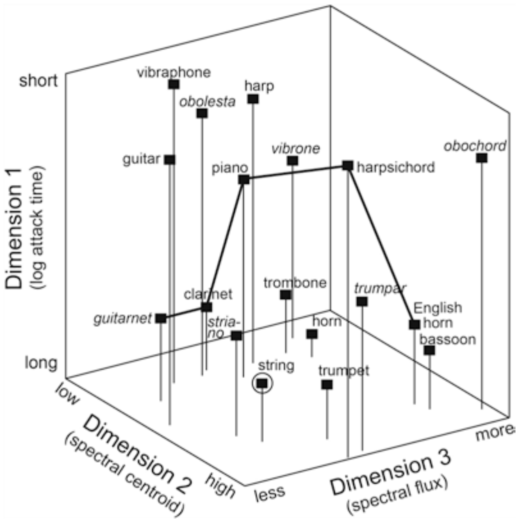

Fig. 8.1 A trajectory (*heavy black line*) of a short timbre melody through timbre space of synthesized sounds intended to mimic acoustical instruments or hybrids (*in italics*). How would one transpose the timbre melody that starts on *guitarnet* to a melody starting on string (*circled*)?

introduced by Schoenberg [1978])? If timbre interval perception can be demonstrated, the door would be opened for the application of some of the operations commonly used on pitch sequences to timbre sequences (Slawson 1985). The perceptual interest of this possibility is that it would extend the use of the timbre space as a perceptual model beyond the dissimilarity paradigm used to construct it in the first place.

Ehresman and Wessel (1978) first conceived of the notion of a timbre interval as the vector between two points in a timbre space. They tested the timbre-vector hypothesis by asking listeners to compare two timbre intervals (A-B versus C-D): A, B, and C were fixed and there were various Ds presented. Listeners ranked the Ds according to how well they fulfilled the analogy: timbre A is to timbre B as timbre C is to timbre D (notated A:B :: C:D; see CD1 vector in Fig. 8.2). The ideal CD vector would be a simple translation of the AB vector in the space with A, B, C, and D forming a parallelogram (shown with dashed lines in Fig. 8.2). Ehresman and Wessel found that the closer timbre D was to the ideal point defined by the parallelogram model, the higher the ranking.

McAdams and Cunibile (1992) subsequently tested the vector model using the three-dimensional space from Krumhansl (1989) and varying the orientation and length of the vector compared to the ideal values. In Krumhansl's timbre-space model, each sound had a position in the three shared dimensions, but they also had a factor specific to each sound that increased its distance from the other sounds, called its "specificity" (see McAdams, Chap. 2). The specificities were ignored in McAdams and Cunibile's calculations. They selected different kinds of Ds (see Fig. 8.2): D1 was near the ideal spot; D2 was about the same distance from C, but was at least 90° in the wrong direction; D3 was in about the right direction from

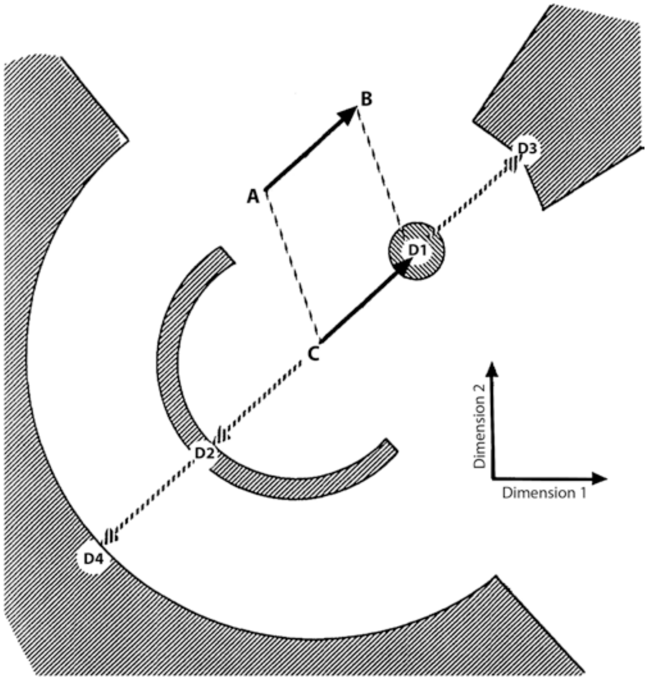

Fig. 8.2 Two-dimensional representation of the different sequence types used by McAdams and Cunibile (1992). The *hashed areas* represent the constraint space for the end points of CD vectors, which are labeled *D1*, *D2*, *D3* or *D4*, accordingly. The ideal point would be at the tip of the arrowhead for the CD vector that forms a parallelogram with AB (connected by dashed lines). For the three-dimensional case, the area would be a small sphere for D1, a shell for D2, part of a cone for D3, and a solid with a hemispherical hollow for D4. (Adapted from figure 2 in McAdams and Cunibile 1992; used with permission of The Royal Society)

C, but its length was at least 1.8 times greater than that of the ideal vector; and D4 was both too far and in the wrong direction. Listeners compared pairs of A:B :: C:D analogies constructed in this way and had to choose for which one C:D seemed most similar to A:B (e.g., compare A:B :: C:D1 with A:B :: C:D4). Five sets of timbres at different places in Krumhansl's timbre space were chosen for each comparison to test for the generality of the results. In general, timbres close to the ideal point predicted by the vector model were preferred as better fulfilling the A:B :: C:D analogy than were timbres that were at some distance from that point. Both nonmusicians and composers of electroacoustic music found the task rather difficult. This shouldn't be too surprising given that even professional composers have had almost no experience with music that systematically uses timbre intervals to build musical structures. Support for the model was stronger for electroacoustic composers than for nonmusicians, however, suggesting some effect of musical training and experience.

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 the sounds in this set were ignored in computing the vectors and selecting the analogies. These specificities would of course 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 corresponded well to what was predicted because the specificities were absent or low in value, whereas others were seriously distorted and thus perceived as less similar to other intervals due to moderate or high specificity values.

The relative lack of generalizability of timbre interval perception across different timbres may be due to a number of factors that were not controlled in McAdams and Cunibile's study. First, there may be a relative instability of judgement strategies given that most listeners have never encountered a listening situation in which perceiving abstract timbral relations was appropriate. Second, there may be effects of the relative magnitude of a given vector and the distance between to-be-compared vectors: it may be difficult to compare with precision very large vectors or small vectors that are very far apart in the space. 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, such as acoustic or electronic musical instruments, because they will, in all probability, have specificities of some kind or another.

It may be difficult to use timbre intervals as an element of musical discourse in a general way in instrumental music given that the timbre spaces of acoustical instruments also tend to be unevenly distributed (see Fig. 8.1), unlike the regular spacing of pitches in equal temperament. If timbre intervals are to be used, in the long run they will most likely need to be limited to synthesized sounds or blended sounds created through the combination of several instruments. 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, which would allow for the perception of similar contours (patterns of relative change along the different dimensions) in trajectories through timbre space. Indeed, McDermott et al. (2008) have shown that timbral brightness contours (patterns of ups and downs) can be recognized irrespective of the exact amount of change and also can be compared to contours in pitch and loudness.

It should be noted, nonetheless, that in a context in which pitch is a structuring factor, timbre may have difficulty imposing itself as a dominant parameter in terms of relational perception, primarily due to a sort of dominance hierarchy favoring duration and pitch relations (rhythm and melody) when several parameters are in play. Research on the conditions under which a given musical parameter plays a significant role in the perceptual structuring of music when varied in the presence of other parameters is limited and rarely goes beyond the royal couple of pitch and duration. A first attempt in this direction, which only used nonmusical sequences, was conducted by Krumhansl and Iverson (1992). They found that the classification of pitches (high versus low) or timbres (bright versus dull) was symmetrically affected by uncorrelated variation in the other parameter: reaction times for pitch were slower when having to ignore random changes in timbre compared to when timbre was held constant and vice versa. This result suggests that it is difficult to ignore either parameter (pitch or timbre) when both are changing and indicates a tight relation between timbral brightness (change in the spectral centroid) and pitch height. This link would be coherent with underlying neural representations that share common attributes such as a tonotopic organization of spectral distribution (for more on pitch-timbre interactions, see McAdams, Chap. 2).

In two other experiments, Krumhansl and Iverson (1992) asked listeners to focus their attention on either the pitch or timbre of a single target event in a short sequence and to decide whether the same event in a second sequence was identical or different with regard to the parameter being tested. In addition, the other notes around the target event could vary either in terms of the attended parameter, the unattended parameter, or both. Globally, timbres were recognized better than pitches. A change in pitch context did not affect recognition of the target timbre and, similarly, a change in timbre context left pitch recognition unaffected. A change in pitch context strongly affected recognition of the pitch of the target event, however, indicating that listeners code relations between pitches (i.e., relative pitch) in memory rather than the absolute pitch value. To the contrary, the effect of variation in timbre context only weakly affected target timbre recognition and only when there was no variation in pitch context. This result suggests that although pitch is coded in relative terms, timbre is more likely to be coded absolutely as a sound source category, and relations among timbres are only coded when pitch does not vary. Krumhansl and Iverson (1992) concluded that relational structures among timbres would be difficult to perceive in the case in which other musical parameters vary independently. Siedenburg and McAdams (2018) presented converging evidence regarding the interference of concurrent pitch variation in a timbre-sequence recognition task (also see Siedenburg and Müllensiefen, Chap. 4). It remains to be seen, however, what the interplay of pitch-based and timbre-based structuring forces would be in instrumental and electroacoustic music that is based more on sound colors and textures and less on melody and harmony.

8.3 Timbre and Auditory Grouping

Figure 8.3 summarizes the grouping processes involved in auditory scene analysis, as well as the resulting perceptual attributes related to orchestral effects produced in music. *Concurrent grouping* determines how components of sounds are grouped together into musical events, a process referred to in psychology as *auditory fusion*. This grouping process precedes, and thus conditions, the extraction of the perceptual attributes of these events, such as timbre, pitch, loudness, duration, and spatial position. The result of combining sounds concurrently in orchestration is *timbral blend* when events fuse together or *timbral heterogeneity* when they remain separate (see Sect. 8.3.1). *Sequential grouping* connects these events into single or multiple auditory streams on the basis of which perception of melodic contours and rhythmic patterns is determined (McAdams and Bregman 1979). The orchestral effect is the integration or segregation of events into streams, textures, and foreground and background layers (see Sect. 8.3.2). And finally, *segmental grouping* affects how events within streams are chunked into musical units, such as motifs, phrases, themes, and sections.

Timbral similarity contributes to the unification of segments of music that are set off from adjacent segments when timbral contrast is introduced. Continuous change in timbre is used in progressive orchestration to create timbral modulations or larger-scale orchestral gestures (see Sect. 8.4). It becomes apparent here that auditory grouping processes are implicated in many aspects of orchestration practice, including the blending of instrument timbres, the segregation of melodies and layers based on timbral differences, and the segmentation of contrasting orchestral materials that results in the creation of perceptual boundaries in musical structures.

8.3.1 Timbre and Perceptual Fusion

As indicated in Fig. 8.3, timbre emerges from the perceptual fusion of acoustic components into a single auditory event, including the blending of sounds produced by separate instruments in which the illusion of a "virtual" sound source is created. The creation of new timbres through blending thus depends on the perceptual fusion of the constituent sound events. Concurrent grouping is affected by sensory cues, such as whether the acoustic components begin synchronously (onset synchrony), whether they are related by a common period (harmonicity), and whether there is coherent frequency and amplitude behavior (McAdams 1984). The coherent behavior cues are related to the *Gestalt principle of common fate*: Sounds that change in a similar manner are likely to have originated from the same source (Bregman 1990).

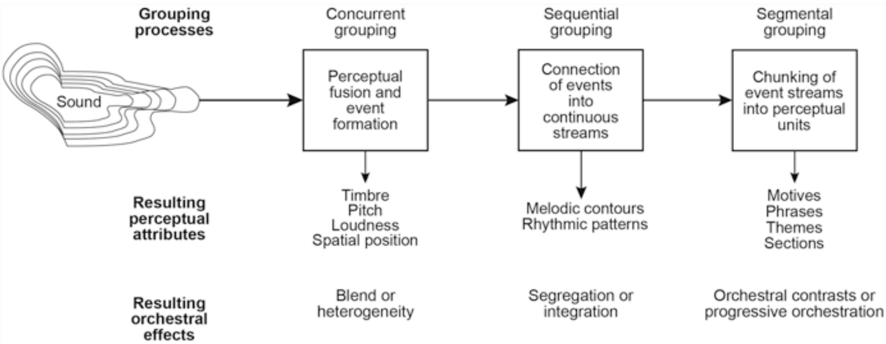

Fig. 8.3 Auditory grouping processes (concurrent, sequential, segmental) that give rise to perceptual attributes and orchestral effects

Lack of synchrony, harmonicity, and parallel change in pitch and musical dynamics (*piano*, *forte*) is likely to signal the presence of two or more sound sources and to provide the information needed to organize and group their respective frequency components separately (McAdams 1984). This means that instrument combinations are more likely to be perceived as blended if they adhere to these principles. This approach can be found in orchestration manuals that suggest strict doubling of a melodic line, most often at the unison or octave intervals at which maximum coincidence of the frequency components would occur. A classic example of the combined use of these cues is Ravel's piece *Boléro*, in which he constructs virtual sound sources with complex timbres by stacking instruments in a harmonic series (fundamental pitch, octave, twelfth, double octave, etc.) and having them play in synchrony and in perfect parallelism in terms of both pitch and dynamics (https://www.youtube.com/watch?v=dZDiaRZy0Ak&frags=pl%2Cwn, instrument combinations start at 4:35 with a trumpet-flute combination or at 7:03 with French horn on the fundamental pitch and celesta and piccolos on pitches corresponding to harmonics 2–5).

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. Sandell (1995) has demonstrated that sounds blend better when they have similar attack envelopes and spectral centroids, as well as a lower composite spectral centroid. He submitted listeners' blend ratings (taken as a measure of proximity) to multidimensional scaling and obtained a "blend space" whose dimensions were correlated with attack time and the spectral centroid, suggesting that the more these parameters were similar for the two combined sounds, the greater their blend. Kendall and Carterette (1993) found a similar trend concerning the role of spectrotemporal similarity in blend for wind instrument combinations. Tardieu and McAdams (2012) reported that greater blending is achieved with lower spectral centroids and slower attacks for combinations of pitched impulsive and sustained sounds (e.g., a vibraphone and bowed cello). However, the timbre resulting from the blend is determined primarily by the attack of the impulsive sound and the spectral envelope of the sustained sound, which create a chimeric sound with the head of one and the tail of the other, respectively.

In addition to global descriptors, such as a spectral centroid, research has also been conducted on the role of local descriptors of *formant structure* (prominent spectral maxima that are invariant with respect to pitch change) on wind and brass instrument blends (Reuter 2003). Lembke and McAdams (2015) extended this approach, characterizing wind-instrument spectra in terms of pitch-generalized spectral envelope descriptions. Some instruments exhibit a formant-like structure with prominent spectral peaks. They conducted two experiments employing blend-production and blend-rating tasks and studied the perceptual relevance of these formants to the blending of dyads composed of a recorded instrument sound and a parametrically varied synthesized sound. Relationships between formant center frequencies influenced blend critically, as did the degree of formant prominence.

These relations of spectral overlap and perceptual fusion seem to hold not only for the blending of sounds of pitched instruments, as mentioned above, but also for vocal sounds (Goodwin 1980) and even noise and pitched sounds as in the case of Burundian Inanga *chuchoté* (whispered Inanga) (https://www.youtube.com/watch?v=94TGtf7PdyE&frags=pl%2Cwn). The Inanga is an African zither that is traditionally accompanied by whispered voice (Fales and McAdams 1994). The latter case is fascinating because the language of that culture is tonal, and the musical instrument helps to communicate the pitch contours that disambiguate the reduced phonetic information provided by the whispered voice. These results taken together demonstrate the importance of spectral overlap in the perception of blend.

Sandell (1995) has proposed three possible perceptual results of instrument combinations. The first is *timbral heterogeneity*: individual sounds are segregated and identifiable. The second is *timbral augmentation*: subservient sounds are blended into a more dominant, identifiable sound whose timbre is then reinforced or highlighted by them. The third is *timbral emergence*: all sounds are blended and unidentifiable. An inverse relation between the degree of blend and the identifiability of the constituent sounds has been documented by Kendall and Carterette (1993). Future work is needed to develop models that can predict: (1) the degree of blend from the underlying perceptual representation, (2) the timbral qualia that emerge from blended sounds, and (3) which timbres are likely to be dominant or remain identifiable in a blend.

Additional factors that also seem to play a role in blend, but which have not been studied systematically, are event duration and spectral density. Punctuated sonorities with many instruments playing across a range of pitches for short durations, as in the opening chord of Ludwig van Beethoven's Third Symphony, the 'Eroica' (https://www.youtube.com/watch?v=nbGV-MVfgec&frags=pl%2Cwn), do not provide enough time to "hear into" the sonority and analyze out the different constituent sounds. In addition, sound masses with many closely spaced pitches are similarly difficult to analyze due to auditory limits in spectral resolution, as one finds in sound mass orchestral pieces such as György Ligeti's *Atmosphères* (https://www.youtube.com/watch?v=9XfefKJRoSA&frags=pl%2Cwn), made popular through the Stanley Kubrick film *2001: A Space Odyssey*.

8.3.2 Timbre and Sequence Perception

The connection of successive sound events into a coherent perceptual 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). Musicians would call these streams musical lines, parts, or voices. An auditory stream is a mental representation of continuous sound activity considered by the perceptual system to be emanating from a single sound source (Bregman 1990). Sequential grouping processes organize successive events into a single stream or multiple streams based on specific cues, which are closely related to *Gestalt principles of proximity* (closeness in time) and *similarity* in auditory properties, such as pitch, timbre, loudness, and spatial position (McAdams and Bregman 1979). One of the main hypotheses behind the theory of auditory scene analysis is that the auditory system operates according to a heuristic that a sequence of events produced by a single sound source will be similar in terms of spectral content (affected by pitch register, instrument, and playing effort), intensity (affected by pitch register and playing effort), and spatial position. Continuities in these cues would thus promote the integration of the events into a stream, and discontinuities would signal the presence of other sound sources, leading to the segregation of the events into different streams within which events are similar. So sequences of sounds from different instruments can be segregated on the basis of timbre (imagine a duo of violin and piano) as can sounds from a single instrument that have very different timbral characteristics, as one might find, for example, in *Nel cor più non mi sento* (*I do not feel my heart anymore*) for solo violin by Niccolò Paganini, in which bowed and plucked sounds form separate streams in counterpoint with each other (https://www.youtube.com/watch?v=OpxwHm_a_Po&frags=pl%2Cwn starting at 1:12). It is important to note that timbre covaries with pitch, playing effort, and articulation in musical instruments and so cannot be considered independently; therefore, changing the pitch or the musical dynamic also changes the timbre.

Once timbre has been formed following concurrent grouping, it plays an important role in determining whether successive sounds are integrated into an auditory stream on the basis of similarities in spectrotemporal properties or segregated into separate streams based on timbral differences that potentially signal the presence of multiple sound sources (McAdams and Bregman 1979; Gregory 1994). This process reflects the fact that individual sources do not generally tend to change their acoustic properties suddenly and repeatedly from one event to the next (for reviews, see McAdams and Bregman 1979; Chap. 2 in Bregman 1990). As the difference between timbres gets larger, the resulting stream segregation gets stronger. Because melody and rhythm are perceptual properties of sequences that are computed within auditory streams (Fig. 8.3), timbre can strongly affect the perception of these musical patterns. A clear demonstration of this principle is depicted in Fig. 8.4.

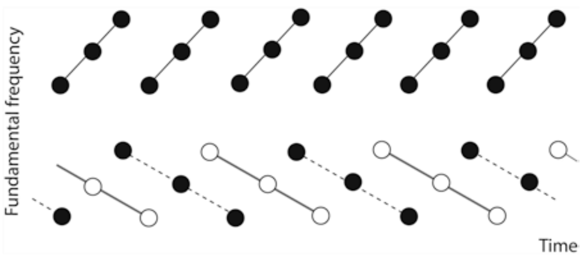

Fig. 8.4 Schematic diagram of the two versions of a melody created by David Wessel (1979) with one instrument (*top*) or two alternating instruments (*bottom*). In the *upper melody*, a single rising triplet pattern is perceived at a particular tempo. In the *lower melody*, if the timbral difference between the sounds of the two instruments (indicated by *open* and *filled circles*) is sufficient, two interleaved patterns of descending triplets at half the tempo of the original sequence are heard, as indicted by the dashed and solid lines

These early demonstrations of auditory streaming on the basis of timbre suggest a link between the timbre-space representation and the tendency for auditory streaming on the basis of the spectral differences that are created. Hartmann and Johnson (1991) argued that aspects of timbre derived from the spectral distribution are primarily responsible for auditory streaming, and 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 (Cusack and Roberts 2000; for a review, see Moore and Gockel 2002). In one study, Iverson (1995) 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* from which acoustic properties that contributed to stream segregation could be determined. He compared the segregation space with a timbre space derived from the same sounds (Iverson and Krumhansl 1993) and showed that both static acoustic cues (such as the spectral centroid) and dynamic acoustic cues (such as attack time and spectral flux) were all implicated in segregation.

Iverson's findings were refined in an experiment by Singh and Bregman (1997). They varied the amplitude envelope and the spectral content independently and measured the relative contributions of these parameters to auditory stream segregation. A change from two to four harmonics (which would change both the centroid and the spread of the spectrum) produced a greater effect on segregation than did a change from a 5 ms attack and 95 ms decay to a 95 ms attack and 5 ms decay. Combining the two gave no greater segregation than was obtained with the spectral change, which suggests a stronger contribution of the spectral property to segregation. However, it should be noted that differences in the attacks of sounds produced by musical instruments involve many more acoustic properties than just a change in the amplitude envelope because they include noisy attack transients and rapid changes in spectral distribution during the attack.

In a slightly more musical task, Bey and McAdams (2003) used a melody discrimination paradigm. They first presented listeners with a target melody interleaved with another melody that served as a distractor such that if the two were not segregated the target melody would be camouflaged by the distractor. This mixed sequence was followed by a test melody that was either identical to the target or differed by two notes that changed the contour. Listeners were asked to decide whether the test melody was present in the previous mixture. Note that with the presentation of the test melody after the mixture, the listener must first organize the mixture into streams and then compare the melody carried by the target timbre with the ensuing test melody with the same timbre. The timbre difference between target and distractor melodies was varied within the timbre space of McAdams et al. (1995). In line with the results of Iverson (1995), melody discrimination increased monotonically with the distance between the target and the distractor timbres, which varied along the dimensions of attack time, spectral centroid, and spectral flux. Here again, the temporal and spectrotemporal properties seem to play a significant role in stream organization in addition to purely spectral properties.

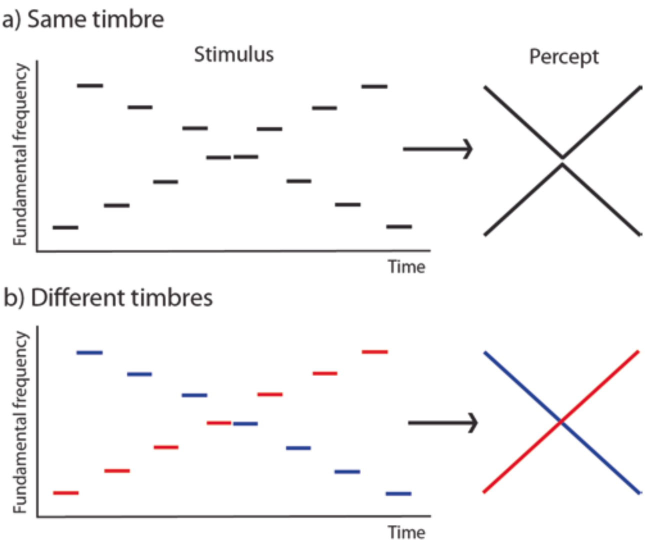

Fig. 8.5 Bouncing and crossing percepts of interleaved ascending and descending melodies depend on timbral differences. (**a**) When the timbre of ascending and descending melodies is the same, a bouncing percept is heard with V-shaped and inverted V-shaped melodies. (**b**) When the timbres are different enough (*represented by color*), the crossing melodies are heard

Timbral difference is also an important cue for following a voice (or musical part) that crosses other voices in pitch or for hearing out a given voice in a polyphonic texture with several independent parts (McAdams and Bregman 1979). Tougas and Bregman (1985) interleaved notes of ascending and descending scales, which are normally perceived as V-shaped and inverted V-shaped melodies that bounce at the crossover point (Fig. 8.5a) when the timbres of the scales are the same (same spectral content in terms of number of harmonics in their case). This percept has been interpreted as a demonstration of the role of the *Gestalt principle of pitch proximity*. However, when the spectral structures, and thus timbres, are different, this bouncing percept is replaced by the complete ascending and descending scales (Fig. 8.5b). So listeners form auditory streams with sounds of similar timbres and segregate different timbres into distinct auditory streams. Similar results are found with continuous glides of simultaneous vocal sounds composed of diphthongs: When the timbres of the vowels were the same at the moment of crossing, a bouncing perception was heard, and when they were different, crossing was perceived (McAdams and Bregman 1979; Culling and Darwin 1993). Timbre can thus play an important role in voice leading in polyphonic music. Voice leading is the connection of successive notes in a musical line, and timbre's role in this process has been largely ignored by music theorists; one exception is Huron (2016), who discusses timbral differentiation in Chap. 8 of his book.

If a composer seeks to create *Klangfarbenmelodien* (the German term for sound color melodies) that change in instrumental timbre from note to note, timbre-based streaming may prevent the listener from integrating the separate sound sources into a single melody if the changes are too drastic. The timbre sequences would not actually have the perceptual status of a melody (understood as an integrated stream within which relations among perceptual properties of successive events can be perceived) but would instead be perceptually fragmented, resulting more in a sort of *Klangfarbenzersplitterung* (sound color fragmentation)! Humans have a predisposition to identify a sound source and follow it through time on the basis of relative similarity in pitch, timbre, loudness, and spatial position (Bregman 1990; Siedenburg and Müllensiefen, Chap. 4). Cases in which such timbral compositions work successfully have used smaller changes in timbre from instrument to instrument (e.g., Anton Webern's orchestration of Bach's Ricercar from *The Muscial Offering*, https://www.youtube.com/watch?v=2cLALT09Y0M&frags=pl%2Cwn) or overlapping of instruments to create a line that cross-fades from one instrument to the next (e.g., Tristan Murail's *Mémoire/Érosion*, https://www.youtube.com/watch?v=dTZDCTzUcbA&frags=pl%2Cwn). However, if pointillistic fragmentation is the composer's desired aim, significant timbre change is indeed effective in inducing perceptual discontinuity.

Goodchild and McAdams (2018) propose two other groupings that are discussed in orchestration treatises but have not yet been studied empirically. They are different from stream segregation in degree and complexity more than in kind. One is *textural integration*, which occurs when two or more instruments that feature contrasting rhythmic figures and pitch materials 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. One might consider it as occupying a middle place between integration and segregation. The emergent property is a musical surface texture. The other one is *stratification*, in which two or more different layers of orchestral material are formed perceptually and are separated into strata of greater and lesser prominence or as foreground and background layers, with one or more instruments in each layer. Integrated textures often occupy an orchestral layer in a middleground or background position, providing a textural atmosphere. An excellent example from Bedřich Smetana's *The Moldau* (measures 187–212) is the intertwining melodies of two flutes and two clarinets in a middleground texture behind the high violin melody alternating with harp arpeggios in the foregrond and horns and other strings in the background (https://www.youtube.com/watch?v=gTKsHwqaIr4&frags=pl%2Cwn starting at 5:35). The beautifully rendered middleground texture in this passage represents the shimmering of moonlight on the Moldau River. A reasonable hypothesis is that similarity of timbre, pitch register, rhythmic patterning, and articulation within layers allow for their grouping together, and differences in these parameters between layers allow for their perceptual separation.

Huron (2016) raises an interesting issue concerning orchestration practice and the effectiveness of timbral differentiation on the segregation of contrapuntal parts. He notes that contrary to what one might expect, composers often adopt more timbrally homogeneous instrument groupings in polyphonic works, such as the extensive repertoire for string quartets, brass ensembles, vocal groups, and solo keyboards. His hypothesis is that such selections of instrumentation by composers may be related to their goal of maintaining balance among the parts because heterogeneous ensembles may present perceptual difficulties due to the differences among the instruments in terms of their acoustic power and their relative salience. Huron (2016) therefore proposes that timbral differentiation is often reserved as a device used to distinguish foreground from background layers. Finally, the elusive notion of *timbral salience*, the properties that capture one's attention and lead to the distinction between foreground and background prominence, needs to be explored empirically in musical settings, although there is some research into environmental settings (Huang and Elhilali 2017).

8.4 Timbre and the Perception of Musical Form

8.4.1 Timbral Contrasts and Segmentation

Having examined how timbre derives from concurrent grouping and plays a role in sequential grouping, let us now consider how timbral discontinuities promote segmental grouping, a process by which listeners segment musical streams into units such as motifs, phrases, themes, and sections (Fig. 8.3). People organize and make sense of continuous streams of acoustic information partly by segmenting them into events, that is, meaningful units. Event segmentation is most likely an ongoing component of everyday perception that composers use. Changes in musical features are a common cue for segmentation, and listeners will indicate segment boundaries in listening experiments if strong enough changes in pitch register, dynamics, instrumentation, and duration occur. The more each feature changes and the more features that change simultaneously, the stronger is the sense of boundary (Hartmann et al. 2016). Furthermore, events are segmented simultaneously at multiple timescales and are grouped in hierarchical fashion with groups over smaller timespans being nested within groups occupying larger timespans. This nesting makes segmentation a crucial component in the formation of a hierarchical mental representation of musical form.

An "event" is some segment of time occupied by sensory information that is conceived by a listener as being bounded by a beginning and an end. For example, notes are events that are grouped into higher-level events of rhythms and melodies, which are in turn grouped into phrases and sections. The parsing of continuously incoming sensory information into events is closely related to the process of updating working memory (the part of short-term memory concerned with immediate conscious perceptual and linguistic processing) and depends on contents stored in long-term memory. Kurby and Zacks (2008) proposed that event segmentation may arise as a side effect of an adaptive mechanism that integrates information over the recent past to improve predictions about what may arrive in the near future. When perceptual features change, it becomes more difficult to predict what will follow, and errors in prediction increase momentarily. At such points, listeners need to update their memory representations of what is actually going on. Kurby and Zacks (2008) hypothesized that two processes give rise to the subjective experience that a new event has begun: (1) the detection of a momentary increase in prediction errors created by the violation of expectations generated by a model of the current event, and (2) the updating of the event model in working memory that results from the expectation violation. One important feature of this approach is the notion that events can be organized at a range of temporal grains, from fine grained (e.g., notes) to coarse grained (e.g., phrases or sections).

Event segmentation may be accomplished contextually on the basis of the internal continuity relations in the music. Generally speaking, a segment is characterized by relative internal continuity as concerns the degree and rate of change of the perceptual properties of the musical materials being heard and by a relative discontinuity at its terminal points (for an application of these principles to music analysis, see Oliver 1967). So change creates prediction errors based on presumed continuity, and the errors in turn cause segmentation. According to the *Gestalt principle of similarity*, sounds that resemble one another are grouped together and are segmented into chunks that are bounded by acoustic dissimilarities. Gradual changes over a given time period would create a sense of continuity, whereas discontinuities promote segmentation into musical units. So musical segments are formed on the basis of similarities in register, texture, and instrumentation (i.e., timbre), and changes in one or more of these musical features signal boundaries at various levels of the musical hierarchy, depending on the cumulative degree of change among them (Deliège 1989).

In their generative theory of tonal music, Lerdahl and Jackendoff (1983) proposed a series of grouping preference rules that reflect how listeners perceptually structure musical sequences. Two *Gestalt principals of temporal proximity* and *qualitative similarity* underlie the rules, most of the latter resulting from a change or discontinuity in one or more auditory attributes, including pitch register, musical dynamics, articulation (staccato, tenuto, legato, mostly related to the duration of gaps between successive notes), note duration, and timbre. Deliège (1987) experimentally tested the extent to which listeners segmented musical phrases in accordance with these grouping rules. She found that timbral discontinuities (changes in instrument timbre) were among the changes that both musician and nonmusician listeners detect most often in short phrases.

Specific evaluation of the role that timbre plays in the segmental structuring of music is limited in music-theoretical and perceptual scholarship. Goodchild and McAdams (2018) propose several types of contrasts that are often found in the orchestral repertoire: (1) antiphonal alternation of instrumental groups in call-and-response phrase structure (antiphonal is from the Greek *antiphonos*, which means "responsive, sounding in answer"); (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; (3) timbral shifts in which musical materials are reiterated with varying orchestrations, passed around the orchestra, and often accompanied by the elaboration or fragmentation of musical motifs; and (4) larger-scale sectional contrasts with major changes in instrumental forces, passing from a full orchestra to a solo violin, for example. The perceptual strengths of these different contrasts depend on the timbral changes used. Furthermore, the timbral differences may play a role in which musical materials are perceived as a call versus a response in call-response patterns, or as an original pattern versus an echoed version of that pattern. Listeners also segment large-scale sections of contemporary works on the basis of marked contrasts in instrumentation and texture (Deliège 1989). Therefore, timbral discontinuities promote the creation of perceptual boundaries, whereas continuities promote the grouping of events into coherent units at various levels of the structural hierarchy. This means that timbral changes can affect both local and global levels of formal organization in music. However, timbral changes interact with changes in other musical parameters in terms of the strength of perceived boundaries. An excellent example of a timbral shift in which a melodic pattern is passed from one instrument to another can be found in Beethoven's *Egmont Overture*, with a sequence from clarinet to flute to oboe and back to clarinet with some fragmentation of the motive in the last two iterations (Fig. 8.6) (https://www.youtube.com/watch?v=2HhbZmgvaKs&frags=pl%2Cwn starting at 4:35). This pattern is repeated three more times, creating a timbral arch each time, which is set off by a two-note punctuation by most of the orchestra each time.

More research in this area would be useful to explore various timbral connections and their perception by listeners. One potential avenue for investigation is the use of timbre to create echo effects in which a repeated pattern sounds farther away on its second occurrence. Rimsky-Korsakov (1964) mentions the notion of echo phrases in which the imitation entails both a decrease in level and an effect of distance, ensuring, however, that the original and echoing instrument or instrument combination possess "some sort of affinity" (p. 110). He cites the example of a muted trumpet as being well suited to echo material in the oboes, and flutes may echo clarinets and oboes. Aside from these suggestions, one might ask: what techniques have composers used to create timbral echoes and how do they relate to perceptual principles? There are several cues to distance perception, including sound level, direct-to-reverberant energy ratio, and spectral filtering. More distant sounds are less intense, have a lower ratio of direct-to-reverberant sound energy, and have less energy in the higher frequencies due to absorption in the air and by sur-

Fig. 8.6 A repeating timbral shift pattern shown by the *blue boxes* from clarinet to flute to oboe back to clarinet in Beethoven's *Egmont Overture*, measures 117–132. All instruments are notated at sounding pitch. The two-note staccato chord at the end of each pattern is also played by other instruments in the orchestra that are not shown in this detail of the score

faces in rooms (Zahorik et al. 2005). The cues that could be simulated with orchestration would be dynamics and spectral properties related to timbre (although some composers use off-stage instruments as well to get the full effect of distance).

Another process by which sequences are segmented into smaller-scale units involves detecting repeating timbral patterns and learning the transition probabilities between timbres over sufficient periods of time. Bigand et al. (1998) presented listeners with artificial grammars of musical sounds for which rules of succession of the sounds were created. After being exposed to sequences constructed with the grammar, listeners heard new sequences and had to decide if the sequence conformed to the learned grammar without having to say why. Indeed, by implicit learning of structures (language and music), any listener can know if a sequence corresponds to the structure in question without knowing why: quite simply, it doesn't "sound" right. The average correct response rate of Bigand and colleague's listeners was above chance, indicating the listeners' ability to learn a timbral grammar.

Tillmann and McAdams (2004) explored this idea further in the direction of segmentation per se based on work by Saffran et al. (1996), who sought to understand how implicit statistical learning of transition probabilities between syllables in language might lead to segmentation into words by infants. The idea is that syllables within words follow each other more often than do syllables in different words, and building up a statistical distribution of such transitions would help segment speech streams into units that correspond to words in a given language. The same research group demonstrated a similar ability in infants with pitched tone sequences, suggesting the ability applies more generally than just to speech (Saffran et al. 1999).

Tillman and McAdams applied this principal to timbre sequences using the sounds from McAdams et al. (1995) with constant pitch, loudness, and roughly equivalent duration. A lexicon of grammatical timbre triplets was created and presented sequentially in random order in an isochronous sequence for about 33 min. The probability of transition from the last sound of one triplet to the first sound of the next triplet was designed to be much lower than was the transition probability between the first and second sounds and the second and third sounds. The listeners were then tested on their recognition of triplets from the timbral grammar with the expectation that they would implicitly learn the transition probabilities among timbres. A control group was tested on a similar task without exposure to the timbral grammar. To examine the role of auditory segmentation on the basis of timbre discontinuity in the learning of timbre sequence regularities, the timbral distance relations among sounds were organized in three different conditions in which the distances between successive timbres of the triplets—as determined from the McAdams et al. (1995) timbre space—were coherent, incoherent, or neutral with respect to the grammatical triplets.

There were significant differences among the sequence types: the coherent type (S1) had the highest choice of grammatical triplets, followed by the neutral type (S3), and then the incoherent type (S2) in both learning and control groups (Fig. 8.7). So the acoustic similarities strongly affected the choices made by listeners: they preferred triplets with smaller distances between them. However, there was no

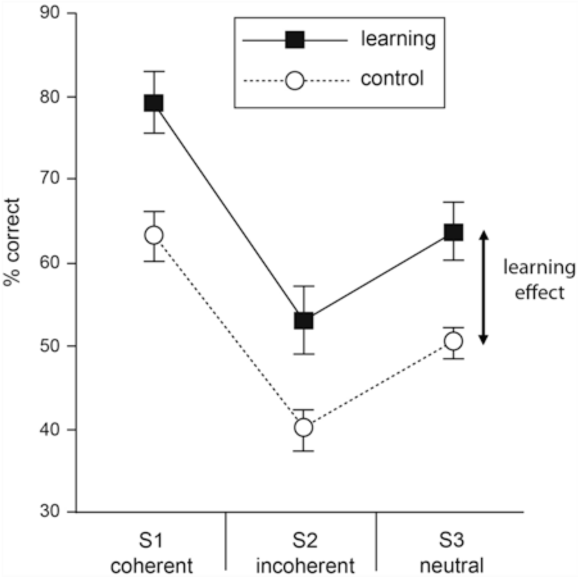

Fig. 8.7 Percentage of correct identification of timbre triplets as belonging to the timbral grammar as a function of listener group (implicit learning through hearing grammatical timbre triplets for 33 min versus control) and sequence type (*S*). The sequence types concern the alignment of acoustic grouping cues on the basis of timbral discontinuity and the implicitly learned grouping on the basis of transition probabilities (*S1: coherent*; *S2: incoherent*; *S3: neutral*). Coherent sequences were better recognized than neutral sequences, which were better than incoherent sequences. However, the amount of implicit learning was the same for all three groups. (Adapted from figure 1 in Tillmann and McAdams 2004; used with permission of The American Psychological Association, Inc.)

interaction between sequence type and listener group (the curves in Fig. 8.7 are parallel). An increase of about 14% in correct choice rate occurred for the learning group compared to the control group in all three sequence types, suggesting that learning of transition probabilities is not affected by segmentation on the basis of acoustic similarity. To summarize, even between dissimilar sounds and despite conflicting perceptual groupings, the cognitive system seems to become sensitive to statistical associative relationships among timbres. In everyday life, this capacity might be rather useful given that associations and statistical regularities (also concerning the temporal ordering of sounds) have to be learned between complex environmental sounds that can differ acoustically.

8.4.2 Timbre and Large-Scale Musical Form

Larger-scale units in music, such as formal functions (e.g., exposition, recapitulation) and types (e.g., sonata, rondo, theme, and variations), have been theorized in Classical music. Although rare, there has been some discussion of how these units can be articulated through orchestration in the music theory and musicology literature, but there is no perceptual research as yet. In classical sonata form, musical material is presented in an exposition section, then elaborated in a development section, and is returned to in a recapitulation section. Cannon (2016) studied contrasts in dynamics and orchestration (primarily instrument density) as key factors that determine 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. Examining several hundred sonata-form movements from nineteenth-century symphonies, he classified the alterations of the main theme on its return in the recapitulation into four broad types: (1) similar, (2) intensified by increased dynamic markings and instrumental forces, (3) attenuated with decreased dynamic markings and instrumental forces, and (4) contradictory with dynamics and instrumentation going in opposite directions. Cannon noted that Beethoven, for example, often intensified the theme in the recapitulation with full orchestra playing the theme at louder dynamic markings (as in the first movement of his First Symphony), a tendency observed in the majority of intensifications in Cannon's corpus. Brahms, however, often obscured the transition from development to recapitulation using lower dynamic markings. Haydn was known to use changes in orchestration in the symphonies composed during his stays in London at the end of the eighteenth century to give new color to a theme in the recapitulation (Wolf 1966). The distinction between a recapitulation being an arrival at a culminating climax point versus being a relaunch or new beginning was captured by both parameters. An arrival was often characterized by a buildup of instrumental forces and crescendo in dynamics that peaked at the onset of the recapitulation, whereas a relaunch was often characterized by a strong contrast in dynamics and instrumental texture between the end of the development and the beginning of the recapitulation. Thus, timbral factors contribute to this large-scale formal feature.

Dolan (2013) examined Haydn's structural and dramatic use of orchestration, including the process of developing variations of thematic materials. She emphasized the essential role played by orchestration in Haydn's articulation of musical form (see Chap. 2 in Dolan 2013). For example, themes that initially appear in one orchestration will return with a different one in order to nuance or transform the musical material subtly, at times keeping all other musical parameters constant, such as harmony, melody, and rhythm. Dolan describes how Haydn used opposing sonorities or textures to lend structure and dramatic impact by introducing interruptions of sections in strings or winds with full orchestral tuttis (all instruments playing together). A particularly instructive example is the second (Allegretto) movement of his "Military" Symphony, no. 100 (https://www.youtube.com/watch?v=6RmwapsXnrg&frags=pl%2Cwn). Figure 8.8 displays the evolution of instrumental involvement over the whole movement with time progressing from left to right, as indicated by the measures in the musical score on the x axis. Instrument families are shown with different hues (green for strings, blues and purples for woodwinds, orange and red for brass, and brown and black for percussion). He initially alternates sections between strings and flute, on the one hand, and single-reed and double-reed woodwinds, on the other, both occasionally punctuated with French horn

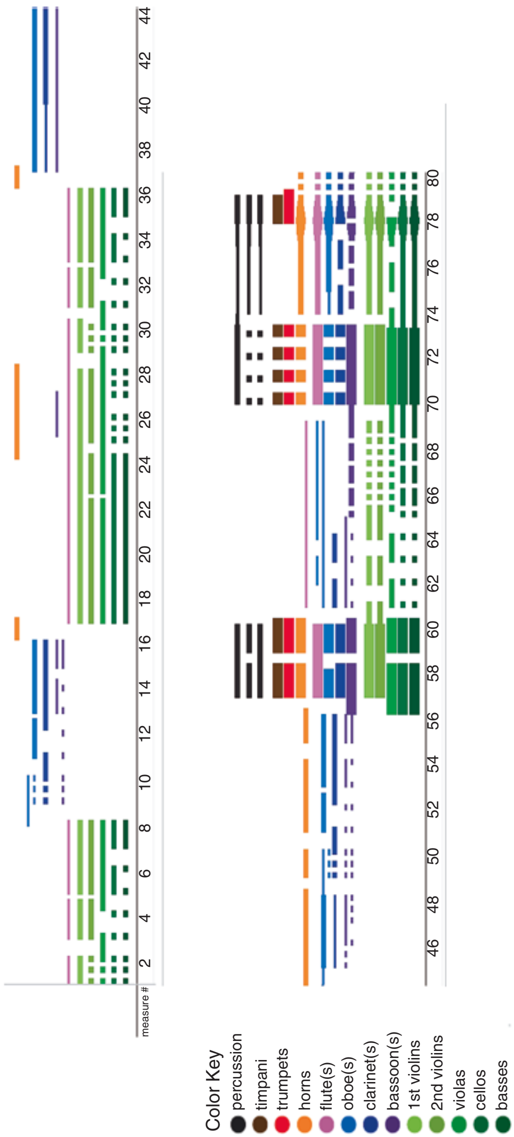

Fig. 8.8 Orchestral graph of Haydn's "Military" Symphony, no. 100, Movement II, up to measure 119. The line thickness represents the musical dynamics in the score (soft to loud). (From www.orchestralrevolution.com, figure 3.5; © 2018 by Emily Dolan, used with permission of the author)

and bassoon. Then in measure 56, a full orchestral tutti bursts in, given particular sonic power by trumpets, timpani and percussion, and the alternation between lesser and greater instrumental forces continues, which outlines the formal structure. The perceptual force of these alternations and interruptions could be tested with online segmentation techniques and perceptual characterization of the constituent sections.

Another aspect of large-scale orchestral shaping is what Goodchild et al. (2019) have called "orchestral gestures," such as the sudden contrast between the full orchestra and a soloist or a large-scale, swelling orchestral crescendo, which contribute to peak emotional experiences in orchestral music (Guhn et al. 2007). Some orchestration treatises mention large-scale gestures on the order of a few minutes. However, a clear taxonomy of techniques and a conceptual framework related to their musical function was lacking until Goodchild (2016) developed a typology of orchestral gestures in which the time course (gradual or sudden) and direction of change (additive or reductive) in instrumentation are the primary factors. She delineated four types: gradual addition, gradual reduction, sudden addition, and sudden reduction (schematized in Fig. 8.9). These gestures are characterized by changes over time in the number and type of instruments involved, as well as in onset density by instrument family, tempo, loudness, and spectral centroid. A visualization of one of the sudden reduction excerpts from the first movement of Bruckner's *Eighth Symphony* is shown in Fig. 8.10.

Goodchild et al.'s (2019) hypothesis was that extended patterns of textural and timbral evolution create orchestral gestures that possess a certain cohesiveness as cognitive units and have a goal-directed sense of motion. Such gestures often give rise to strong emotional experiences due to a confluence of change along many timbral dimensions, particularly timbral brightness as captured by the spectral centroid, but also changes in loudness, tempo, and registral extent (see upper panels in Fig. 8.10). Listeners' continuous ratings of emotional intensity were recorded while listening to excerpts from the ninteenth and twentieth century orchestral repertoire.

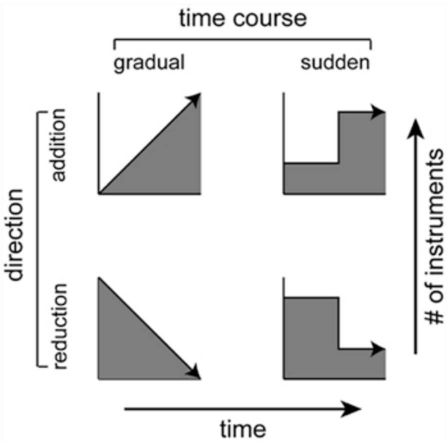

Fig. 8.9 The four types of orchestral gestures proposed by Goodchild et al. (2019). The gestures are categorized in terms of gradual or sudden change and the addition or reduction of instruments over time (From figure 1 in Goodchild et al. 2019; used with permission of Sage Publishing)

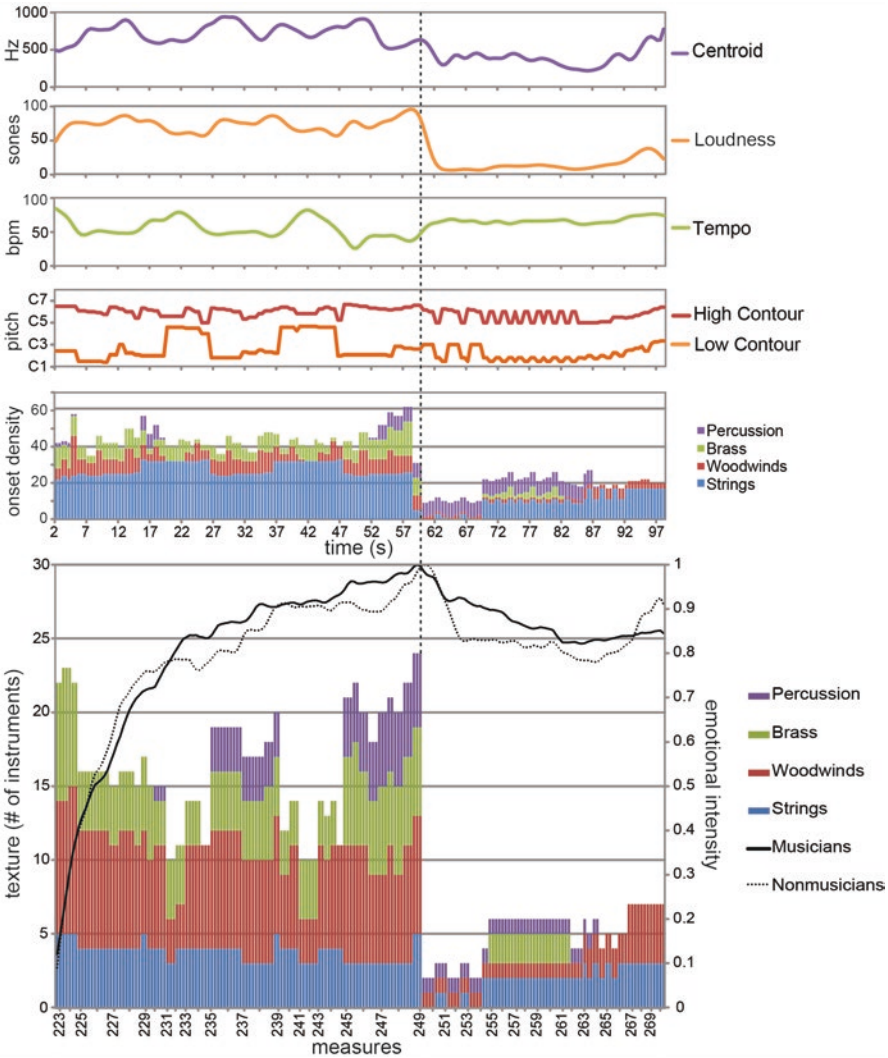

Fig. 8.10 Visualization of Bruckner's *Eighth Symphony*, first movement, measures 221–270. In the *upper panels*, spectral centroid (Hz), loudness (sones), tempo (in beats per minute), pitch range (ambitus), and onset density within instrument families are shown. The *bottom panel* graphs the instrumental texture, overlaid with the emotional intensity ratings for musician (*solid line*) and nonmusician (*dotted line*) listeners. In the bottom two panels, the *colors* represent the number of instruments of a given family that are involved. The *vertical dotted line* indicates the moment of sudden change in instrumental texture. (From figure A.11 in Goodchild 2016; used with permission of the author)

The data revealed different response profiles for each gestural type. For the gradual addition type, the emotional intensity ratings climbed steadily following the increasing growth of instrumental texture (number of instrumental parts) and loudness. For the sudden addition gestures, there was a slight tendency for musicians, but not nonmusicians, to anticipate the moment of sudden change with heightened emotional responses. Additional knowledge acquired through explicit musical training or greater experience with orchestral music may have led the musicians to develop anticipatory schemas for such changes. The responses to the gradual and sudden reductive excerpts featured a plateau of lingering high emotional intensity, despite the decrease of most musical features, including loudness, the spectral centroid (timbral brightness), instrumental texture, and onset density (number of attacks per beat). This response pattern is evident in Fig. 8.10 wherein the sudden reduction in instrumental forces at measure 250 is accompanied by a decrease in spectral centroid, loudness, and onset density along with a slight increase in tempo, and yet the average emotional intensity rating only makes a slight dip at that point.

Using re-orchestration and digital orchestral rendering as tools for testing hypotheses concerning the role of timbral brightness in emotional valence, Goodchild (2016) also showed with psychophysiological measures that the brightness of the orchestration (measured as spectral centroid) leading up to an expressive event dramatically shaped the resulting experience. Future research in this area could explore instances in which orchestral shaping (such as an abrupt change in texture or timbre) does or does not coordinate with other musical processes (such as phrase structure) to explain the interaction between formal structure based on melodic and harmonic elements and structure based on orchestration. The music-theoretical meanings and the resulting perceptual effects have yet to be explored, but this kind of work demonstrates the fertile ground that is possible in timbre cognition research through an interdisciplinary approach uniting music analysis and experimental psychology.

8.4.3 Timbre and Musical Tension

Larger-scale changes in timbre can also contribute to the expression of other higher-level structural functions in music, such as the ebb and flow of musical tension and relaxation, a type of process in music that many music theorists consider to be one of the primary bases for the perception of larger-scale form in music. When instruments composing a vertical sonority are strongly blended, timbral roughness and brightness become major components of musical tension. Nevertheless, they depend to a great degree on how the incoming acoustic information has been parsed into events and streams by auditory grouping processes. One might suppose that orchestration, in addition to pitch and rhythmic patterns, can play a major role in the structuring of musical tension and relaxation patterns that are an important component of a listener's aesthetic response to musical form. A feeling of tension accompanies a moment at which the music must continue, and a sense of relaxation signals the completion of the musical phrase or unit. In such cases, the structuring and sculpting of timbral changes and relations among complex auditory events provide myriad possibilities that composers have been exploring for decades in contemporary orchestral music but also in electroacoustic music (see Risset 2004). Musicologists have begun to address these issues, particularly as concerns timbre's potential role in what one might characterize as "musical syntax" (Roy 2003; Nattiez 2007), but psychologists have yet to tackle this area.

Experimental work on the role of harmony in the perception of musical tension and relaxation suggests that an important component of perceived tension is an attribute of timbre that is referred to as roughness (Bigand et al. 1996). The impression of timbral roughness seems to be based on the sensation of rapid fluctuations in the amplitude envelope that are correlated across peripheral auditory channels (Daniel and Weber 1997; Saitis and Weinzierl, Chap. 5). It can be generated by proximal frequency components that beat with one another. Dissonant intervals that generate an impression of roughness, like major sevenths (eleven semitones) and minor seconds (one semitone), tend to have more such beating than do consonant intervals such as octaves (twelve semitones) and fifths (seven semitones). As such, a fairly direct relation between sensory dissonance and timbral roughness has been demonstrated (cf. Plomp 1976; reviewed by Parncutt 1989).

To explore how timbre, through orchestration, might contribute to musical tension, Paraskeva and McAdams (1997) measured the effect of a change in orchestration on the inflection of tension and relaxation by comparing piano and orchestral versions of two pieces. Listeners were asked to make ratings based on the perceived degree of completion of the music at several points at which the music was stopped. What resulted was a completion profile (Fig. 8.11), which was used to infer musical tension by equating completion with release and lack of completion with tension. They tested two pieces: an excerpt from the six-voice fugue in the Ricercar from the *Musical Offering* by J. S. Bach (a tonal piece) and the first movement of the *Six Pieces for Orchestra, op. 6* by Webern (a nontonal piece, https://www.youtube.com/watch?v=NUCp4QvZxE8&frags=pl%2Cwn). Each piece was played both in an orchestral version (Webern's orchestration of the *Musical Offering* was used for the Bach piece; see link in Sect. 8.3.2) and in a direct transcription for piano of the original orchestral version of the Webern movement. Both versions were realized with a digital sampler to ensure that the performance nuances (timing, phrasing, etc.) were similar between the two. There were only very small differences between the completion profiles for musicians and nonmusicians, indicating that musical training didn't affect the completion ratings. Both tonal and atonal pieces produced significant fluctuations in musical tension, which is interesting given that some theorists feel that atonal music is devoid of this particular dimension of musical experience because it does not follow the standard tonal schemas (Lerdahl 1992). The important result here is that there were significant differences between the piano and orchestral versions, indicating an effect of timbre change on perceived musical tension. Notably, when the two versions *were* significantly different at a given stopping point (asterisks in Fig. 8.11), the orchestral version was always more relaxed than the piano version.

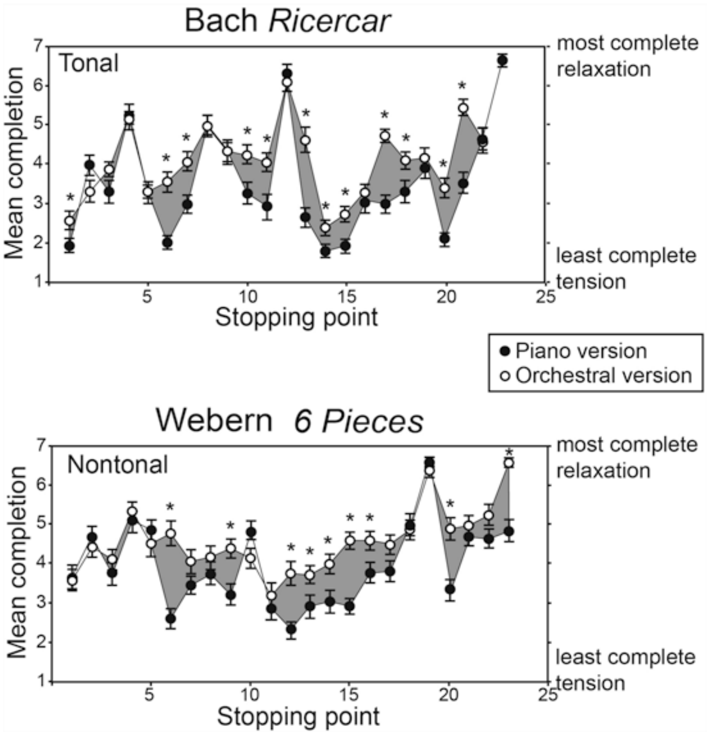

Fig. 8.11 Profiles of average completion ratings for piano and orchestral versions of tonal and nontonal pieces. The music was played to a certain point and stopped, at which point listeners made the completion rating. The next trial played the music from the beginning to the next stopping point. *Grey areas* highlight the differences between versions. *Asterisks* indicate significant differences between versions at a given stopping point. Error bars represent one standard error of the mean. (Adapted from Paraskeva and McAdams 1997; used with permission of the authors)

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 auditory roughness on the results of such processes. Wright and Bregman (1987), for example, illustrate several ways in which concurrent and sequential grouping processes interact and affect the perception of dissonance and tension in polyphonic music. 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 (Fig. 8.3). It may be that the same is true of sensory dissonance or auditory roughness, if we consider it to be a property of a concurrently grouped sound event. Piano sounds, being percussive in nature, have a rather sharp attack compared to most sounds from bowed and blown instruments. If several notes occur at the same time in the score and are played with a piano sound, they will be quite synchronous (particularly on a digital sampler). Because they all start at the same time and have similar amplitude envelopes and similar spectral distributions, 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 grouped together, although the effect of concurrent grouping on roughness perception does not seem to have been explicitly tested.

The situation is likely to be quite different for the orchestral version. For one, the same timing was used for piano and orchestra versions in the digital sampler. In the orchestral version, instruments with both slower and faster attacks were used. In other words, there is a greater range of attack times across wind and string instruments, depending on articulation, than would be the case with piano tones. Therefore, greater asynchrony could occur between the instruments in terms of perceived attack time, and the attack time difference is likely to reduce the perceptual fusion (Tardieu and McAdams 2012). Furthermore, the timbres of these instruments are often quite different. If several musical lines with different timbres arrive at the same moment on different pitches of a chord, the simultaneity may not be perceived as such because the listener may continue to track individual instruments sequentially in separate auditory streams on the basis of both the timbral similarity of notes from the same instrument and relative pitch proximity—what music theorists call voice leading (see Sect. 8.3.2).

Bregman and Pinker (1978) have demonstrated the interplay of concurrent fusion and sequential stream formation and conceived a sort of competition between the two auditory organization processes. Therefore, the attack asynchrony and the decomposition of simultaneities into separate auditory streams whose events are timbrally similar would work together to reduce the degree of perceptual fusion. A reduction in fusion would lead to greater segregation, and any 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 be much less than those of the piano version, which get grouped together more strongly. So once again, the perceptual effects of orchestration can have a very tight interaction with the processes of auditory scene analysis.

8.5 Reflections on Timbre and Musical Structure

Listeners are able to implicitly learn grammars built on rules for the probability of transition between timbres without explicit training. However, as Tillmann and McAdams (2004) demonstrated, listeners prefer certain relations among timbres that form coherent musical patterns and that distinguish among patterns. This result opens a vast field of possibilities for the construction of a veritable musical syntax based at least partially on timbre. For the syntactical use of timbre to have meaning in music, listeners must be able to learn rules of transition between timbres, as they do with durations and pitches. This learning has to be achieved implicitly by simply listening to the music without explicit training. Although the necessity of learning musical relations is obvious if one hopes to comprehend the resulting musical structures, the only explicit and experimentally controlled demonstrations of this capacity for timbre have been by Bigand et al. (1998) and Tillmann and McAdams (2004) (mentioned in Sect. 8.4.1). These findings raise the possibility of employing timbre as a primary parameter and structuring force in music.

Nattiez (2007) has critiqued Meyer's (1989) distinction between primary and secondary musical parameters and has questioned Meyer's relegation of timbre to secondary status. In Meyer's conception, primary parameters, such as pitch and duration, are able to carry syntax. (Meyer probably really meant inter-onset intervals, which define rhythm, rather than note duration, because duration per se is probably a secondary parameter related to articulation—staccato and legato.) According to this proposal, syntactic relations are based on implications for what follows next (expectations) and the realization (or not) of those implications, which is not possible with secondary parameters because they are not organized in discrete units or in clearly recognizable categories. Snyder (2000) proposes that we hear secondary parameters (including timbre) simply in terms of their relative amounts (on more of an ordinal scale), making them more useful for musical expression and nuance than for building grammatical structures.

Contrary to this position, Nattiez (2007) claims that timbre can be used to create syntactic relations that depend on expectations that lead to a perception of closure. He based his claim on his own analyses of Western and non-Western musical traditions, as well as Roy's (2003) analyses of electroacoustic music. Nattiez (2007) concluded that the main limit of Meyer's stance concerning timbre was that he confined his analyses to works composed in terms of pitch and rhythm (what current scholars of contemporary classical music call "pitch-based music"). In most cases in these styles of music, timbre is indeed only allowed to play a secondary functional role. Nattiez argued that timbre can be used to create syntactic relations that: (1) depend on expectations, leading to a perception of closure; or (2) are quite simply learned by a community of listeners as serving a given musical function within a system of hierarchical relations. He presented convincing cases supporting this hypothesis in analyses of the timbral structures in music as diverse as orchestral pieces by Debussy, Japanese drumming, and the throat singing tradition of Inuit women in northern Québec.

This debate recalls the distinction by composer and re-orchestrator John Rea between *prima facie* and *normative* orchestration (personal communication, October 26, 2011). Normative orchestration refers to situations in which the musical materials and structure are conceived in terms of pitch, harmony, duration, rhythm, and the formal structures based on them. Orchestration consists of highlighting, reinforcing, or cosmetically coloring these structures, although many orchestration decisions may be related to various programmatic topics (rapid string tremolos for storms, horn calls for the hunt or forest scenes, triumphant brass and percussion for military references) or emotional states (deep, dark melancholy of low cellos, bassoons and bass clarinets versus joyous country dance celebration with higher register woodwinds). Prima facie orchestration, to the contrary, concerns composition in which aspects of timbre are conceived at the outset as an integral part of the musical materials and forms. Examples from the electroacoustic music of Robert Normandeau, such as the piece *Tangram* (https://www.youtube.com/watch?v=KVB

RdbjQJbM&frags=pl%2Cwn), orchestral music such as *Polymorphia* by Krzysztof Penderecki (https://www.youtube.com/watch?v=9mYFKJBgxbM&frags=pl%2Cwn), or music that mixes acoustical instruments and computer-generated sounds such as *Archipelago* by Roger Reynolds are excellent examples to understand these possibilities. But even in the orchestral music of Haydn, Mozart, and Beethoven in the high Classical period, timbre plays a structuring role at the level of sectional segmentation induced by changes in instrumentation. These segmentations distinguish individual voices or orchestral layers that are composed of similar timbres and structure orchestral variations in symphonic forms (Dolan 2013).

In addition to contributing to the organization of auditory streams and orchestral layers, to contrasting materials that evoke segmentation at various levels of musical structure and to form large-scale orchestral gestures, timbre can also play a role in the ebb and flow of musical tension and relaxation and can thus contribute to the inherent expression of musical form as experienced by listeners. When instruments fuse into a musical sonority, the resulting auditory roughness, as an aspect of timbre, constitutes a major component of musical tension. However, perceived roughness strongly depends on the way the auditory grouping processes have parsed the acoustic information into events and streams (Wright and Bregman 1987) and also depends on the musical texture (homophony, polyphony, or heterophony) (Huron 2016). As a factor structuring tension and relaxation, timbre has been used effectively by electroacoustic composers such as Francis Dhomont. Roy's (2003) analyses of his music demonstrated that he employs timbre to build expectancies and deceptions in a musical context that is not "contaminated" by strong pitch structures. Roy's work implies that in a context in which pitch is a structuring factor, timbre may have trouble imposing itself as a dominant parameter as mentioned above. The interaction of musical parameters in the sculpting of the experience of musical form could be a vast and rich field if both perceptual experimentation and music analysis work together in an interdisciplinary setting to get at the essence of how orchestration—in the broadest sense of the choice, combination, and juxtaposition of sounds—actually works in the music of many styles and cultures.

A last point to consider for future experimental and musicological research on timbre concerns the crucial roles of performers and sound engineers in the final result of a sought-after timbre effect. Many parameters that affect both the timbre produced directly by an instrument (temporal and spectral properties of sound events) and the fusion of the sound of an instrument with those of other instruments (onset synchrony, pitch tuning, adjustment of timbre, and relative levels of instruments) are under the control of performers. In the end, all of these factors condition the timbre that emerges and how timbres connect sequentially and create segmental contrasts. Lembke et al. (2017), for example, showed that performers' success in achieving blend depends on both the possibilities of timbral modulation of the instrument itself (bassoon and French horn in their case, with the horn providing more room for timbral modulation) and what the role of each instrumentalist is in the musical scenario (leader or follower). Generally, followers who are trying to blend into the sound of a leader tend to darken the timbre of their instrument.

Beyond these effects, one might consider the role of sound recording and mixing, which intervene before the final result on analog or digital media. Along these lines, the composer John Rea described an experience he had in 1995 with a project of re-orchestrating Alban Berg's opera *Wozzeck* for an ensemble of twenty-one musicians in the place of the full orchestra specified by Berg (personal communication, October 26, 2011). He listened to five commercial recordings of *Wozzeck* precisely because, on the one hand, he wanted to hear how sounds fused and to see if the score presented this information in a particular way; on the other hand, he had to choose an ensemble of instruments to orchestrate the harmonies in order to best "re-present" the original score. He arrived at the following devastating conclusion: "The commercial recordings contribute to the dissimulation (the 'lie' if you will) that Art requires in order to carry on a discourse." There was indeed fusion, but it differed in each case, in each recording. It was clear that each conductor, but also each sound engineer or producer, had decided what was appropriate as a blend, as fusion, and as the projection of these qualities. Some performances of the passages in question were often in paradoxical contradiction with other performances/recordings of the same passages. To make interpretive Art always implies a confluence of several (at times conflictual?) sources of imagination and comprehension of the artistic goal.

Psychoacoustics and cognitive psychology can potentially reveal a large number of possibilities for the use of timbre in music. Composers may take profit from these scientific endeavors in the composition of their works. Music theorists and musicologists may explore, through analyses of orchestration in scores and recordings of notated music and in sound materials from electroacoustic works or field recordings of unnotated music, the ways in which composers and performers use timbre as a structuring force in music.

Acknowledgements The writing of this chapter was supported by grants from the Canadian Social Sciences and Humanities Research Council (890-2014-0008), the Fonds de Recherche du Québec—Société et Culture (SE-171434), the Canada Research Chairs Program (#950-223484), and the Killam Research Fellow Program of the Canada Council for the Arts.

Compliance with Ethics Requirements Stephen McAdams declares that he has no conflict of interest.

References

Adler S (2002) The study of orchestration, 3rd edn. W. W. Norton and Company, New York

Berlioz H, Strauss R (1948) Treatise on instrumentation (trans: Front T from 1904 German edn.). Edwin F. Kalmus, New York

Bey C, McAdams S (2003) Postrecognition of interleaved melodies as an indirect measure of auditory stream formation. J Exp Psychol Hum Percept Perform 29(2):267–279. https://doi.org/10.1037/0096-1523.29.2.267

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. Percept Psychophys 58(1):124–141. https://doi.org/10.3758/BF03205482

Bigand E, Perruchet P, Boyer M (1998) Implict learning of an artificial grammar of musical timbres. Curr Psychol Cogn 17(3):577–600
Bregman AS (1990) Auditory scene analysis: the perceptual organization of sound. MIT Press, Cambridge, MA
Bregman AS, Campbell J (1971) Primary auditory stream segregation and perception of order in rapid sequences of tones. J Exp Psychol 89(2):244–249. https://doi.org/10.1037/h0031163
Bregman AS, Pinker S (1978) Auditory streaming and the building of timbre. Can J Psychol Rev Can Psychol 32(1):19–31. https://doi.org/10.1037/h0081664
Cannon SC (2016) Altered dynamics and instrumentation at the onset of recapitulation in the nineteenth-century symphony. Analitica-Rivista online di studi musicali 8(1) http://www.gatm.it/analiticaojs/index.php/analitica/article/view/141
Culling JF, Darwin CJ (1993) The role of timbre in the segregation of simultaneous voices with intersecting Fo contours. Percept Psychophys 34(3):303–309. https://doi.org/10.3758/BF03205265
Cusack R, Roberts B (2000) Effects of differences in timbre on sequential grouping. Percept Psychophys 62(5):1112–1120. https://doi.org/10.3758/BF03212092
Daniel P, Weber R (1997) Psychoacoustical roughness: implementation of an optimized model. Acta Acustica united with Acustica 83(1):113–123
Deliège I (1987) Grouping conditions in listening to music: an approach to Lerdahl & Jackendoff's grouping preference rules. Music Percept 4(4):325–360. https://doi.org/10.2307/40285378
Deliège I (1989) A perceptual approach to contemporary musical forms. Contemp Music Rev 4(1):213–230. https://doi.org/10.1080/07494468900640301
Dolan EI (2013) The orchestral revolution: Haydn and the technologies of timbre. Cambridge University Press, Cambridge, UK
Ehresman D, Wessel D (1978) Perception of timbral analogies. Rapport IRCAM, vol 13. IRCAM-Centre Pompidou, Paris
Fales C, McAdams S (1994) The fusion and layering of noise and tone: implications for timbre in African instruments. Leonardo Music J 4:69–77. https://doi.org/10.2307/1513183
Goodchild M (2016) Orchestral gestures: music-theoretical perspectives and emotional responses. Thesis, McGill University, Montreal, QC. http://digitool.library.mcgill.ca/webclient/DeliveryManager?pid=141286&custom_att_2=direct
Goodchild M, McAdams S (2018) Perceptual processes in orchestration. In: Dolan E, Rehding A (eds) The Oxford handbook of timbre. Oxford University Press, New York. https://doi.org/10.1093/oxfordhb/9780190637224.013.10
Goodchild M, Wild J, McAdams S (2019) Exploring emotional responses to orchestral gestures. Musicae Scientiae 23(1):25–49. https://doi.org/10.1177/1029864917704033
Goodwin AW (1980) An acoustical study of individual voices in choral blend. J Res Music Educ 28(2):119–128. https://doi.org/10.1177/002242948002800205
Gregory AH (1994) Timbre and auditory streaming. Music Percept 12(2):161–174. https://doi.org/10.2307/40285649
Guhn M, Hamm A, Zentner M (2007) Physiological and musico-acoustic correlates of chill response. Music Percept 24(5):473–484. https://doi.org/10.1525/MP.2007.24.5.473
Hartmann M, Lartillot O, Toiviainen P (2016) Interaction features for prediction of perceptual segmentation: effects of musicianship and experimental task. J New Music Res 46(2):1–19. https://doi.org/10.1080/09298215.2016.1230137
Hartmann WM, Johnson D (1991) Stream segregation and peripheral channeling. Music Percept 9(2):155–183. https://doi.org/10.2307/40285527
Huang N, Elhilali M (2017) Auditory salience using natural soundscapes. J Acoust Soc Am 141(3):2163–2176. https://doi.org/10.1121/1.4979055
Huron D (2016) Voice leading: the science behind a musical art. MIT Press, Cambridge, MA
Iverson P (1995) Auditory stream segregation by musical timbre: effects of static and dynamic acoustic attributes. J Exp Psychol Hum Percept Perform 21(4):751–763. https://doi.org/10.1037//0096-1523.21.4.751
Iverson P, Krumhansl CL (1993) Isolating the dynamic attributes of musical timbre. J Acoust Soc Am 94(5):2595–2603. https://doi.org/10.1121/1.407371

Kendall R, Carterette EC (1993) Identification and blend of timbres as a basis for orchestration. Contemp Music Rev 9(1–2):51–67. https://doi.org/10.1080/07494469300640341
Krumhansl CL (1989) Why is musical timbre so hard to understand? In: Nielzén S, Olsson O (eds) Structure and perception of electroacoustic sound and music. Excerpta Medica, Amsterdam, pp 43–53
Krumhansl CL, Iverson P (1992) Perceptual interactions between musical pitch and timbre. J Exp Psychol Hum Percept Perform 18(3):739–751. https://doi.org/10.1037/0096-1523.18.3.739
Kurby CA, Zacks JM (2008) Segmentation in the perception and memory of events. Trends Cogn Sci 12(2):72–79. https://doi.org/10.1016/j.tics.2007.11.004
Lembke S, McAdams S (2015) The role of spectral-envelope characteristics in perceptual blending of wind-instrument sounds. Acta Acustica united with Acustica 101(5):1039–1051. https://doi.org/10.3813/AAA.918898
Lembke S-A, Levine S, McAdams S (2017) Blending between bassoon and horn players: an analysis of timbral adjustments during musical performance. Music Percept 35(2):144–164. https://doi.org/10.1525/mp.2017.35.2.144
Lerdahl F (1992) Cognitive constraints on compositional systems. Contemp Music Rev 6(2):97–121. https://doi.org/10.1080/07494469200640161
Lerdahl F, Jackendoff RS (1983) A generative theory of tonal music. MIT Press, Cambridge, MA
McAdams S (1984) The auditory image: a metaphor for musical and psychological research on auditory organization. In: Crozier WR, Chapman AJ (eds) Cognitive processes in the perception of art. North-Holland, Amsterdam, pp 289–323. https://doi.org/10.1016/S0166-4115(08)62356-0
McAdams S (1989) Psychological constraints on form-bearing dimensions in music. Contemp Music Rev 4(1):181–198. https://doi.org/10.1080/07494468900640281
McAdams S, Bregman AS (1979) Hearing musical streams. Comput Music J 3(4):26–43
McAdams S, Cunibile J-C (1992) Perception of timbral analogies. Philos T Roy Soc B 336(1278):383–389. https://doi.org/10.1098/Rstb.1992.0072
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. Psychol Res-Psych Fo 58(3):177–192. https://doi.org/10.1007/Bf00419633
McDermott JH, Lehr AJ, Oxenham AJ (2008) Is relative pitch specific to pitch? Psychol Sci 19(12):1263–1271. https://doi.org/10.1111/j.1467-9280.2008.02235.x
Meyer LB (1989) Style and music: theory, history, and ideology. University of Chicago Press, Chicago
Moore BCJ, Gockel H (2002) Factors influencing sequential stream segregation. Acta Acustica united with Acustica 88(3):320–332
Nattiez J-J (2007) Le timbre est-il un paramètre secondaire? [Is timbre a secondary parameter?]. Les Cahiers de la Société Québécoise de Recherche en Musique 9(1–2):13–24
Oliver H (1967) Structural functions of musical material in Webern's op. 6, no. 1. Perspectives of New Music 6(1):67–73. https://doi.org/10.2307/832407
Paraskeva S, McAdams S (1997) Influence of timbre, presence/absence of tonal hierarchy and musical training on the perception of musical tension and relaxation schemas. In: Rikakis T (ed) Proceedings of the 1997 International Computer Music Conference. Aristotle University, Thessaloniki, Greece, [CD-ROM]
Parncutt R (1989) Harmony: a psychoacoustical approach. Springer, Berlin
Plomp R (1976) Aspects of tone sensation: a psychophysical study. Academic, London
Reuter C (2003) Stream segregation and formant areas. In: Kopiez R, Lehmann R, Wolther AC, Wolf C (eds) Proceedings of the 5th triennial ESCOM conference. epOs-music, Osnabrück, pp 329–331
Rimsky-Korsakov N (1964) Principles of orchestration: with musical examples drawn from his own works (trans: Agate E from 1912 Russian ed.). Dover, New York
Risset J-C (2004) Le timbre [Timbre]. In: Nattiez J-J (ed) Musiques, une encyclopédie pour le XXIe siècle. Tome 2: Les savoirs musicaux. Actes Sud/Cité de la Musique, Arles, pp 134–161
Roy S (2003) L'analyse des musiques électroacoustiques: modèles et propositions [Analysis of electroacoustic musics: models and proposals]. L'Harmattan, Paris

Saffran JR, Johnson EK, Aslin RN, Newport EL (1999) Statistical learning of tone sequences by human infants and adults. Cognition 70(1):27–52. https://doi.org/10.1016/S0010-0277(98)00075-4

Saffran JR, Newport EL, Aslin RN (1996) Word segmentation: the role of distributional cues. J Mem Lang 35(4):606–621. https://doi.org/10.1006/jmla.1996.0032

Sandell GJ (1995) Roles for spectral centroid and other factors in determining “blended” instrument pairing in orchestration. Music Percept 13(2):209–246. https://doi.org/10.2307/40285694

Schoenberg A (1978) Theory of harmony (trans: Carter RE from 1911 original German publication). University of California Press, Berkeley

Siedenburg K, McAdams S (2018) Short-term recognition of timbre sequences: effects of musical training, pitch variability, and timbral similarity. Music Percept 36(1):24–39. https://doi.org/10.1525/MP.2018.36.1.24

Singh PG, Bregman AS (1997) The influence of different timbre attributes on the perceptual segregation on complex-tone sequences. J Acoust Soc Am 102(4):1943–1952. https://doi.org/10.1121/1.419688

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

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

Tardieu D, McAdams S (2012) Perception of dyads of impulsive and sustained instrument sounds. Music Percept 30(2):117–128. https://doi.org/10.1525/Mp.2012.30.2.117

Thoret E, Goodchild M, McAdams S (2018) Timbre 2018: timbre is a many-splendored thing. McGill University, Montreal. https://www.mcgill.ca/timbre2018/files/timbre2018/timbre2018_proceedings.pdf

Tillmann B, McAdams S (2004) Implicit learning of musical timbre sequences: statistical regularities confronted with acoustical (dis)similarities. J Exp Psychol Learn Mem Cogn 30(5):1131–1142. https://doi.org/10.1037/0278-7393.30.5.1131

Tougas Y, Bregman AS (1985) Crossing of auditory streams. J Exp Psychol Hum Percept Perform 11(6):788–798. https://doi.org/10.3758/BF03205976

Wessel DL (1979) Timbre space as a musical control structure. Comput Music J 3(2):45–52. https://doi.org/10.2307/3680283

Wolf EK (1966) The recapitulations in Haydn’s London symphonies. Music Q 52(1):71–79. https://doi.org/10.1093/mq/LII.1.71

Wright JK, Bregman AS (1987) Auditory stream segregation and the control of dissonance in polyphonic music. Contemp Music Rev 2(1):63–92. https://doi.org/10.1080/07494468708567054

Zahorik P, Brungart DS, Bronkhorst AW (2005) Auditory distance perception in humans: a summary of past and present research. Acta Acustica united with Acustica 91(3):409–420