CHAPTER 7

ACOUSTIC STRUCTURE AND MUSICAL FUNCTION: MUSICAL NOTES INFORMING AUDITORY RESEARCH

MICHAEL SCHUTZ

Introduction and Overview

Beethoven's Fifth Symphony has intrigued audiences for generations. In opening with a succinct statement of its four-note motive, Beethoven deftly lays the groundwork for hundreds of measures of musical development, manipulation, and exploration. Analyses of this symphony are legion (Schenker, 1971; Tovey, 1971), informing our understanding of the piece's structure and historical context, not to mention the human mind's fascination with repetition. In his intriguing book *The first four notes*, Matthew Guerrieri deconstructs the implications of this brief motive (2012), illustrating that great insight can be derived from an ostensibly limited grouping of just four notes. Extending that approach, this chapter takes an even more targeted focus, exploring how groupings related to the harmonic structure of individual notes lend insight into the acoustical and perceptual basis of music listening.

Extensive overviews of auditory perception and basic acoustical principles are readily available (Moore, 1997; Rossing, Moore, & Wheeler, 2013; Warren, 2013) discussing the structure of many sounds, including those important to music. Additionally, several texts now focus specifically on music perception and cognition (Dowling & Harwood, 1986; Tan, Pfordresher, & Harré, 2007; Thompson, 2009). Therefore this chapter focuses on a previously under-discussed topic within the subject of musical sounds—the importance of temporal changes in their perception. This aspect is easy to overlook, as the perceptual fusion of overtones makes it difficult to consciously recognize their individual contributions. Yet changes in the amplitudes of specific overtones excited by musical instruments as well as temporal changes in the relative strengths of those overtones play a crucial role in musical timbre. Western music has traditionally focused on properties such as pitch and rhythm, yet contemporary composers are increasingly interested in timbre, to the point where it can on occasion even serve as a composition's primary focus (Boulez, 1987; Hamberger, 2012). And although much previous scientific research on the neuroscience of music as well as music perception has focused on temporally invariant tones, there has been increasing recognition in the past decade that broadening our toolbox of stimuli is important to elucidating music's psychological and neurological basis. Consequently, understanding the role of temporal changes in musical notes holds important implications for psychologists, musicians, and neuroscientists alike.

Traditional musical scores give precise information regarding the intensity of each instrument throughout a composition in the form of dynamic markings. But for obvious practical reasons scores never specify the rapid intensity changes found in each overtone of an individual note. At most, composers hint at their preferences through descriptive terms such as "sharper/duller," vague instructions ("as if off in the distance"), and/or performers use stylistic considerations to make such decisions—e.g., by following period-specific performance practice. And to a large extent, both the harmonic structure of a note as well as changes in its harmonic structure over time are natural consequences of an instrument's physical structure. For example, the rapid decay of energy in harmonics shortly after the onset of a vibraphone note contrasts with the long sustain of its fundamental—contributing to its characteristic sound.

Musical notation clearly reflects changes in the intensity of collections of notes (e.g., *crescendos*, *sfz*) but never on the changes *within* notes themselves. While understandable, this decision mirrors the lack of attention to changes in overtone intensity in many psychophysical descriptions of sound—as well as perceptual experiments with auditory stimuli. This is unfortunate, as these intensity changes play an important role in efforts to synthesize "realistic" sounding musical notes—an issue of great relevance to composers creating electronic music. These also play an important role in discussions of tone quality so crucial to music educators training young ears, not to mention sound editors/engineers exploring which dynamic changes are important to capture and preserve when recording/mixing/compressing high quality audio. This chapter summarizes research on both the perceptual grouping of overtones and their rapid temporal changes, placing it in a broader context by highlighting connections to another important topic—how individual notes are perceptually grouped into chords. Finally, it concludes with a discussion of mounting evidence that auditory stimuli devoid of complex temporal changes may lead to experimental outcomes that fail to generalize to world listening—and on occasion can suggest errant theoretical frameworks and basic principles.

Grouping Notes: Deconstructing Chords and Harmonies

The vertical alignment of notes gives rise to musical harmonies ranging from lush to biting—from soothing to scary. Consequently, composers carefully design complex groupings whose musical effects hinge on small changes in their arrangement. For example, major and minor chords differ significantly in their neural processing (Pallesen et al., 2005; Suzuki et al., 2008) and evoke distinct affective responses (Eerola, Friberg, & Bresin, 2013; Heinlein, 1928; Hevner, 1935). Yet from the standpoint of acoustic structure this change is small—a half-step in the third (i.e., "middle note") of a musical chord (Aldwell, Schachter, & Cadwallader, 2002). In absolute terms, this represents a relatively small shift in the raw acoustic information—moving one of three notes the smallest permissible musical distance. From a raw acoustic perspective, this is particularly unremarkable in a richly orchestrated passage, yet the shift from major to minor can lead to significant changes in a passage's character. Individuals with cochlear implants—which offer relatively coarse pitch discrimination—are often unable to hear these distinctions, and often find music listening problematic (Wang et al., 2012). Fortunately most hear these changes quite readily, as evidenced by a literature on the detection of "out of key" notes shifted by a mere semi-tone (Koelsch & Friederici, 2003; Pallesen et al., 2005). Although musical acculturation occurring at a relatively young age (Corrigall & Trainor, 2010, 2014) aids this process, even musically untrained individuals are capable of detecting small changes (Schellenberg, 2002).

Notes of different pitch are often grouped together into a single musical object—a chord. Typically consisting of three or more individual notes, chords function as a "unit" and together lay out the harmonic framework or backbone of a musical passage. The specific selection of simultaneous notes (i.e., harmonically building chords) has profound effects on the listening experience of audiences, forming one of the key building blocks of strong physiological responses to music (Lowis, 2002; Sloboda, 1991). The masterful selection of notes, rhythms, and instruments requires both intuition and craft, and basic principles are articulated in numerous treatises on composition (Clough & Conley, 1984), and guidelines to orchestration (Alexander & Broughton, 2008; Rimsky-Korsakov, 1964). Yet another aspect of musical sound's vertical structure plays a crucial role in the listening experience, even if it is under less direct control by composers—the "vertical structure" (i.e., harmonic content) of individual notes—as well as the time-varying changes to these components. This topic forms the primary focus of this chapter, for much as study of individual notes can lend insight into our perception of musical passages, studying the rich, time-varying structure of concurrent harmonics can lend insight into our understanding of their perception.

Grouping Harmonics: Deconstructing Individual Notes

The complexities in composers' grouping of individual notes into chords are well known (Aldwell et al., 2002), yet the musical importance of individual harmonics is less transparent, even though single notes produced by musical instruments contain incredible sophistication and nuance (Hjortkjaer, 2013). Musical instruments produce sounds rich in overtones, which for pitched instruments generally consist of harmonics at integer multiples of the fundamental (Dowling & Harwood, 1986; Tan et al., 2010), as well as other non-harmonic energy (particularly during a sound's onset). The lawful structure of these overtones serves as an important binding cue, triggering a decision by the perceptual system to blend overtones such that "the listener is not usually directly aware of the separate harmonics" (Dowling & Harwood, 1986, p. 24). Although some musicians develop the ability to "hear out" individual components of their instruments (Jourdain, 1997, p. 35), in general this collection of frequencies fuses into a single musical unit. Consequently for practical matters the complex structure of individual notes is of less musical interest than the composer's complex selection of structural cues (Broze & Huron, 2013; Huron & Ollen, 2003; Patel & Daniele, 2003; Poon & Schutz, 2015), or the performer's interpretation of those cues (Chapin, Jantzen, Kelso, Steinberg, & Large, 2010).

Although the musical importance of small note-to-note variations in amplitude with respect to phrasing and expressivity (Bhatara, Tirovolas, Duan, Levy, & Levitin, 2011; Repp, 1995) is widely recognized, the small moment-to-moment amplitude variations in individual overtones have received less research attention. Musical sounds contain overtones shifting in their relative strength over time (Jourdain, 1997, p. 35), and some textbooks explicitly note the importance of these dynamic changes (Thompson, 2009, p. 59). Yet the role of spectra is often presented as time-invariant and described through summaries of spectral content irrespective of temporal changes in a note's spectra.

Musical instruments produce notes rich in temporal variation—not only in their overall amplitudes, but even with respect to the envelopes of individual harmonics. For example, Fig. 1 visualizes a musical note performed on the trumpet (left panel) and clarinet (right panel), based on instrument sounds provided by the University of Iowa Electronic Music studios (Fritts, 1997). The intensity (z axis) of energy extracted from each harmonic (x axis) is graphed over time (y axis). These 3D visualizations illustrate the temporal complexity of harmonics bound into the percept of a single note. In fact, divorced from its context in a melody, expressive timings in musical passages, discussion of performer's intentions regarding phrasing and numerous other considerations, analysis of isolated notes affords invaluable insight. Small temporal variations in each overtone play a key role in the degree to which synthesized notes sound "real" rather than "artificial." Highly trained musicians can routinely produce different variations on

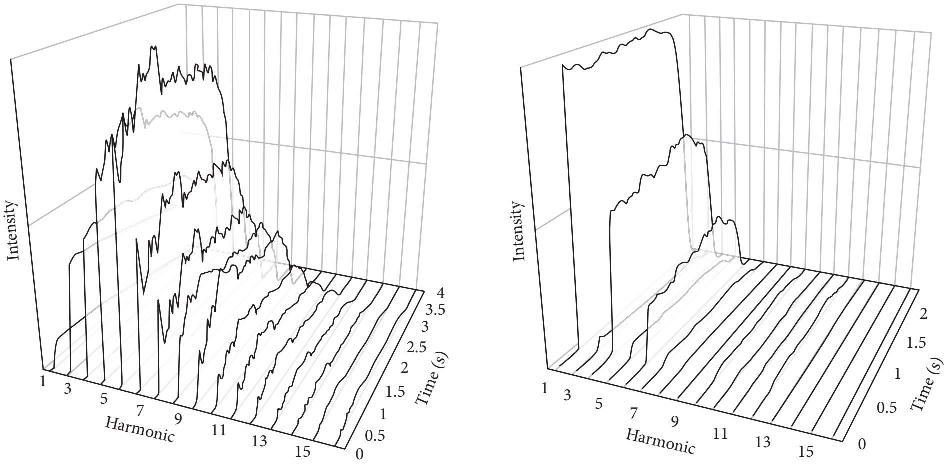

FIGURE 1. Visualization of single notes produced by a trumpet (left) and clarinet (right), illustrating their complex temporal structure. Although the trumpet spectrum changes more dynamically than the clarinet, each partial is in constant flux.

The goal of these 3D figures is to illustrate the dynamic nature of the harmonic structure of musical tones. Consequently they are not complete acoustical analyses (which are readily available elsewhere), but serve to highlight information lost in temporally invariant power spectra.

a single note ("brighter" or "more legato," "shimmery," etc.), which involve intentionally varying both the balance and temporal changes in a note's overtones.

As tones synthesized without adequate temporal changes often sound uninteresting or "fake," composers of electronic music, producers, instrument manufacturers, and other musical professionals pay top dollar for high quality audio samplings of instruments needed for their artistic purposes. Some creators of electronic music prefer samples of real musical sounds over efforts to synthesize these sounds (Risset & Wessel, 1999), in part due to the temporal complexity of accurately realizing the temporal changes in individual musical notes, as well as our sensitivity to small changes (or the lack thereof) in electronically generated tones. From a psychological perspective, what is so crucial about the structure of individual notes? What are the acoustic differences between life-like and dull renditions of individual instruments?

The importance of dynamic changes in an individual note's harmonics can be most usefully understood within the context of musical timbre—a complex, multidimensional property that has proven incredibly challenging to even define, let alone explain. Unfortunately for timbre enthusiasts, this property is often treated as a "miscellaneous category" (Dowling & Harwood, 1986, p. 63) accounting for the perceptual experience of "everything about a sound which is neither loudness nor pitch" (ANSI, 1994; Erickson, 1975). In other words, timbre is often defined less by what it *is* than what it *is not* (Risset & Wessel, 1999). This oppositional approach is sensible given the multitude of acoustic factors known to play a role in its perception (Caclin, McAdams, Smith, & Winsberg, 2005; McAdams, Winsberg, Donnadieu, de Soete, & Krimphoff, 1995).

Acoustic Structure and Musical Timbre

One particularly useful technique for studying musical timbre is multidimensional scaling (MDS), which allows for exploration absent of assumptions about which acoustic properties are most important. Many studies using this approach will present a variety of individual notes matched for pitch and intensity, asking participants to rate their similarity (or more often, dissimilarity). Analysis of dissimilarity ratings affords construction of a multidimensional space allowing for visualization of the "perceptual distance" between different pairs of notes. Early studies found spectral properties play a crucial role (Miller & Carterette, 1975), and subsequent work has refined our understanding of their role on both the neural (Tervaniemi, Schröger, Saher, & Näätänen, 2000) and perceptual (Grey & Gordon, 1978; Trehub, Endman, & Thorpe, 1990) levels. Consequently, the role of spectra in timbre is well explained in numerous textbooks on auditory perception and music cognition (Dowling & Harwood, 1986; Tan et al., 2010; Thompson, 2009, p. 48), typically through visualizations of power spectra, similar to Fig. 2.

Power spectra provide a useful, time-invariant summary of the relative harmonic strength. By collapsing along the temporal dimension shown in Fig. 1, Fig. 2 summarizes one of the characteristic distinctions between brass and woodwind instruments—that trumpets produce energy at all harmonics, whereas clarinets primarily emphasize alternate harmonics. Yet power spectra fail to capture the dynamic changes prominent in natural musical instruments, and the perceptual difference between synthesizing the information represented in Fig. 1 and Fig. 2 is striking. For interactive demonstrations of these differences, pedagogical tools useful for both teaching and research purposes are freely available from www.maplelab.net/pedagogy.

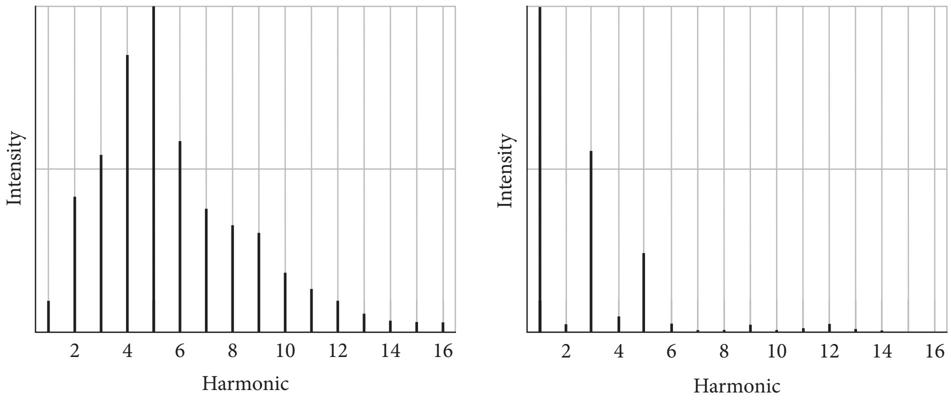

FIGURE 2. Power spectra of trumpet and clarinet. These plots accurately convey the trumpet's energy at many harmonics in contrast to the clarinet's energy primarily at odd numbered harmonics. However, power spectra fail to convey any information about the temporal changes in harmonic amplitude so crucial to a sound's timbre.

The shortcomings of power spectra are clear in cases where temporal cues play key roles not only in the realism of a musical sound, but in the distinction between different musical timbres. For example, the top row of Fig. 3 shows power spectra for notes produced on the trombone vs. cello.[1] This visual similarity in power spectra is somewhat surprising, given the markedly different methods of sound production in these instruments—a brass tube driven by lips on a mouthpiece vs. a bow drawn across a string. Additionally, cellos and trombones function differently in most musical compositions, suggesting their perception is distinct. Although this distinction is not apparent from their power spectra, it is clear in the middle row of Fig. 3 showing changes in harmonic strength over time. The bottom row provides a visualization of tones synthesized using the power spectra in the first row—illustrating what is retained and what is lost in time-invariant visualizations of musical sounds.

Certain aspects of temporal dynamics are recognized as playing an important role in musical timbre. For example, both the rise time (initial onset) of notes (Grey, 1977; Krimphoff, McAdams, & Winsberg, 1994) as well as gross temporal structure—amplitude envelope—have been shown to be important (Iverson & Krumhansl, 1993). As an extreme example, reversing the temporal structure of a note qualitatively changes its timbre, such that a piano note played "backwards" sounds more like a reed-organ than a piano (Houtsma, Rossing, & Wagennars, 1987). It is important to note that in this case the power spectra for piano notes played either forwards or backwards are identical—yet the experience of listening to these renditions differs markedly. Even beyond dramatic changes such as backwards listening, temporal changes are known to play an important role in sounds from natural instruments. However, interest in the connection between temporal dynamics and timbre has largely focused on a sound's *onset* (Gordon, 1987; Strong & Clark, 1967) rather than changes throughout its sustain period. For example, past studies have shown that insensitivity to a tone's onset correlates with reading deficits (Goswami, 2011). Tone onset is also crucial to distinguishing between musical timbres (Skarratt, Cole, & Gellatly, 2009), and their removal leads to confusion of instruments otherwise easily differentiable (Saldanha & Corso, 1964).[2]

The Use of Temporally Varying Sounds in Music Perception Research

Although temporal changes in the strengths of individual harmonics clearly play an important role in musical sounds, these changes are rightly recognized by experimental psychologists as potentially confounding (or at least introducing noise into) perceptual

[1] All analyses of notes in this chapter are based on additional samples from the University of Iowa Electronic Music studios (Fritts, 1997).

[2] However, presenting notes without transients as part of a melodic sequence (rather than as isolated tones) may mitigate this confusion (Kendall, 1986).

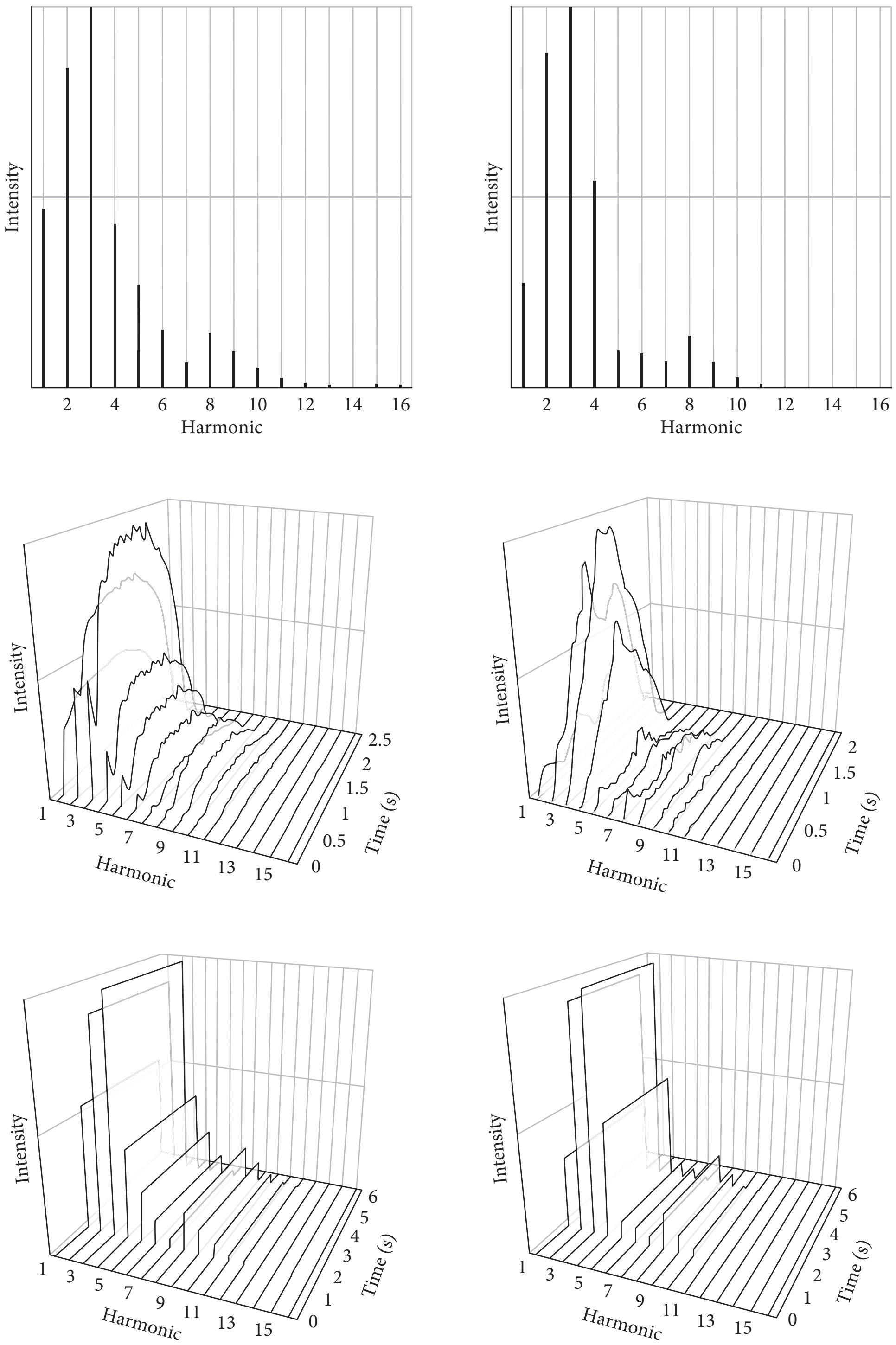

FIGURE 3. Visualizations of trombone (left) and cello (right). Panels in top row illustrate similarity in these instruments' power spectra, despite the clear acoustical differences shown in the middle panels. Bottom panels visualize tones synthesized using static power spectra (i.e., ignoring temporal changes in the strength of individual harmonics).

experiments. Not only will different instruments (along with variations in mouthpieces, mallets, bows, etc.) make consistency challenging when using natural musical tones, the complexity of changes in recordings of nominally steady-state notes runs contrary to the level of control desirable for scientific experimentation. If an experimenter's goal is to explore the role of pitch difference in auditory stream segregation, short pure tones with minimal amplitude variation offer clear benefits for drawing strong, replicable conclusions elucidating some aspects of our auditory perceptual organization. Consequently, the high degree of emphasis placed upon tightly constrained, easily reproducible stimuli incentivizes the use of simplified tones lacking temporal variation beyond simplistic onsets and offsets. This raises important questions about what kinds of stimuli are used to assess auditory perception. Although simplified sounds aid researchers in avoiding problematic confounds, their over-use could lead to challenges with generalizing their findings to natural sounds with the kinds of temporal variations shown in Fig. 1.

In order to explore the kinds of sounds used in research on music perception, my team surveyed 118 empirical papers published in the journal *Music Perception* from experiments dating back to its inception in 1983, based on a previous comprehensive bibliometric survey (Tirovolas & Levitin, 2011). Primarily interested in determining the amount of amplitude variation found in the temporal structures of auditory stimuli, we classified every stimulus used in each of the 212 surveyed experiments as either "flat" (i.e., lacking temporal variation), "percussive" (decaying notes such as those produced by the piano, cowbell, or marimba), or "other"—sounds such as those produced by sustained instruments like the French horn or human voice. Fig. 4 illustrates examples of each stimulus class.

The most surprising outcome from this survey was that although most articles included a wealth of technical information on spectral structure, duration, and the exact model of headphones or speakers used to present the stimuli, about 35 percent failed to define the stimuli's temporal structure. This finding is not unique to *Music Perception*—my team found similar problems with under-specification in the journal *Attention, Perception & Psychophysics* (Gillard & Schutz, 2013). More important than under-specification, both surveys revealed a strong bias against sounds with the kinds of temporal variations common to musical instruments. Although flat tones lend themselves well to tight experimental control and consistent replication amongst different labs, they fail to capture the richness of the sounds forming the backbone of the musical listening experience. Yet they remain prominent in a wide range of research on auditory perception on tasks purportedly designed to illuminate generalizable principles of auditory perception.

Prominent researchers have noted that the world is "[not] replete with examples of naturally occurring auditory pedestals [i.e., flat amplitude envelopes]" (Phillips, Hall, & Boehnke, 2002, p. 199). Yet flat tones appear to be the normative approach to research on auditory perception, which are clearly far removed from the complexity of natural musical sounds—as shown in Fig. 5. Note that each of the three musical instruments visualized not only exhibits constant temporal changes, but temporal changes in the

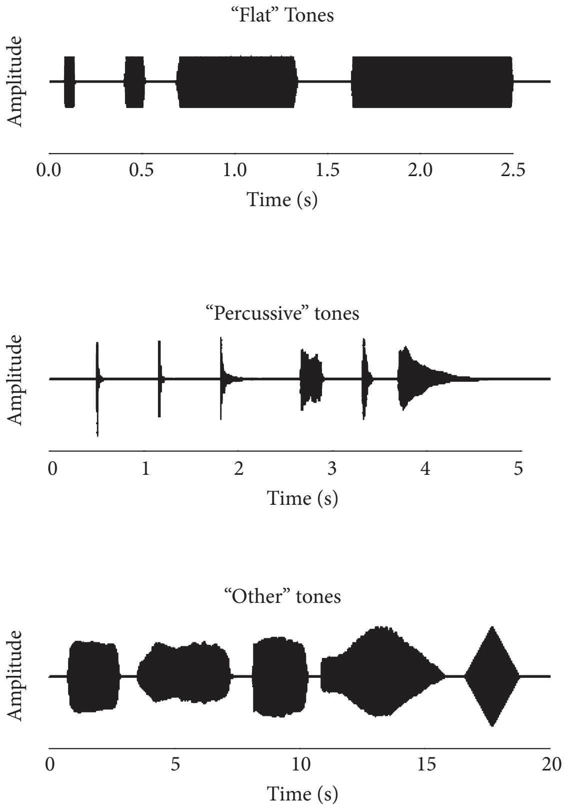

FIGURE 4. Wave forms of different sounds found in the survey of stimuli used in *Music Perception* (Schutz & Vaisberg, 2014).

Reproduced from *Music Perception: An Interdisciplinary Journal 31*(3), Michael Schutz and Jonathan M. Vaisberg, Surveying the temporal structure of sounds used in music perception, pp. 288–296, doi:10.1525/mp.2014.31.3.288, Copyright © 2014, The Regents of the University of California.

amplitudes of each individual harmonic. This dynamic fluctuation contrasts starkly with the flat tones favored in auditory perception research shown in the bottom right panel. This over-fixation on sounds lacking meaningful amplitude variation is not confined to behavioral work; a large-scale review of auditory neuroscience research concluded with a note of caution that important properties of functions of the auditory system will only be fully understood when researchers begin employing envelopes that "involve modulation in ways that are closer to real-world tasks faced by the auditory system" (Joris, Schreiner, & Rees, 2004, p. 570). The acoustic distance between the temporally dynamic musical sounds and temporally constrained flat tones common in auditory perception and neuroscience research raises important questions about the degree to which theories and models derived from these experiments generalize to musical listening. The complexities of balancing competing needs for experimental control and ecological relevance are significant, and will serve as the focus of the following section.

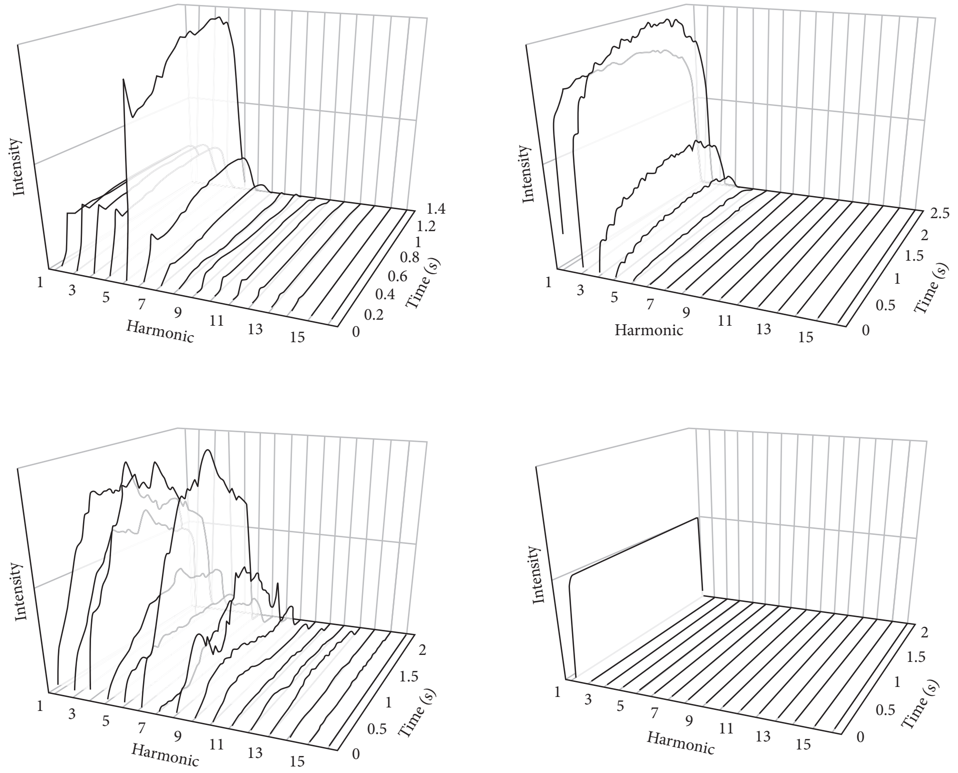

FIGURE 5. Single notes produced by an oboe (upper left), French horn (upper right), and viola (lower left) illustrate their temporal complexity. Although their specific mix of harmonics varies, these instruments all exhibit constant changes in the strength of each harmonic over the tone's duration. This temporal complexity contrasts strongly with the temporal simplicity of the flat tone depicted in the lower right panel, which lacks temporal variation beyond abrupt onsets/offsets, and no change in relative strength of harmonics.

On the Methodological Convenience of Simplified Sounds

This focus on tightly constrained stimuli is not necessarily problematic; control of extraneous variables is essential to researchers' ability to draw strong conclusions from individual experiments. Consistency in the synthesis of stimuli amongst different labs holds many advantages with respect to replication, an issue of increasing importance to the field as a whole. And in some circumstances the real-world associations inherent in temporally complex sounds can pose obstacles to answering key questions. For example, researchers exploring acoustic attributes of unpleasant sounds illustrate that frequency range (Kumar, Forster, Bailey, & Griffiths, 2008), spectral "roughness" (Terhardt, 1974), and the relative mix of harmonics-to-noise (Ferrand, 2002) are key factors—issues important for engineers designing human–computer auditory interfaces. Yet a direct ranking of sounds shows that vomiting is regarded as one of the most unpleasant (Cox, 2008), an outcome related less to its specific *acoustic properties* than the obvious *real-world associations* (McDermott, 2012). In some cases these real-world associations may be regarded as confounds obfuscating the general principles at hand.

Therefore, in some inquiries aimed at understanding the relationship between acoustic structure and perceptual response, it is not only reasonable but actually necessary to use sounds devoid of referents. This issue of disentangling the effects attributable to associations vs. acoustic features is of particular importance in the perception of music, given the rich and complex relationship between music, memory, and emotion. Familiar compositions can evoke memories as a result of past associations—for example from a history of personal listening/performance (Schulkind, Hennis, & Rubin, 1999) or use in film sound tracks (e.g., those used by Vuoskoski and Eerola, 2012). Indeed songs from popular television shows are so familiar they have even been used to assess the pervasiveness of absolute pitch amongst the general population (Schellenberg & Trehub, 2003). Consequently, synthesized tones lacking real-world associations serve a useful purpose in advancing our understanding of auditory perception.

However, although artificial sounds devoid of real-world associations that afford precise control/replication offer advantages in certain circumstances, their simplicity can pose barriers to fully understanding music perception. In fact, auditory psychophysics' focus on "control" (Neuhoff, 2004) and the study of isolated parameters absent their natural context (Gaver, 1993) is an issue of long-standing concern in some corners of the auditory perception community. This is of particular importance to understanding music, as composers, performers, conductors, and recording engineers focus great attention to slight nuances of musical timbre. Yet the same differences so useful in artistic creation often serve as confounds within the realm of auditory psychophysics. This raises important questions about the types of stimuli that should be used in experiments designed to address questions related to music listening. Can artificial sounds abstracted from our day-to-day musical experiences lead to experimental outcomes that generalize to listening outside the laboratory?

Perceptual experiments exploring audio-visual integration in musical contexts offer a useful case study in the consequences of ignoring the role of musical sounds' dynamic temporal structures. A large body of audio-visual integration research using temporally simplistic sounds has concluded that vision rarely influences auditory evaluations of duration[3] (Fendrich & Corballis, 2001; Walker & Scott, 1981; Welch & Warren, 1980). However, a musical experiment exploring ongoing debate amongst percussionists led to a surprising break with widely accepted theory. In that series of studies an internationally acclaimed musician attempted to create long and short notes on the marimba—a tuned, wooden bar instrument similar to the xylophone. Notes on the marimba are percussive (Fig. 4, middle panel)—with continuous temporal variation in their structure

[3] Provided that the acoustic information is of sufficient quality (Alais & Burr, 2004; Ernst & Banks, 2002).

as the energy transferred into the bar (by striking) gradually dissipates as a result of friction, air resistance, etc. Whether or not the duration of these notes can be intentionally varied has been long debated in the percussion community (Schutz & Manning, 2012). However, an assessment of an expert percussionist's ability to control note duration demonstrated that these gestures are in fact acoustically inconsequential, but trigger an *illusion* in which the longer physical gesture used to strike the instrument affects perception of the resulting note's duration (Schutz & Lipscomb, 2007). Musical implications (Schutz, 2008) aside, this finding represents a clear break from previously accepted views on the integration of sight and sound (Fendrich & Corballis, 2001; Walker & Scott, 1981; Welch & Warren, 1980).

The surprising ability of percussionists to shape perceived note duration despite previous experimental work to the contrary stems in large part from a bias in the temporal structure of stimuli used in auditory research. Subsequent experiments illustrate that movements derived from the percussionists' gesture (Schutz & Kubovy, 2009b) integrate with sounds exhibiting decaying envelopes (e.g., piano notes, produced from the impact of a hammer on string), but failed to integrate with the sustained tones produced by the clarinet or French horn (Schutz & Kubovy, 2009a). As the clarinet differs in many properties from the marimba and piano, a direct test of temporal structure using pure tones (i.e., sine waves) shaped with decaying vs. amplitude invariant amplitude envelopes found visual information integrated with the temporally dynamic percussive tones, but not the temporally invariant flat tones previously used in audio-visual integration experiments (Schutz, 2009).

This distinction between the outcomes of experiments with tones using temporally dynamic vs. static amplitude envelopes is important in assessing the degree to which lab-based tasks inform our understanding of listening in the real world. For example, temporal structure can play a key role in the well-known audio-visual bounce effect (ABE), in which two circles approach each other, overlap, and then move to their original starting point. Although this ambiguous display can be perceived as depicting circles either "bouncing off" or "passing through" one another, a brief tone coincident with the moment of overlap enhances the likelihood of seeing a bounce (Sekuler, Sekuler, & Lau, 1997). However, not all sounds affect this integrated perception in the same way. Sounds synthesized with decaying envelopes mimicking impact events trigger significantly more bounce percepts than their mirror images (Grassi & Casco, 2009). The temporal structure of individual tones also plays a role in a variety of "general" perceptual tasks assessed primarily using tones lacking dynamic temporal changes, leading to different experimental outcomes in tasks ranging from learning associations (Schutz, Stefanucci, Baum, & Roth, 2017) to perceiving pitches (Neuhoff & McBeath, 1996), assessing event duration (Vallet, Shore, & Schutz, 2014), and segmenting auditory streams (Iverson, 1995).

Overlooking the importance of temporal structure in auditory perception can even lead to misguided theoretical claims used to inform ongoing research programs. For example, as discussed previously a great deal of audio-visual integration research involves temporally simplified tones ensuring experimental control. However, interest in the role of the natural connection between sight and sound has been considered in discussions regarding the "unity assumption" (Welch, 1999) and/or "identity decision" (Bedford, 2004). That research explores the idea that event unity between sight and sound plays an important role in the binding decision, such that stimuli perceived as "going together" are more likely to bind. For example, in the well-known "ventriloquist effect" the sound of a ventriloquist's voice is perceptually bound with concurrent lip movements of their puppets (Abry, Cathiard, Robert-Ribes, & Schwartz, 1994; Bonath et al., 2007). Unfortunately, the natural real-world relationships between sights and sounds often pose challenges for the controlled manipulations so important to experimental research. For example, tightly controlled, psychophysically inspired studies of multimodal speech help clarify the importance of event unity in multisensory integration. Gender matched faces and voices—the sound of a male producing syllable paired with the lip movements of either male or female articulating that syllable—bind more strongly than gender mis-matched faces and voices (Vatakis & Spence, 2007). This finding offers strong evidence for the unity assumption raising important questions about the degree to which it applies to auditory stimuli beyond speech.

A series of experiments assessing the role of the unity assumption with musical stimuli involved pairing the sound of a piano note and plucked guitar string with video recordings of the movements used to produce these sounds. Following their earlier procedures, this approach found no evidence of the unity assumption playing a role in this non-speech musical task (as well as other stimuli such as a hammer striking ice vs. a bouncing ball). This outcome contributed to the conclusion that the unity assumption applied only to speech stimuli (Vatakis, Ghazanfar, & Spence, 2008). However, as summarized below, subsequent research found strong evidence for the unity assumption in non-speech tasks—considering the importance of auditory temporal structure.

The piano and guitar sounds used by Vatakis et al. (2008) exhibited similar amplitude envelopes—a property defining the gross temporal structure of a sound (i.e., the summation of changes in the amplitudes of spectral components). Building upon their approaches to assessing binding using musical notes produced by the marimba and cello, my team found evidence for the unity assumption when assessing sounds that involved clearly differentiable amplitude envelopes (Chuen & Schutz, 2016). Although in hindsight, the traditional focus on flat tones in auditory psychophysics research helped obfuscate the obvious similarity in temporal structure of the guitar and piano notes used by Vatakis et al. (2008). Given the relatively small proportion of auditory perception studies using natural sounds, this oversight is understandable as the use of natural sounds in psychophysics experiments is laudable given the general focus on temporally invariant stimuli, which "often seems to have limited direct relevance for understanding the ability to recognize the nature of complex natural acoustic source events" (Pastore, Flint, Gaston, & Solomon, 2008, p. 13).

From these examples, it is clear that the time-varying structure of natural sounds (or lack thereof) can meaningfully influence the outcomes of psychological experiments.

This is true whether researchers' goals are to explore natural listening or attempting to better understand the theoretical structure and function of the auditory system. This issue holds important implications even for experiments aimed at elucidating generalized principles of perceptual processing rather than explicitly assessing the role of dynamic temporal changes. Together, these concerns are consistent with those raised previously by proponents of ecological acoustics such as John Neuhoff, who argue that "the perception of dynamic, ecologically valid stimuli is not predicted well by the results of many traditional experiments using static stimuli" (2004, p. 5).

Conclusions

Traditional studies of specific sequences of notes such as the four note opening of Beethoven's Fifth Symphony provide useful insight into both the theoretical structure of musical passages, as well as their larger cultural relevance. Much as the constant movement of pitches and rhythms gives rise to lively melodies, the continual variations in temporal structure (for multiple simultaneous harmonics) play an important role in musical listening. However, as this information is not notated in musical scores and is often under-emphasized in scientific discourse, the importance of these dynamic changes is not always fully recognized. This "insight" is well understood amongst those involved in sound synthesis and virtual modeling of musical instruments. However, the need for tight experimental control for stimuli used in experimental work on auditory perception and auditory neuroscience has incentivized the use of simple time-invariant flat tones. Although they offer important methodological benefits, their distance from musical sounds can pose limitations on their ability to inform our understanding of natural listening. With modern recording and sound synthesis approaches we now have the ability to generate auditory stimuli exhibiting the rich temporal variation of natural musical sounds, while also affording the precise control so crucial for avoiding confounds—raising exciting new possibilities for future innovation and discovery. Looking toward the future, research assessing core questions of auditory perception using temporally complex sounds will help clarify the degree to which existing theories and models apply to our perception of natural sounds such as those produced by musical instruments.

Acknowledgments

Funding supporting this research was provided by the Natural Sciences and Engineering Research Council of Canada (NSERC), Social Sciences and Humanities Research Council of Canada (SSHRC), and the Ontario Early Researcher Award (ERA). I would like to thank Maxwell Ng for his assistance in creating the visualizations of the instrument sounds used throughout this chapter.

References

Abry, C., Cathiard, M. A., Robert-Ribes, J., & Schwartz, J. L. (1994). The coherence of speech in audio-visual integration. *Current Psychology of Cognition 13*, 52–59.

Acoustical Society of America Standards Secretariat (1994). Acoustical Terminology ANSI S1.1–1994 (ASA 111-1994). *American National Standard*. ANSI/Acoustical Society of America.

Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology 14*(3), 257–262.

Aldwell, E., Schachter, C., & Cadwallader, A. (2002). *Harmony & voice leading* (3rd ed.). Boston, MA: Schirmer.

Alexander, P. L., & Broughton, B. (2008). *Professional orchestration: The first key. Solo instruments & instrumentation note, volume 1* (3rd ed.). Petersburg, VA: Alexander Publishing.

Bedford, F. L. (2004). Analysis of a constraint on perception, cognition, and development: One object, one place, one time. *Journal of Experimental Psychology: Human Perception and Performance 30*(5), 907–912.

Bhatara, A., Tirovolas, A. K., Duan, L. M., Levy, B., & Levitin, D. J. (2011). Perception of emotional expression in musical performance. *Journal of Experimental Psychology: Human Perception and Performance 37*(3), 921–934.

Bonath, B., Noesselt, T., Martinez, A., Mishra, J., Schwiecker, K., Heinze, H.-J., & Hillyard, S. A. (2007). Neural basis of the ventriloquist illusion. *Current Biology 17*(19), 1697–1703.

Boulez, P. (1987). Timbre and composition—timbre and language. *Contemporary Music Review 2*(1), 161–171.

Broze, Y., & Huron, D. (2013). Is higher music faster? Pitch–speed relationships in Western compositions. *Music Perception: An Interdisciplinary Journal 31*(1) 19–31.

Caclin, A., McAdams, S., Smith, B. K., & Winsberg, S. (2005). Acoustic correlates of timbre space dimensions: A confirmatory study using synthetic tones. *Journal of the Acoustical Society of America 118*(1), 471–482.

Chapin, H., Jantzen, K., Kelso, J. A. S., Steinberg, F., & Large, E. W. (2010). Dynamic emotional and neural responses to music depend on performance expression and listener experience. *PLoS ONE* 5, 1–14.

Chuen, L., & Schutz, M. (2016). The unity assumption facilitates cross-modal binding of musical, non-speech stimuli: The role of spectral and amplitude cues. *Attention, Perception, & Psychophysics 78*(5), 1512–1528.

Clough, J., & Conley, J. (1984). *Basic harmonic progressions*. New York: W. W. Norton.

Corrigall, K. A., & Trainor, L. J. (2010). Musical enculturation in preschool children: Acquisition of key and harmonic knowledge. *Music Perception: An Interdisciplinary Journal 28*(2), 195–200.

Corrigall, K. A., & Trainor, L. J. (2014). Enculturation to musical pitch structure in young children: Evidence from behavioral and electrophysiological methods. *Developmental Science 17*(1), 142–158.

Cox, T. J. (2008). Scraping sounds and disgusting noises. *Applied Acoustics 69*(12), 1195–1204.

Dowling, W. J., & Harwood, D. L. (1986). *Music cognition*. Orlando, FL: Academic Press.

Eerola, T., Friberg, A., & Bresin, R. (2013). Emotional expression in music: Contribution, linearity, and additivity of primary musical cues. *Frontiers in Psychology 4*, 1–12. Retrieved from https://doi.org/10.3389/fpsyg.2013.00487

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

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature 415*(6870), 429–433.

Fendrich, R., & Corballis, P. M. (2001). The temporal cross-capture of audition and vision. *Perception & Psychophysics 63*(4), 719–725.

Ferrand, C. T. (2002). Harmonics-to-noise ratio: An index of vocal aging. *Journal of Voice 16*(4), 480–487.

Fritts, L. (1997). *University of Iowa Electronic Music Studios. University of Iowa*. Retrieved from http://theremin.music.uiowa.edu/MIS.html

Gaver, W. (1993). What in the world do we hear? An ecological approach to auditory event perception. *Ecological Psychology 5*(1) 1–29.

Gillard, J., & Schutz, M. (2013). The importance of amplitude envelope: Surveying the temporal structure of sounds in perceptual research. In *Proceedings of the Sound and Music Computing Conference* (pp. 62–68). Stockholm, Sweden.

Gordon, J. W. (1987). The perceptual attack time of musical tones. *Journal of the Acoustical Society of America 82*(1) 88–105.

Goswami, U. (2011). A temporal sampling framework for developmental dyslexia. *Trends in Cognitive Sciences 15*(1) 3–10.

Grassi, M., & Casco, C. (2009). Audiovisual bounce-inducing effect: Attention alone does not explain why the discs are bouncing. *Journal of Experimental Psychology: Human Perception and Performance 35*(1), 235–243.

Grey, J. M. (1977). Multidimensional perceptual scaling of musical timbres. *Journal of the Acoustical Society of America 61*(5), 1270–1277.

Grey, J. M., & Gordon, J. W. (1978). Perceptual effects of spectral modifications on musical timbres. *Journal of the Acoustical Society of America 63*(5), 1493–1500.

Guerrieri, M. (2012). *The first four notes: Beethoven's Fifth and the human imagination*. New York: Alfred A. Knopf.

Hamberger, C. L. (2012). *The evolution of Schoenberg's Klangfarbenmelodie: The importance of timbre in modern music*. The Pennsylvania State University. Retrieved from https://etda.libraries.psu.edu/files/final_submissions/8130

Heinlein, C. P. (1928). The affective characters of the major and minor modes in music. *Journal of Comparative Psychology 8*, 101–142.

Hevner, K. (1935). The affective character of the major and minor modes in music. *American Journal of Psychology 47*(1), 103–118.

Hjortkjaer, J. (2013). The musical brain. In J. O. Lauring (Ed.), *An introduction to neuroaesthetics: The neuroscientific approach to aesthetic experience, artistic creativity, and arts appreciation* (pp. 211–244). Copenhagen: Museum Tusculanum Press.

Houtsma, A. J. M., Rossing, T. D., & Wagennars, W. M. (1987). Auditory demonstrations on compact disc. *Journal of the Acoustical Society of America*. New York: Acoustical Society of America/Eindhoven: Institute for Perception Research.

Huron, D., & Ollen, J. (2003). Agogic contrast in French and English themes: Further support for Patel and Daniele (2003). *Music Perception: An Interdisciplinary Journal 21*(2), 267–271.

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

Iverson, P., & Krumhansl, C. L. (1993). Isolating the dynamic attributes of musical timbre. *Journal of the Acoustical Society of America 94*, 2594–2603.

Joris, P. X., Schreiner, C. E., & Rees, A. (2004). Neural processing of amplitude-modulated sounds. *Physiological Reviews 84*, 541–577.

Jourdain, R. (1997). *Music, the brain, and ecstasy: How music captures our imagination*. New York: William Morrow and Company.

Kendall, R. A. (1986). The role of acoustic signal partitions in listener categorization of musical phrases. *Music Perception 4*(2), 185–213.

Koelsch, S., & Friederici, A. D. (2003). Toward the neural basis of processing structure in music. *Annals of the New York Academy of Sciences 999*, 15–28.

Krimphoff, J., McAdams, S., & Winsberg, S. (1994). Caractérisation du timbre des sons complexes. II. Analyses acoustiques et quantification psychophysique. *Journal de Physique IV Colloque 4*, 625–628.

Kumar, S., Forster, H. M., Bailey, P., & Griffiths, T. D. (2008). Mapping unpleasantness of sounds to their auditory representation. *Journal of the Acoustical Society of America 124*(6), 3810–3817.

Lowis, M. J. (2002). Music as a trigger for peak experiences among a college staff population. *Creativity Research Journal 14*(3–4), 351–359.

McAdams, S., Winsberg, S., Donnadieu, S., de Soete, G., & Krimphoff, J. (1995). Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes. *Psychological Research 58*(3), 177–192.

McDermott, J. (2012). Auditory preferences and aesthetics: Music, voices, and everyday sounds. In R. J. Dolan & T. Sharot (Eds.), *Neuroscience of preference and choice: Cognitive and neural mechanisms* (pp. 227–257). London: Academic Press.

Miller, J. R., & Carterette, E. C. (1975). Perceptual space for musical structures. *Journal of the Acoustical Society of America 58*(3), 711–720.

Moore, B. C. J. (1997). *An introduction to the psychology of hearing* (4th ed.). London: Academic Press.

Neuhoff, J. G. (2004). *Ecological psychoacoustics* (J. G. Neuhoff, Ed.). Amsterdam: Elsevier/Academic Press.

Neuhoff, J. G., & McBeath, M. K. (1996). The Doppler illusion: The influence of dynamic intensity change on perceived pitch. *Journal of Experimental Psychology: Human Perception and Performance 22*(4), 970–985.

Pallesen, K. J., Brattico, E., Bailey, C., Korvenoja, A., Koivisto, J., Gjedde, A., & Carlson, S. (2005). Emotion processing of major, minor, and dissonant chords: A functional magnetic resonance imaging study. *Annals of the New York Academy of Sciences 1060*, 450–453.

Pastore, R. E., Flint, J., Gaston, J. R., & Solomon, M. J. (2008). Auditory event perception: The source–perception loop for posture in human gait. *Perception & Psychophysics 70*(1), 13–29.

Patel, A. D., & Daniele, J. R. (2003). Stress-timed vs. syllable-timed music? A comment on Huron and Ollen (2003). *Music Perception: An Interdisciplinary Journal 21*(2), 273–276.

Phillips, D. P., Hall, S. E., & Boehnke, S. E. (2002). Central auditory onset responses, and temporal asymmetries in auditory perception. *Hearing Research 167*(1–2), 192–205.

Poon, M., & Schutz, M. (2015). Cueing musical emotions: An empirical analysis of 24-piece sets by Bach and Chopin documents parallels with emotional speech. *Frontiers in Psychology 6*, 1–13. Retrieved from https://doi.org/10.3389/fpsyg.2015.01419

Repp, B. H. (1995). Quantitative effects of global tempo on expressive timing in music performance: Some perceptual evidence. *Music Perception: An Interdisciplinary Journal 13*(1), 39–57.

Rimsky-Korsakov, N. (1964). *Principles of orchestration* (M. Steinberg, Ed.). New York: Dover.

Risset, J.-C., & Wessel, D. L. (1999). Exploration of timbre by analysis and synthesis. In D. Deutsch (Ed.), *The Psychology of Music* (pp. 113–169). San Diego, CA: Gulf Professional Publishing.

Rossing, T. D., Moore, R. F., & Wheeler, P. A. (2013). *The science of sound* (3rd ed.). London: Pearson Education.

Saldanha, E. L., & Corso, J. F. (1964). Timbre cues and the identification of musical instruments. *Journal of the Acoustical Society of America 36*(11), 2021–2026.

Schellenberg, E. G. (2002). Asymmetries in the discrimination of musical intervals: Going out-of-tune is more noticeable than going in-tune musical intervals. *Music Perception: An Interdisciplinary Journal 19*(2), 223–248.

Schellenberg, E. G., & Trehub, S. E. (2003). Good pitch memory is widespread. *Psychological Science 14*(3), 262–266.

Schenker, H. (1971). Analysis of the first movement. In E. Forbes (Ed.), *Beethoven Symphony No. 5 in C minor* (pp. 164–182). New York: W. W. Norton.

Schulkind, M. D., Hennis, L. K., & Rubin, D. C. (1999). Music, emotion, and autobiographical memory: They're playing your song. *Memory & Cognition 27*(6), 948–955.

Schutz, M. (2008). Seeing music? What musicians need to know about vision. *Empirical Musicology Review 3*(3), 83–108.

Schutz, M. (2009). *Crossmodal integration: The search for unity* (Dissertation). University of Virginia.

Schutz, M., & Kubovy, M. (2009a). Causality and cross-modal integration. *Journal of Experimental Psychology: Human Perception and Performance 35*(6), 1791–1810.

Schutz, M., & Kubovy, M. (2009b). Deconstructing a musical illusion: Point-light representations capture salient properties of impact motions. *Canadian Acoustics 37*(1), 23–28.

Schutz, M., & Lipscomb, S. (2007). Hearing gestures, seeing music: Vision influences perceived tone duration. *Perception 36*(6), 888–897.

Schutz, M., & Manning, F. (2012). Looking beyond the score: The musical role of percussionists' ancillary gestures. *Music Theory Online 18*, 1–14.

Schutz, M., Stefanucci, J., Baum, S. H., & Roth, A. (2017). Name that percussive tune: Associative memory and amplitude envelope. *Quarterly Journal of Experimental Psychology 70*(7), 1323–1343.

Schutz, M., & Vaisberg, J. M. (2014). Surveying the temporal structure of sounds used in music perception. *Music Perception: An Interdisciplinary Journal 31*(3), 288–296.

Sekuler, R., Sekuler, A. B., & Lau, R. (1997). Sound alters visual motion perception. *Nature 385*(6614), 308.

Skarratt, P. A., Cole, G. G., & Gellatly, A. R. H. (2009). Prioritization of looming and receding objects: Equal slopes, different intercepts. *Attention, Perception, & Psychophysics 71*(4), 964–970.

Sloboda, J. (1991). Music structure and emotional response: Some empirical findings. *Psychology of Music 19*(2), 110–120.

Strong, W., & Clark, M. (1967). Perturbations of synthetic orchestral wind-instrument tones. *Journal of the Acoustical Society of America 41*(2), 277–285.

Suzuki, M., Okamura, N., Kawachi, Y., Tashiro, M., Arao, H., Hoshishiba, T., . . . Yanai, K. (2008). Discrete cortical regions associated with the musical beauty of major and minor chords. *Cognitive, Affective, & Behavioral Neuroscience 8*(2), 126–131.

Tan, S.-L., Pfordresher, P. Q., & Harré, R. (2007). *Psychology of music: From sound to significance*. New York: Psychology Press.

Terhardt, E. (1974). On the perception of periodic sound fluctuations (roughness). *Acta Acustica United with Acustica 30*, 201–213.

Tervaniemi, M., Schröger, E., Saher, M., & Näätänen, R. (2000). Effects of spectral complexity and sound duration on automatic complex-sound pitch processing in humans: A mismatch negativity study. *Neuroscience Letters 290*, 66–70.

Thompson, W. F. (2009). *Music, thought, and feeling: Understanding the psychology of music*. New York: Oxford University Press.

Tirovolas, A. K., & Levitin, D. J. (2011). Music perception and cognition research from 1983 to 2010: A categorical and bibliometric analysis of empirical articles in *Music Perception*. *Music Perception: An Interdisciplinary Journal 29*(1), 23–36.

Tovey, D. F. (1971). The Fifth Symphony. In E. Forbes (Ed.), *Beethoven Symphony No. 5 in C minor* (pp. 143–150). New York: W. W. Norton.

Trehub, S. E., Endman, M. W., & Thorpe, L. A. (1990). Infants' perception of timbre: Classification of complex tones by spectral structure. *Journal of Experimental Child Psychology 49*(2), 300–313.

Vallet, G., Shore, D. I., & Schutz, M. (2014). Exploring the role of amplitude envelope in duration estimation. *Perception 43*(7), 616–630.

Vatakis, A., Ghazanfar, A. A., & Spence, C. (2008). Facilitation of multisensory integration by the "unity effect" reveals that speech is special. *Journal of Vision 8*(9), 1–11.

Vatakis, A., & Spence, C. (2007). Crossmodal binding: Evaluating the "unity assumption" using audiovisual speech stimuli. *Perception & Psychophysics 69*(5), 744–756.

Vuoskoski, J. K., & Eerola, T. (2012). Can sad music really make you sad? Indirect measures of affective states induced by music and autobiographical memories. *Psychology of Aesthetics, Creativity, and the Arts 6*, 1–10.

Walker, J. T., & Scott, K. J. (1981). Auditory-visual conflicts in the perceived duration of lights, tones and gaps. *Journal of Experimental Psychology: Human Perception and Performance 7*(6), 1327–1339.

Wang, S., Liu, B., Dong, R., Zhou, Y., Li, J., Qi, B., . . . Zhang, L. (2012). Music and lexical tone perception in Chinese adult cochlear implant users. *The Laryngoscope 122*, 1353–1360.

Warren, R. M. (2013). *Auditory perception: A new synthesis*. Amsterdam: Elsevier.

Welch, R. B. (1999). Meaning, attention, and the "unity assumption" in the intersensory bias of spatial and temporal perceptions. In G. Aschersleben, T. Bachmann, & J. Musseler (Eds.), *Cognitive contributions to the perception of spatial and temporal events* (pp. 371–387). Amsterdam: Elsevier.

Welch, R. B., & Warren, D. H. (1980). Immediate perceptual response to intersensory discrepancy. *Psychological Bulletin 88*(3), 638–667.