THE UNIVERSITY OF CHICAGO

PERCEPTION, PITCH, AND MUSICAL CHORDS

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE HUMANITIES
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

DEPARTMENT OF MUSIC

BY
MARK JEROME YEARY

CHICAGO, ILLINOIS
AUGUST 2011

To my family

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGMENTS

This dissertation would not have been possible without the assistance of a number of people. The continuous support and insight offered by my adviser, Lawrence Zbikowski, helped me transform a potpourri of concepts, theories, and proposals into a coherent monograph. The chalkboard sessions that Larry and I conducted in the early and middle stages of my research were essential to developing the conceptual framework of the dissertation, and his nearly constant attention to the details of this work was critical to bringing it to its present state.

My interests in the issues underlying this dissertation may be traced back to the seminars and discussions I had with faculty and colleagues at the University of California, Santa Barbara. Pieter van den Toorn provided a foundation of scholarship and discourse that was invaluable to my development as a theorist; it was in his seminars that I was introduced to the writings of Carl Dahlhaus, the use of perception and cognition in musical analysis, and the works of Igor Stravinsky. The history of theory seminars of Lee Rothfarb provided an introduction to the *Tonpsychologie* of Hermann von Helmholtz and Carl Stumpf that continues to inform my research. Jonathan Pearl took me under his wing during my time at UCSB, and his interdisciplinary approach to music scholarship demonstrated how I might forge a similar path. Lastly, discussions with Timothy Cooley and Stefanie Tcharos were essential to my early development as a music scholar; my work with them as members of the Center for Interdisciplinary Study of Music (CISM) introduced me to the broad scholarly approach that is the cornerstone of my continuing scholarship, and both Stefanie and Tim went out of their way to offer support and guidance as I made the transition to Chicago.

I have benefited greatly from working with the faculty of the University of Chicago Department of Music. The early music research of scholars present and past stoked my interest in the pre-modern and early modern Western practices that continue to be a fruitful resource for my research; I would like to recognize the seminars of Robert Kendrick, Thomas Christensen, Martha Feldman, Anne Walters Robertson, and Margaret Bent as being specifically influential. Seminars with Berthold Hoeckner, Travis Jackson, and Philip Gossett kept me connected with the more recent practices that inspired my study of chords. Steven Rings and Thomas Christensen played essential roles in reviewing the proposal that would eventually lead to this dissertation; their insight and suggestions allows me to identify the core issues that drove my later research.

As my work took an increasingly empirical direction following the proposal, Larry suggested that I discuss my ideas with Howard Nusbaum, whose continuing support was invaluable as I came up to speed with the principles and methods of cognitive psychology. The members of Howard's lab have always been very welcoming, and their guidance and feedback was instrumental in the design and execution of the experiments that form a part of this dissertation. I would like to give thanks and recognition in particular to Shannon Heald for giving me a crash course in statistics for psychology, and for her overall guidance throughout my empirical work; to Serena Klos and Peter Hu for their assistance with the software and equipment used during these experiments; to Steven Hedger for his specifically musical insight into experimental design and procedures; to Olivier Lescop for many discussions on signal processing and the use of MATLAB; and to David Casey and Bruno Cabral for their assistance with running the experimental trials—and for taking over the reins during my week of "family leave" following the birth of my second daughter.

The Franke Institute for the Humanities offered valuable support during the later stages of my research. Our periodic meetings of Franke Fellows were a productive venue for discussion topics ranging across the humanities, and my writing has benefited greatly from the comments of my colleagues outside the music department. The graduate workshop on the Perception and Understanding of Music provided a similarly ideal environment for presenting the materials that would eventually make up the first two chapters of the dissertation. Many of the people involved with forming and sustaining this workshop have been mentioned above; I would like to thank in particular Jonathan De Souza for coordinating the workshop, and Berthold Hoeckner for his comments and suggestions both within and outside of the workshop setting.

The camaraderie of my fellow graduate students was an essential counterbalance to the isolating tendencies of dissertation writing. It was reassuring to know that, at whatever time I decided to set up my mobile office in the Regenstein Library, I would find at least one other student plugging away at his or her proposal, presentation, or dissertation.

Lastly, on a more personal note, I would like to recognize the enduring support and encouragement from my wife Sharon and our daughters Ellen and Sarah. As much as I look to frame my research within the human experience, these efforts are trivial when compared to the personal development that comes with being a husband and father. I love you all; Dad's book is finally done.

ABSTRACT

In this dissertation, I argue that hearing a musical chord—a simultaneity of two or more notes perceived as a single object—is perceptually different from hearing separate concurrent tones, and that the object status of chords shapes our experience of listening to harmonic music. The listener's tendency to perceive a given sonority as a chord is affected by the presence of salient acoustic patterns within the sonority, by the musical context in which the sonority appears, and by his or her experience with listening to harmonic music.

Following an outline of the acoustic and contextual cues that promote chordal listening, I offer a series of performance strategies based on these cues that maximize the likeliness of hearing a sonority as a chord. I then argue that these strategies played a role in the development of the Western practice of harmonic tonality, and that the design and use of polyphonic instruments in the late Renaissance period enabled many of these strategies to be applied within musical practice. A further investigation of contextual and experience-based factors in chord perception is conducted in a pair of experiments, in which the listener is asked to recognize or "hear out" a tone from within a three-tone sonority: the first experiment uses familiar and unfamiliar sonorities to examine the effect of listening experience on this task, whereas the second measures the effect of different harmonic textures presented as a pre-sonority context.

A listener who perceives a sonority as a chord is better able to perceive its emergent features, which are defined as properties of the whole that are not necessarily properties of its parts. I examine the emergent feature of pitch—a familiar property of the musical tone in both perceptual and theoretical descriptions—using the virtual pitch model proposed by Ernst Terhardt, and I outline the conditions in which a listener might perceive a chord as bearing an emergent pitch. An analysis of the opening sonority of Igor Stravinsky's *Symphony of Psalms* gives an example of how chord pitch may be used as a compositional resource. Drawing upon the conclusions of this analysis, I suggest how further research on perceiving chords' emergent features—in particular the perceptual correlate of the music-theoretical concept of chord quality—could be applied to develop a more complete understanding of how we experience chords.

CHAPTER 1

THE OBJECT AND THE CHORD

Carl Dahlhaus begins his survey of emergent harmonic tonality—the dominant Western musical practice of the past four centuries—with the following caveat:

> It would be futile [*vergeblich*] to attempt the separate definition of such basic concepts of tonal harmony as "chord" or "*basse fondamentale*," or to name specific criteria by which one could determine whether a sonority is or is not a chord. For terms like "chord" and "*basse fondamentale*" do not designate objective facts that one can point to in a musical score. Rather, these terms denote cofactors in a particular mode of musical perception [*musikalische Hörweise*], factors that receive their full meaning only in relation to other factors. (Dahlhaus 1968/1990, 67)

In making this leap into musical listening, Dahlhaus emphasizes the need to understand the thoughts and categories that gave rise to the elements of harmonic tonality—namely, the chord.[1] Identifying a stack of notes within a musical score as "a chord" is unfounded without a corresponding account of one's experience of hearing the simultaneity. To claim a series of sonorities as chords, continues Dahlhaus, is to claim that a hypothetical listener hears these sonorities "not as resultants—as combinations of tones and intervals—but as directly perceived unities [*unmittelbar gegebene Einheiten*]" (Dahlhaus 1968/1990, 67).

The refinement that Dahlhaus proposes—defining the chord by conceptualization as well as notation—brings with it a reduced ability to objectively define the terms and concepts that we use to describe music: the chord, a staple of musical analyses and

---

[1] Dahlhaus situates his search for the origins harmonic tonality within the tradition of Besseler (1950) and Lowinsky (1961), though without characterizing his efforts as superior based on his contextual approach.

harmony textbooks, becomes a subjective experience that resists definition according to strictly music-analytical terms. We may understand Dahlhaus's reference to the "futility" of defining the chord as stressing the need to consider the musical and cultural contexts that are intertwined with the historical repertory: calling a given sonority an "inversion," for instance, presupposes that the musicians creating and hearing this sonority would have found the concept of inversions useful or applicable in some way.

I would like to propose that Dahlhaus's statement also points to a second sort of futility: the lack of objective access to a listener's perceptual responses suggests that even a careful consideration of treatises and concepts would be insufficient to describe the experience of musical listening that we claim as an essential element of harmonic tonality. Yet Dahlhaus's resignation coincides historically with the cusp of a renewed research interest in the ways humans perceive and process information in the world (Miller 1960; Gibson 1966; Neisser 1967; Rosch et al. 1976). And with respect to the capacities that might help us determine whether a sonority is or is not a chord, more recent work on auditory perception (van Noorden 1975; Bregman 1990) provides a foundation from which we may begin to peer into the mind (and ear) of the listener.

In this dissertation I challenge the futility of defining the chord by developing a perceptually informed framework for conceptualizing sonorities. I propose a model for hearing chords that draws up recent research in psychoacoustics, auditory perception, cognition, and musical listening; this model is then examined in my own empirical contribution. In turn, I use this model to propose an experiential account of chords: how perceiving a chord differs from hearing a simultaneity of notes or tones, and how the experience of hearing chords contributes to the "way of hearing" (*Hörweise*) that we adopt in listening to chordal music.[2]

1.1 Defining the chord

The central claim of this dissertation is that a chord is a single perceptual entity in musical practice. Hearing a chord as *unmittelbar*—immediate or direct—implies that we do not assemble the chord *ex post facto* from a collection of separately perceived tones; rather, the chord appears as a Gestalt, an integral entity that, in the words of one writer, forms "the primary foundation of my experience" (Wertheimer 1923/1938, 5).

This claim requires that I refine the commonly understood definition of chord. Calling a chord "the simultaneous sounding of two or more notes" (*Grove music online*, s.v. "Chord") does not specify the way in which this simultaneity is experienced—as a single unified entity, or as individual notes related to one another by the concept "chord." I define the **chord**,[3] therefore, as a simultaneity of two or more notes that are heard as a single musical entity. In place of the definition of chord cited above—as multiple simultaneous notes—I use the term **sonority**; a chord is a sonority, but a sonority need not be heard as a chord.

Although redefining a familiar term such as "chord" may seem unnecessarily confusing, I propose that this more refined definition is already present within various musical practices within Western culture. In describing a performance of rock music, for

---

[2] The perceptual bent of Dahlhaus's term *Hörweise* is later taken up by Clarke (2005), who draws upon the ecologically based perception theories of Gibson (1966) to examine modern "ways of listening."

[3] Terms in **boldface** are collected and defined in a glossary at the end of the dissertation.

instance, we might say that the guitarist plays chords, whereas the backup singers sing "harmonies." Although both guitarist and singers may make use of the same sonorities common to harmonic tonality (such as major and minor triads), we differentiate the sounds they produce based on our experience of hearing (and seeing) a performance; whereas the singers' "harmonies" are heard as a particular blend of separate sounds, the chord is heard as a single "thing." To locate this chordal "thing" within our listening experience, I turn to the concept of the perceptual object.

1.2 Perceptual and auditory objects

A **perceptual object** is a mental representation or image of some part of the environment, composed of sensory information (Griffiths and Warren 2004). The need to parse the environment into objects is apparent when considering what William James called the "blooming, buzzing confusion" (James 1890, 488) of a newborn's perceptual experience: we learn to form objects as a means of experiencing, organizing, and categorizing the massive amount of incoming information or **stimuli** available in a typical environment. The concept of objects as perceptual operations—as something we *do* to sensory information—is essential to our modern understanding of perception, specifically in the field of categorization (Rosch et al. 1976; Tversky and Hemenway 1984): how we make decisions about our environment depends upon how we parse the environment into objects.

An **auditory object**, consequently, is a perceptual object formed of auditory sensations. Although the concept of auditory objects is a foundation of recent hearing research, it has previously appeared under the guise of several different terms, such as auditory image (McAdams 1984), auditory stream (Bregman 1990), or auditory event (Rosenblum 2004). I adopt the term auditory object in this dissertation because it facilitates connections and comparisons with another common form of perceptual object—the visual object (Adams and Janata 2002; Kubovy and Schutz 2010). These connections in turn permit the adaptation of research in visual perception—such as the concept of Gestalt touched on in the previous section—to further define and situate the auditory object within our perceptual experiences.

We commonly associate auditory objects with their physical sources: if we hear a "car crash" auditory object, we are quick to associate the object with a real-world event, and we may turn our heads toward the sound source in anticipation of further actions or events (Neisser 1976, chapter 2). But we may also experience a car crash auditory object in the absence of a physical car crash: we may hear a recorded version of a car crash, played over a pair of speakers, and derive the same object from its surrounding sound stimuli. Furthermore, we may later imagine or "simulate" (Barsalou 1999) this car crash object, in the absence of the material stimuli that prompted our initial percept.

Indeed, any attempt to identify an auditory object as such must go beyond external (objective) reference: what one listener hears as "a sound" may constitute several sounds for another listener, and a portion of "a sound" for a third listener. Claims of objecthood, therefore, must be grounded in our understanding of how, why, and under what conditions we form objects. Accordingly, I propose that there are three broad factors to be considered in our search for the acoustic object. Although these factors may not necessarily be sympathetically aligned within a given listening scenario, we may explore each in turn to suggest which factors may be most influential within the listener's object-forming processes.

1.2.1 Acoustics

Auditory objects are formed of acoustic stimuli that are received by our **auditory system**—broadly defined as the collection of mental and physical entities, from the outer ear to the brain, that enable the perception of sounds. Stimuli are organized according to the dimensions of our auditory system; within these dimensions, we perceive features that enable object formation.

Dimensions are the media in which our perceptual systems to organize and differentiate stimuli; the dimensions of auditory perception are *frequency* and *time* (Van Valkenburg and Kubovy 2003). We commonly form objects according to perceived boundaries or discontinuities within these dimensions: where multiple stimuli overlap in dimensions, it is difficult to separate the stimuli into multiple objects. With auditory objects in particular, these boundaries are typically manifested as changes in **intensity** (amplitude) in the physical sound stimulus, such that we might examine the dimensions of "the object" with the aid of sound visualization equipment (Griffiths and Warren 2004).

Features are patterns of stimuli for which our perceptual systems have acquired distinct detection abilities (Treisman and Gelade 1980; Treisman 1993). The Gestalt laws of perception (Wertheimer 1923/1938) may be understood as laws of feature detection: we perceive stimuli based on the features they afford, and we form objects based on the consistency and alignment of features that we detect within the sound stimuli. Common features of auditory objects include pitch, timbre, loudness, and spatial location.

1.2.2 Context

Objects are formed within a particular **context** or environment. Whereas the term context may refer to both external (environmental) and internal (cognitive) elements, we will use the term **domain** to emphasize the necessity of the cognitive aspect.

A domain is an environment or context with which we associate one or more perceptual tasks; it is closely related to the concepts of schema (Rumelhart 1980) and script (Schank and Abelson 1977), in that a domain implies a set of expected behaviors or events that are likely to occur. We form objects that are appropriate to the tasks or goals of a given domain: two people having a conversation will form objects appropriate to a speech domain, such as words or phrases, but a third person trying to ignore the conversation may simply form a single "background speech" object.

1.2.3 Experience

Objects are formed according to our **experience**—the conscious and unconscious learning and conditioning that follows our perceptual actions. We are likely to form objects that correspond with our most common actions (Tversky and Hemenway 1984)—or, to frame this relationship more causally, our common actions shape how we form objects. A listener who hears the sounds of a revving automobile engine is likely to perceive "engine noise" as a single object; an auto mechanic hearing the same stimulus has learned to differentiate the sound of a loose timing belt, and hears it as a distinct "timing belt" object. In this regard, different listeners may draw upon different domains to shape their percepts: the mechanic's "auto repair" domain is learned through training, such that it is unavailable to listeners without similar experience.

Experience permits the listener to use his or her **attention** to assist in forming objects. Attention operates as a perceptual spotlight: a listener may selectively perceive one object in the environment while "tuning out" others, reducing their capacity to interfere with the attended object (Fritz et al. 2007). An object that is the focus of attention is perceived in greater detail, with a richer set of features, than unattended objects; attending to an object situates it as a perceived figure among an undifferentiated background of stimuli (Kubovy and Van Valkenburg 2001). Attention may be directed or shifted from one object to another, or from a composite object to one of its members, such that a listener is able to "scan" an auditory environment.

With these factors in mind, we may define the object in relation to both the sound stimulus—what its acoustic composition suggests—and the listener—how familiar the sound is, and what he or she wants to do with it. Although the object remains in the mind of the listener, the conditions under which it is likely to exist may be described such that two or more listeners with common (culturally shared) listening experiences may discuss "the object," just as I am about to discuss with you "the object in music"—in a somewhat contrived setting—in the following section.

1.3 The object in music

I will now ask you to produce a sound that is likely to be perceived as some form of auditory object. Say an /i/vowel, as in "beet," in full voice at a comfortable pitch, for about half a second in duration. Although you are likely to experience the resulting sound as unitary, the physical manifestation of the sound may be described as multiple: your vibrating vocal cords produce a **complex tone**, consisting of multiple **partials** or **harmonics**. Yet these partials share two features in particular that suggest they belong to a single object: **harmonicity**, in which their frequencies are multiples of a common **fundamental frequency** (**F0**), and **onset synchrony**, in which the partials begin to sound at the same time. The combination of these features is a strong **grouping** cue to their common sound source (Bregman 1990, chapter 3; Darwin and Carlyon 1995); the listener therefore groups these partials together to form a single sound, a "voice object."

Although this voice object corresponds with a sort of "everyday" mode of listening (Gaver 1993), we may conceive of more refined or "higher" levels of perception that correspond with more specific listening tasks—such as musical listening. A typical musical percept is not simply an auditory object, but an object perceived within a musical domain; it is a **musical object**. Our ability to experience music as music—as the interplay of melodies, rhythms, textures, and forms, rather than as an unorganized collection of sounds—relies upon our transformation of everyday auditory objects into musical objects, or what Scruton (1997, chapter 1) refers to as the transformation from "sound" to "tone."

We may differentiate these two kinds of object not only by their perceptual domains, but by the respective features that they bear. A listener hearing the /i/sound as occurring outside of any particular context—as an auditory object—hears the sound without any particular expectation of its behavior. Accordingly, he or she perceives features of the sound that are correspondingly general: for instance, the sound is perceived as having an approximate pitch height, duration, and loudness. But were the listener to hear this same /i/sound as a *musical* object, such as the first note of a familiar vocal melody, he or she would perceive features that are appropriate for musical listening. Instead of pitch height, he or she would hear a specific pitch location within a scale, known as **scalar pitch**; instead of raw duration, he or she would hear the musical object within a metrical context, as (for instance) “half a beat.”

The domain-dependence of features suggests that we may have different percepts of the same sound stimulus: by “applying” a different domain, we hear the sound as having different features. This change in percepts is demonstrated in the “speech-to-song illusion” examined by Diana Deutsch and colleagues (Deutsch, Lapidis, and Henthorn 2008; see also Falk and Rathcke 2010): by applying a typically musical behavior—metric repetition—to an otherwise unremarkable speech segment, the listener is more likely to judge the stimulus as “song” rather than “speech.” Accordingly, the listener’s ability to reproduce the scalar pitches of the stimulus is enhanced by this apparent domain switch; the musical feature of scalar pitch is enabled by the adoption of a musical domain.

1.3.1 Tone as musical object

This transition from auditory to musical object—a one-to-one process in the examples discussed above—provides several different listening strategies when applied to a musical sonority. In perceiving the sonority, the formation of auditory objects acts as a type of pre-processing event: low-level features that apply to auditory objects in general, such as harmonicity and onset synchrony, may be used to identify the tones of the sonority. The listener may attend to one or more of these "proto-objects" to form a musical object and perceive its higher-level, music-specific features, such as scalar pitch (Scholl 2001).

A visual interpretation of this scenario is shown in Figure 1.1. At the far left, the cloud representes the aggregate of sound stimuli, including the three tones of a musical sonority, as well as potential non-tone or "background" sounds. The tones of the sonority bear low-level features—shown in the figure as rectangles of various orientations—that enable them to be perceptually separated from any non-tone sounds; the listener attends to a proto-object to imbue it with the features of a musical tone, such as pitch and timbre. The non-attended tones, in contrast, are not perceived as having higher-level (musical) features, though they may in turn become objects of attention; the dashed lines in the figure represent their potential for perception as separate musical objects.

The listening strategy modeled in Figure 1.1 corresponds with Dahlhaus's "combinations of tones" mode of hearing sonorities: the listener forms a musical object of one of the sonority's component tones, while the other tones are left unattended in a not-yet-musical state. This mode of hearing is appropriate to a musical passage in which one tone or "voice" is understood to have greater musical significance than the others—as with a melody against a harmonic accompaniment, for instance. The listener's attention to this privileged voice permits him or her to experience its musical features in greater detail. However, this detail comes at the expense of **emergent features** that are shared among combinations of tones; for instance, the ability to perceive pitch distance between two tones is more difficult when the listener attends to only one of the two tones (Borchert, Micheyl, and Oxenham 2011).

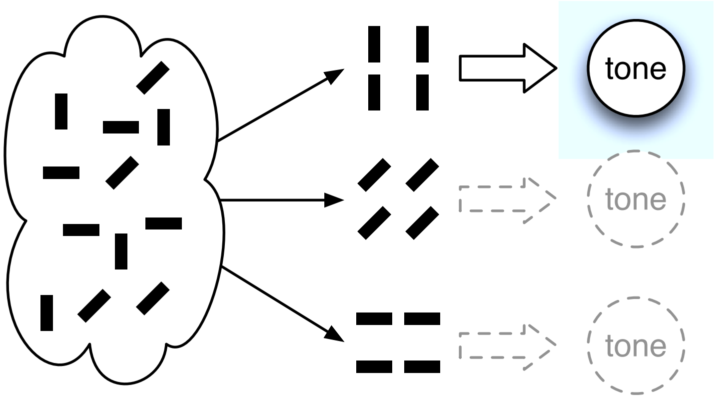

Figure 1.1: Attending to a single tone as a musical object

Although the listener in the scenario above is less able to perceive emergent features among the sonorities' component tones, this strategy need not be classified as suboptimal: in many cases, the features of an individual tone may be more musically relevant than those of possible tone combinations. In fact, the control of emergent musical features such as dissonance is a common compositional ideal in Western musical practice; consequently, composers of Western polyphony often "softened" the perceptual salience of dissonant intervals by staggering the onsets of the interval's two tones—thereby ensuring that the tones were perceived as separate musical objects (Wright and Bregman 1987; Wright 2008).

1.3.2 *Chord as musical object*

Figure 1.2 depicts a different listening strategy, in which the listener forms an aggregate percept of two of the three tones of the sonority, grouping them into a musical object—a dyadic chord.[4] As with the strategy shown in Figure 1.1, this approach highlights certain features while obscuring others. Perceiving the dyadic chord allows the listener to better perceive its emergent features; in this case, the chord contributes an "interval" feature derived from the pitch distance of its two component tones. The ability to directly perceive an interval feature would be useful in a musical culture where such intervals were associated with other syntactic elements. To find such a culture, we need look no further than the contrapuntal practice of Western medieval music, in which specific "imperfect" intervals (minor third, major third, and major sixth) were understood as implying continuation or motion toward a more reposeful, "perfect" interval (unison, perfect fifth, and octave, respectively) (Cohen 2001b).

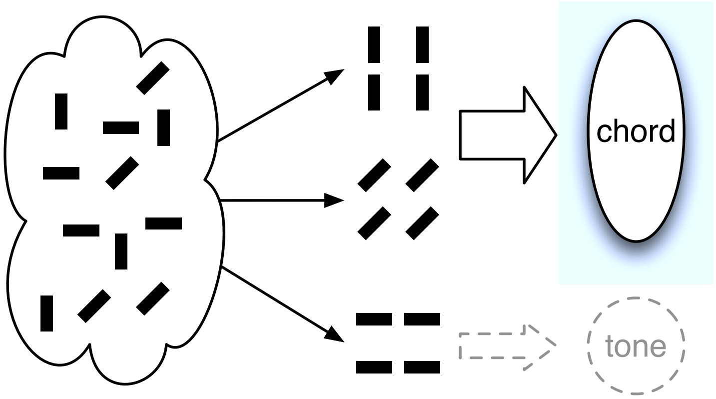

Figure 1.2: Attending to a dyad as a musical object

In turn, the listener who perceives the dyadic chord is less able to discern the musical features associated with an individual tone. If we assume that the two tones of the chord are similar in many respects—such as loudness, timbre, and location—then the most musically significant feature to be affected by this mode of perception is pitch:

---

[4] In this discussion I choose the term "dyadic chord" over the more common "dyad" to highlight its holistic nature; the definition of chord, provided in section 1.1, includes two-tone sonorities.

hearing the dyadic chord makes it harder to hear the pitches of its component tones. Yet as we understand both pitch *and* interval quality to be significant within most Western musical practices, we must account for the listener's need to perceive both: the two features should not be mutually exclusive. Accordingly, the listener may shift between these two perceptual scenarios—at one time hearing a chord, a moment later hearing its constituent tones—when it is musically advantageous to do so. The listener accomplishes this shift by redirecting his or her attention toward the desired object—effectively changing the resulting percept through a change in listening task. This attentional shift is not instantaneous: the control of attention to move between two percepts bears a small time cost, usually in the range of 100–200 milliseconds. However, it appears that Western contrapuntal practice has made a concession toward this time cost to encourage such shifts: the musical moments when interval quality is most syntactically significant—at and preceding the cadence—are also the moments at which multiple voices typically align to form a long, synchronous sonority that permits the listener to perform this shift.

Alternatively, we may also theorize that the dyadic chord affords its own pitch-like features: as both pitch and interval are derived from the same frequency dimension, it is plausible that an interval percept would afford some form of pitch as well. In the medieval Western contrapuntal practice discussed above, the listener would have a musical reason to prioritize one of the two tones comprising the dyadic chord: a likely candidate would be the tenor voice, which is understood to have the largest role in regulating between-voice intervals. The listener might then "draw out" the **harmonic series** of the tenor and perceive its pitch as a feature of the dyadic chord.[5] In this scenario, the listener

---

[5] The claim that chords may afford a pitch percept is taken up further in chapter 4.

hearing the dyad may perceive it as having the musical features "D-with-major-sixth-above"—taking in both pitch and interval—instead of simply hearing a major sixth.

This strategy of perceiving tone aggregates may be extended to encompass all three tones, as in the case of a triadic chord. As with perceiving the dyadic chord, the listener's ability to perceive the chord depends upon its musical utility: in a musical practice that makes no syntactic use of three-tone sonorities, the listener would have no musical reason to take this aggregate as musically meaningful. And it is this point that Dahlhaus uses to differentiate harmonic tonality from intervallic or melodic forms of musical organization: although both medieval and modern Western musical cultures use triadic relationships as a means of organizing multiple tones, only those musical practices that employ the *concept* of triadic chords—of tone aggregates as bearers of musical information—may be said to include triadic chords.

This claim need not imply that the medieval listener never heard chords of three or more pitch classes, as in a triadic chord. The frequent use of sustained pre-cadential or "tendency" sonorities in fourteenth-century *ars nova* musical practice (Fuller 1992) suggests that a listener could perceive a holistic aggregate of tones when hearing a familiar sonority. As with the dyadic chord discussed above, the listener may shift his or her attention away from the tone level to take in the chord percept of a musically meaningful sonority. In this regard, Dahlhaus's claim for the role of chords in harmonic tonality is one of degree, not of kind: whereas earlier listening practices might have taken only a few types of sonorities as "unitary," a listener within the musical culture of harmonic tonality may hear a broader range of sonorities in this manner.

Similarly, a listener disposed to hearing chords may shift his or her attention from the chord level to the tone level, though this is often a more difficult path: as with the partials of a complex tone, the component tones of a chord often share several acoustic features—onset timing, timbre, and dynamics—that provide a significant amount of compositional consistency. The listener attempting to "hear out" a tone within the chord may therefore rely upon the existence of contextual cues—such as those provided by voice-leading guidelines of Western contrapuntal music—or upon his or her training in perceptually separating the note from the chord.[6]

1.4 Conclusion

We may now paraphrase Dahlhaus's opening argument in terms of the factors and models presented above. To make a claim about a musical concept such as the chord, we must take three factors into consideration: the acoustic stimuli that are the products of performance, the musical context that informs the listener's goals, and the listener's experience in forming objects in accordance with these acoustic and contextual factors. The remaining chapters of this dissertation grapple with two questions that arise from this exposition: how we might tell when a sonority is likely to be heard as a chord, and how the chord percept might differ from the percept of multiple notes.

In chapter 2 I investigate the chord as a product of our perceptual processes, drawing upon Bregman's landmark monograph on auditory perception (Bregman 1990) to frame my exploration. I then explore the particularly musical factors that shape and

---

[6] Eric Clarke notes that the ability to perceive the individual tones of a triadic chord is acquired through training; "untrained" listeners "tend to regard a chord as a single entity" (Clarke 2005, 24).

direct our listening, and I conclude by proposing a set of production strategies under which a listener is likely to hear a sonority as a chord. Chapter 3 examines the factors of experience and context on the listener's tendencies to hear a sonority as a chord. The first of these experiments presents familiar (triadic) and unfamiliar (non-triadic) sonorities to examine the effect of listening experience, whereas the second uses a variable pre-sonority context to condition the listener's level of object formation.

The listening models given above not only account for the possibility of multiple perceptual approaches, but they also suggest how these approaches differ experientially—particularly with respect to how we hear the frequency content of the perceived objects. In chapter 4 I examine the perception of pitch—a familiar feature of the musical tone in both perceptual and theoretical descriptions—and suggest conditions in which pitch might be perceived as a feature of the chord as well; I locate these conditions within musical practice in chapter 5, in which I claim that chord pitch can be used as a compositional resource. I conclude with a proposal for further empirical work in chord perception, in the mold of the experiments presented in chapter 3, whereby we may begin to locate the other musical features that characterize our experience of chordal listening.

CHAPTER 2

PRODUCING THE CHORD

The advent of chords in Western music is commonly understood as a gradual broadening of the musician's harmonic palette. As discussed in chapter 1, Dahlhaus's distinction between individual and chordal listening forms one dimension of this broadening: musicians learned to use groups of multiple tones in consistent and musically meaningful ways. A similar development may be observed in the dimension of preferred intervallic content, or consonance: whereas earlier contrapuntal practices favored sonorities composed of perfect intervals (octave, fifth, and fourth), later practices expanded their category of consonances to include other intervals such as thirds and sixths.

If we attempt to trace the changes in these two dimensions—group size and intervallic content—within Western musical culture, we would note a historical lag between the two: triadic sonorities appear to have been in use for several generations before the appearance of chordal thinking in writing and practice. It is certainly plausible to conclude that this lag between harmonic content and harmonic concept may simply be "a matter of time"—that musicians required a period of acculturation before they were able to grasp the triad's potential for forming integral musical structures. Such is the proposal of Helmholtz (1877) in his assessment of the pre-chordal nature of Renaissance polyphony:

> Great, then, as was the artistic advance in rhythm and the progression of parts, during this period, it did little more for harmony and the tonal system than to accumulate a mass of *not-yet-orderly experiences* [*noch ungeordneter Erfahrungen*]. Since the involved progression of the parts gave rise to chords in extremely varied transpositions and sequences, the musicians of this period

> could not but hear these chords and become acquainted with their effects, however little skill they showed in making use of them. At any rate, the experience of this period prepared the way for harmonic music proper, and made it possible for musicians to produce it, when external circumstances *invited such a discovery* [*Erfindung hindrängten*]. (Helmholtz 1877, 246, alternate translation in italics).

In this chapter, I attempt to augment this scenario by examining the perceptual basis of the concept of chords—the factors that lead to the cognitive and cultural uptake of triadic chords as musical objects. I begin by examining the perceptual **processes**—the operations occurring along the path from stimulus to object—that act upon acoustic and contextual cues toward a holistic percept. These processes are situated within their respective domains: the perceptual recruitment of different listening domains enables the distinction between "sound" and "tone" introduced in the previous chapter. I then refine the definition of musical object introduced in section 1.3 as the result of a particular combination of musical and "everyday" listening tasks. With this definition in hand, I go on to discuss the musical utility of the chord, concluding with a hypothesis on the perceptual and cultural factors that contributed to the introduction of chords in Western music.

2.1 Processes of auditory perception

The analysis of perceptual processes is limited by the same conditions that shape our understanding of objects: as perception is a mental act, the elements of perception resist objective identification. Although recent research has identified patterns of brain activity, known as event-related potentials (ERPs), that offer traces of perceptual processes such as object formation (Winkler, Denham, and Nelken 2009), the results of this work have little value without the *a priori* psychological concepts we use to understand them. With these limitations in mind, we may proceed by exploring those listening scenarios that afford the greatest insight toward the capabilities of our auditory system.

The complexity involved in making sense of our auditory world is best captured by the "cocktail party problem," initially posed by Colin Cherry in an influential article (Cherry 1953; see also Bronkhorst 2000): how are two people able to maintain a conversation in a room filled with the sounds of other conversations? We may consider this problem to be a question of how to form auditory objects: the listener, in attending to one person's speech among many, must form a "desired talker object" that is separable from the similar sounds of speech from other talkers. At the same time, he or she must also be able to follow and group the various phonemic sounds of the talker's speech into a single speech object.

Albert Bregman's landmark monograph (Bregman 1990) characterizes the task of the cocktail party listener as *auditory scene analysis*—assembling auditory objects that correspond with their real-world sources. Bregman identifies two kinds of processes that make this analysis possible.

- **Sequential** processes seek out regularities or differences in sounds across a span of time (Bregman 1990, chapter 2). A sound that exhibits moderate changes in its properties over time, such as a sounding siren's rising and falling pitch, may still be heard as a single object. Similarly, sequential processes may detect regularities within a series of short repetitive sounds, such a footsteps or raindrops: despite the presence of discontinuities, we group or **stream** the sounds into a single object.

- **Simultaneous** processes, in contrast, seek out regularities or differences among multiple sounds within a single "slice" of time (Bregman 1990, chapter 3). Sounds that possess noticeably different features are deemed to come from different sources; the listener is then able to **segregate** these contrasting sounds to form different objects. Described in terms of auditory scene analysis, the cocktail party listener streams one talker's speech while segregating it from other speech.

Let us now consider each of these processes in a bit more detail.

2.1.1 Sequential processes

The mechanics of our sequential processes are best observed in an experimental paradigm employed by van Noorden (van Noorden 1975; Micheyl and Oxenham 2007). In this experiment, the listener hears a loop of two partials or **pure tones** at different frequencies, labeled A and B; the tones are repeated or "looped" in an A–B–A–silence pattern (with each element having an equal duration), as shown in Figure 2.1. When the repetition rate (tempo) of this pattern is slow, or when the difference in frequencies of the A and B tones is small, listeners were likely to hear a single stream or auditory object, composed of the A–B–A pattern—known as the "horse" pattern for its galloping sound. At faster repetition rates, or with greater frequency differences, listeners were more likely to hear two separate streams of A and B tones—known as the "morse" pattern (similar to two streams of Morse code). In between these extremes, the perception of one stream ("horse") or two streams ("morse") seems to alternate somewhat randomly over an extended listening period.

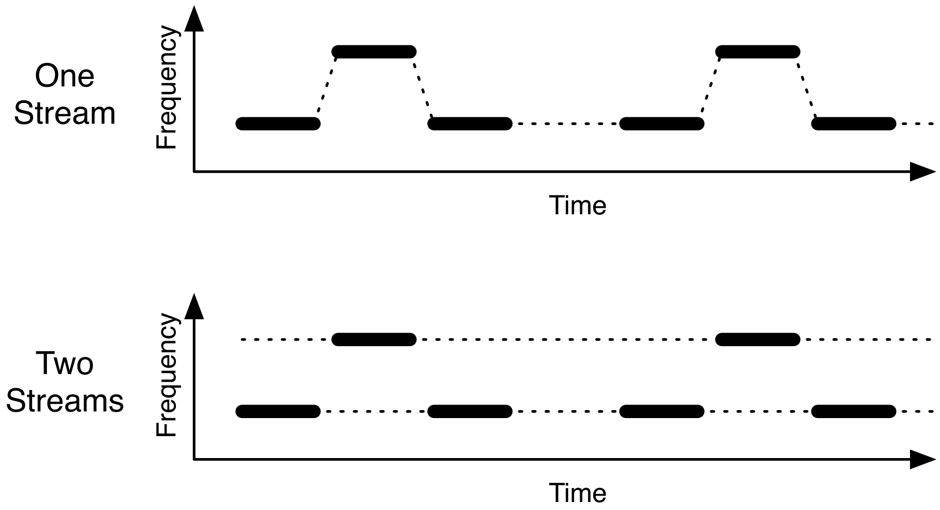

Figure 2.1: Pattern of repeating tones, adapted from van Noorden (1975)

Because these sequential or streaming processes seem to operate reflexively on the acoustic features of the stimulus, Bregman labeled them "automatic"; their role is to group sounds in time so that the listener might attend to one or the other (Bregman 1990, chapter 1). Further research in this sequential-tone paradigm by Robert Carlyon, Rhodri Cusack and colleagues (Carlyon et al. 2001; Cusack and Carlyon 2004) suggested that attention may also be active in stream *formation* as well as stream selection; attention need not simply select from an array of pre-formed objects. One significant implication of this research is that sounds outside of our attentional focus are not elaborated: they may remain somewhat unprocessed or undifferentiated in comparison with the object of our attention. Placed within the cocktail party paradigm, the listener doesn't segregate other conversations that are taking place in the room; they are heard together as a "speech noise" background.

Not all sequential processes operate at the "automatic" level of time and frequency dimensions: the listener may apply contextual experience-based processes—called *schema-based* processes by Bregman (1990, chapter 1)—to detect regularities within higher-level features. Experience-based processes are at work when listening to speech, for example: in hearing a sequence of speech sounds that differ in frequency content (such as vowels and consonants), the listener uses his or her knowledge of common speech patterns to override these low-level features that might be taken as segregation cues in a non-speech context. More broadly, recent research suggests that we may attend to any salient features—such as pitch, timbre, or spatial location—that exhibit Gestalt "good continuity," and use them as the basis for streaming (Xiang, Simon, and Elhilali 2010).

2.1.2 Simultaneous processes

In the cocktail party scenario, simultaneous processes act to separate multiple concurrent objects, so that the listener may follow or stream the object of his or her choosing. Although many of the sequential cues introduced above may also be used as segregation cues—the features that characterize one sound often differ from the features of others—there are two widely observed cues that appear to be specific to simultaneous processing. Onset synchrony, in which two or more sounds begin simultaneously, is a grouping or **fusion** cue to hear these sounds as a single object: because unrelated sounds are unlikely to start at the same time (Bregman 1993), we take this condition as a cue to perceptual integrality. Harmonicity, in which the frequencies of two or more tones form multiples of a common fundamental, is also taken as an ecological cue to objecthood, since most real-world sound sources (including the human voice) produce harmonic complex tones.

An experiment by Bregman and Pinker (1978) extended the van Noorden "horse or morse" paradigm to examine the effects of onset synchrony and harmonicity on object formation; the authors combined a repeating A–B pattern with an added third tone (labeled C), such that the B and C tones occurring simultaneously. As in the van Noorden experiment, adjusting the A–B frequency gap altered the likeliness of forming an A–B stream—leaving the C tone to form a separate stream. More significantly, adjusting the temporal onset of the C tone also affected listeners' percepts: as the C tone was made more "out of sync" with the B tone, listeners were more likely to connect the temporally shifted B tone with the A tone, leaving a "pure" C tone as a separate object.

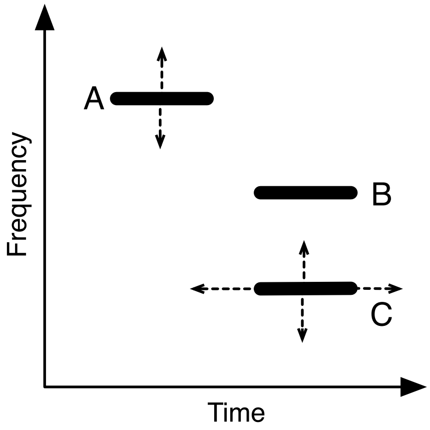

Figure 2.2: Pattern of repeating tones, adapted from Bregman and Pinker (1978)

The authors hypothesized that adjusting the frequency of the C tone might reveal the effects of harmonicity: aligning the frequencies of the B–C pair in a simple harmonic ratio would make the B–C grouping effect more likely, compared to pairs in which their frequencies were inharmonic or "dissonant." As it turned out, however, adjusting the frequency of the C tone did not have a significant effect on the streaming or alternating tendencies otherwise present (due to the other two adjustments). In discussing the lack of a significant harmonicity effect, Bregman (1990, chapter 3) suggests that the use of only two tones was not sufficient to elicit an effect, and that complexes with more pure tones would be more likely to perceptually fuse.

Another way to explain the results of Bregman and Pinker's experiment would be that harmonic ratios are by themselves insufficient to promote segregation. If we frame the two grouping processes corresponding to onset synchrony and harmonicity as *segregation* processes, we may interpret Bregman and Pinker's results to imply that differences in both frequency *and* time are necessary for segregation (Kubovy and Van Valkenburg 2001; Van Valkenburg and Kubovy 2003).[1] Support for this theory is found in the work of Elhilali and colleagues (Elhilali et al. 2009; Shamma and Micheyl 2010), whose *temporal coherence theory* reframes the role of simultaneous processes in initial object formation: in order for two sounds to be perceptually segregated, they must exhibit salient differences in both onset and frequency. The temporal coherence theory effectively states that there are no "simultaneous processes": instead, objects are formed instantaneously based on perceivable differences (boundaries) in onset time and frequency, while other perceivable features may be used to segregate or stream sounds over time. The grouping or fusion cue of harmonicity may therefore be explained as a sequential cue to pitch; the commonality of harmonic sounds leads the listener to favor percepts that produce a pitch.

---

[1] A similar experiment led by Turgeon, Bregman, and Roberts (2005) examined the harmonicity effect of fusion in harmonic complex tones that either had the same or different fundamental frequency; the authors found that harmonicity did have a significant effect in segregation—but only when coupled with differences in onset timing.

2.1.3 Perceptual domains

Bregman's distinction between primitive and schema-based processes reflects the seemingly automatic nature of many of the primitive processes analyzed in the preceding section. Whereas this distinction is often rephrased in terms of **bottom-up** (acoustic or stimulus-based) and **top-down** (cognitive or knowledge-based) processes, Bregman later proposed that this distinction is more properly about the environments in which they operate: "By calling them primitive, I mean that instead of depending on knowledge of specific types of sound, such as voices, musical instruments, or machines, they depend on general acoustic properties that can be used for decomposing all types of mixture" (Bregman 1993, 14). Accordingly, recent research uses the concept of domain (introduced in section 1.2.2), to indicate the environment, context, or task in which the listener would use a given process. A process that applies to all perceptual environments or tasks is called **domain-general**; processes that appear to be developed or specialized for use in a specific context or task are **domain-specific**.

The concept of domain may be applied with greater or lesser specificity, such that, on the one hand, a given stimulus may be the subject of multiple layers of domain-specificity, and, on the other hand, the breadth of each successive domain becomes successively narrower, down to the limit of the individual—the "personal domain".[2] A listener hearing a melodic passage, such as the fragment shown in Figure 2.3, may call upon several levels of specificity in parsing the melody into musical objects. At the lowest, domain-general level, there are observable regularities or patterns that allow the

---

[2] See Hannon and Trainor (2007) on the process of learning in a music domain, and the mix of domain-specific and domain-general development that occurs in musicians.

listener to connect the tones of the melody to form an object, despite the brief silences that may separate each individual tone. At a basic musical level, he or she may experience the melody as existing in a musical scale of pitches, noting that the melody begins and ends on the same pitch. At a more culturally defined musical level, he or she may recognize an interleaved melody that forms a Western nursery rhyme, "Mary Had a Little Lamb." Further specific domains are also possible: the listener with a knowledge of Western harmony could, for instance, segregate the melody based on their membership in tonic or dominant harmonies, for instance.
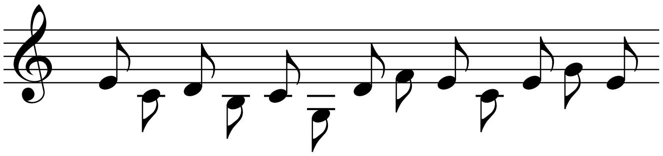
Figure 2.3: A musical fragment with interleaved melody

2.2 Characteristics of musical practice

The introduction of multiple levels of domain-specificity further complicates our attempt to capture the listener's formation of musical objects. Although we may acknowledge the possibility of highly individual modes of listening based on a listener's learning and experience, it is also desirable to propose a working definition of musical practice—a collection of behaviors that distinguishes music from other human activities—such that we may explore listening tasks that are appropriate to the musical domain. For the purposes of this exercise, I adopt a broad but pragmatic definition of musical practice: it is the use of *patterned non-linguistic sounds*. From this definition I propose three characteristics that are specific to musical practice; these characteristics are then used to identify musical listening tasks that may be combined or contrasted with the domain-general listening goals discussed earlier.

2.2.1 *Organization by pitch and rhythm*

The use of pitch or rhythm—the music-perceptual correlates of frequency or time, respectively—to organize music is common to all known musical cultures; the majority of musical practices use both (Nettl 2000; Bispham 2009). There are several reasons for the prominence of pitch and rhythm in music. First, as music may lack the semiotic anchors present in other sound-producing practices such as speech, it is practical to organize sounds along those features that are closely correlated with the acoustic stimuli. Second, the organizational features of pitch and rhythm match two of the fundamental behaviors considered essential to music making: the ability to match pitch with a perceived sound, and the ability to entrain or synchronize sound production across multiple participants (Bispham 2009). Last, the physical structure of our hearing system permits fine-grained distinctions in pitch and rhythm—features derived from the dimensions of frequency and time, respectively—but only cruder distinctions within other potential organizational features such as loudness or timbre (Patel 2008, chapter 2).[3]

2.2.2 *Emphasis of features over source identification*

Our most common listening activities—domain-general as well as domain-specific—are often oriented toward the source of a perceived object. This behavior

---

[3] This does not prohibit the organizational use of other attributes in musical listening; Patel (2008, chapter 2) notes that the music of the *tabla*, a set of hand drums originally used in Indian music, may be described as being organized in timbre.

is desirable in the speech domain, in which the intelligibility of speech depends upon its segregation from concurrent noise and other speech. Likewise, a listener often hears a "real world" auditory object as providing information about its source: perceiving the spatial location of a sound permits the listener to orient his or her response toward (or away from) the sound source (Neisser 1976, chapter 2). Music, in contrast, need not rely upon source identification: although the listener may segregate what he or she perceives as music from "background noise," he or she may organize patterns of pitch and rhythm irrespective of the features' sound sources.

The use of pitch in particular to group sequential sounds from separate sources—known as a hocket in medieval Western practice—is a common example of this emphasis. Figure 2.4, an excerpt from an anonymous 14th-century English motet, contains a highly salient hocket between the two tenor voices as they approach the final cadence. In this passage, pitch proximity and alternating note/rest patterns are used to give the impression of a single melodic line; the listener, experienced with hearing melodies that have comparatively "smooth" pitch motion, streams the two tenors' contributions into a single melody.

2.2.3 Complex sounds

Overall, listeners show a domain-general aesthetic or "hedonic" preference for complex sounds (Brattico, Brattico, and Jacobsen 2009): the perceptual richness of complex sounds facilitates perception and organization according to its salient features, and complex sounds are more robustly encoded in auditory short-term memory than (ecologically rare) pure tones (McKeown and Wellsted 2009). Within a musical setting, the
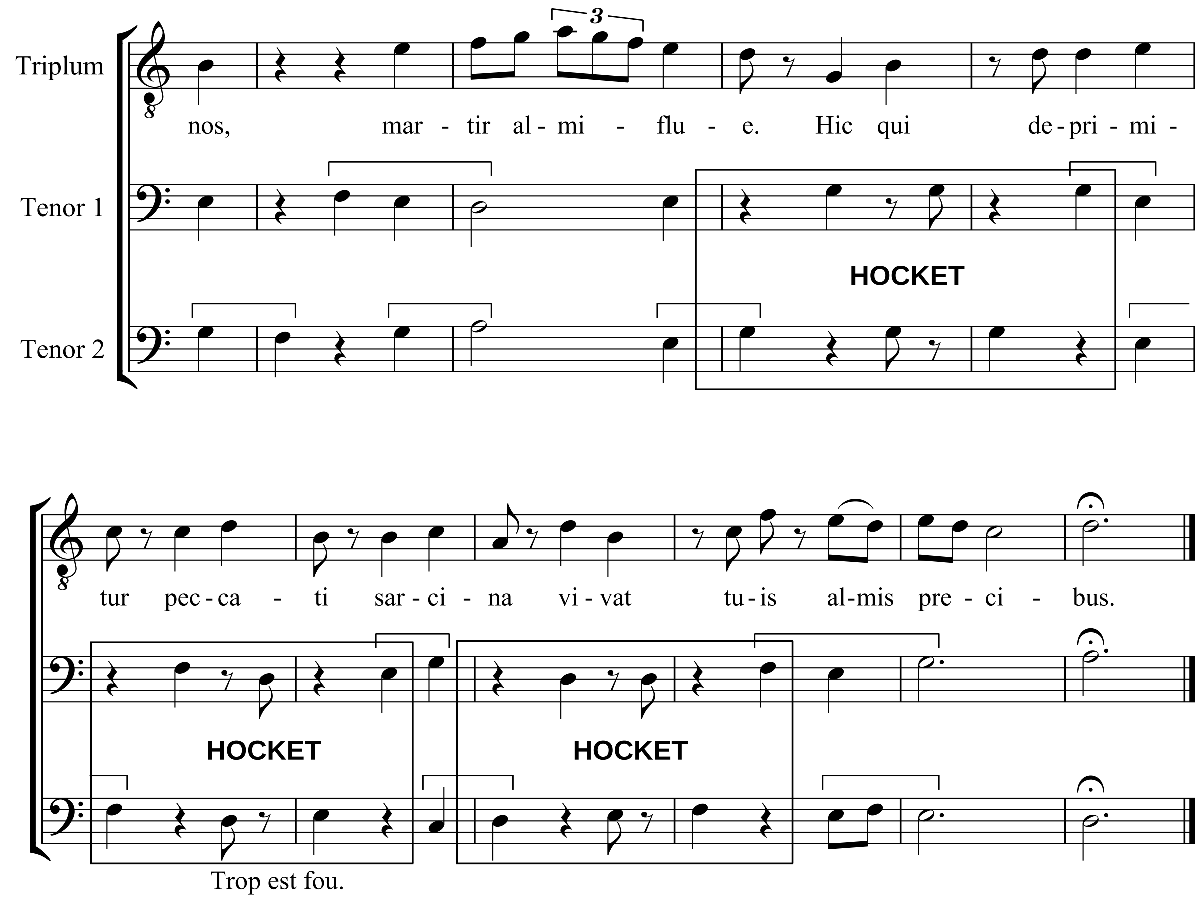
Figure 2.4: Anonymous (14th century), *Triumphat hodie /Trop est fol /Si qe la nuit*, adapted from Harrison (1980)

listener may prefer to form objects that are rich in non-organizational or "secondary" musical features, such as timbre or dynamics; such features may contribute to the aesthetic experience of musical listening, so long as they do not obscure the sound's primary organizational features (Schneider 2001).

Musical practices are also characterized by their creation of complex sounds through more synthetic means. Whereas the use of hocket in Figure 2.4 shows how pitch may be used to create a temporally extended complex object, a much more common musical practice is the alignment of multiple sounds in the dimensions of frequency or time. Bregman (1990) discusses the musical utility of such combinatorial complex sounds, which he refers to as **auditory chimerae**:

> We use the word chimera metaphorically to refer to an image derived as a composition of other images. An example of an auditory chimera would be a heard sentence that was created by the accidental composition of the voices of two persons who just happened to be speaking at the same time. Natural hearing tries to avoid chimeric percepts, but *music often tries to create them*. It may want the listener to accept the simultaneous roll of the drum, clash of the cymbal, and brief pulse of noise from the woodwinds as a single coherent event with its own striking properties. The sound is chimeric in the sense that it does not belong to any single environmental object. (Bregman 1990, 459–60, emphasis mine)

The auditory chimera percept in Bregman's example is a result of **energetic masking**: the onsets of the chimera's component sounds are so closely aligned that our auditory systems lack the resolution to use them as segregation cues. As we shall see in the following section, the creation of auditory chimerae is one of the more fundamental characteristics of music making: it is the musical behavior that creates the chord.

2.3 The musical object

With the above characteristics in mind, we may find numerous examples of objects that are well-suited for musical hearing. Although it is possible to hear any sound stimulus as musical given the proper environment—a fact to which John Cage's *4′33″* stands as the most notorious testament—I would like to highlight one particular kind of object as being essentially and fundamentally musical, as it possesses each of the characteristics outlined in the preceding sections: this object is the musical unison. And as the unison constitutes a "trivial case" of the chord—it is a chord in which all tones happen to have the same pitch—we may trace a path from the unison to the chord within musical practice.

2.3.1 The unison as musical object

Instances of the musical unison are ubiquitous, from the sound of an orchestral string section to an impromptu chorus of "Happy Birthday." Yet the production of this seemingly ordinary sound requires both of the fundamentals of musical behavior noted in section 2.2.1 above: matching complex pitch and entrainment to a metric pulse. In turn, the unison preserves the "primary" organizational features of Western music, pitch and rhythm: the unison bears the same pitch and duration as one of its component tones. And as the unison's component sounds are aligned in both time and frequency dimensions, it effectively masks the acoustic signals to source identification: it is doubly chimeric.

It is, however, in the realm of "secondary" musical features, discussed in section 2.2.3 above, that the musical value of the unison is best realized. The combination of multiple tones into a single object effectively increases the loudness of the object, which often has a desired ecological effect of increasing the range and number of listeners who may hear the sound. In addition, the slight spectral offsets within this object create patterns of variation and interference, known as **modulations**. As these modulations are uncommon to sounds produced by a single source, they introduce into the resulting object an unusually complex pattern of spectral activity—creating a sound that appeals to our domain-general aesthetic preference for complexity.

2.3.2 *The chord as musical object*

The chord, an auditory chimera like the unison, affords "secondary" musical characteristics that are not attributable to a single tone and that provide the listener with an aesthetic reason to retain the initial holistic percept of a single object. However, the chord differs from the unison in one significant regard: it does not necessarily preserve the availability of a holistic pitch percept.

As the combined partials from a chord's tones are unlikely to form a single harmonic series, any possible pitch percept of the chord would be "weaker" or less salient in quality than that of a single tone. And since pitch is one of the organizational dimensions of Western music, listeners familiar with Western music will prefer to hear sounds as having a defined pitch: the listener has incentive to parse the chord into its pitch-bearing notes. In other words, the chord—regarded as an essential element of Western harmony—may be regarded as an **inharmonic** sound, in that it does not afford a single harmonic series.[4]

In this respect, perceiving multiple tones is similar to perceiving multiple speakers: the lack of meaningful features in the aggregate—pitch in the musical domain, speech content in the speech domain—encourages the perception of its components. Yet as prominent a role as pitch plays in Western musical organization, it need not be the sole factor. We may hypothesize that the chord, as a complex sound, may afford other emergent features that may be used for musical organization, and that the listener within a

---

[4] Chords built from octaves, perfect fifths, and their compounds form a notable exception to this claim: these sonorities are mostly harmonic due to their overlapping harmonic series. The pitch-bearing nature of these octave-only or "open-fifth" chords is taken up in the following section and examined in detail in chapter 4.

chord-using musical culture would have reason to perceive and organize passages of harmonic music at the chord level as well as the pitch level. The nature of these chordal properties—that is, the perceptual correlates of the typology we apply to chords—will be discussed in more detail in the final three chapters of the dissertation; at this time we may simply observe that the listener may draw upon experience to recognize one or more emergent chordal features across a mixture of initially segregated notes, and shift his or her percept accordingly to the level of the chord.

2.4 Performing the chord

We have now outlined the domain-general and music-specific goals and processes that are most commonly active when hearing a sonority: although the chord is a chimeric percept that bears an aesthetically pleasing complex sound, the listener's use of pitch to organize music encourages the perception of its pitch-bearing component tones. This knowledge may now be directed toward the performance of sonorities, such that we may propose a number of compositional strategies that promote hearing a sonority as a chord. In the following, I outline a number of musical cues—both domain-general and music-specific—that are likely to promote a chord percept. Taking the chord level and the note level as the primary perceptual "axis" of sonorities, we may propose two general listening scenarios involving the perception of a chord:

1. The listener initially perceives a chord, which may or may not be retained throughout its salient (audible) duration.

2. The listener initially perceives one or more notes, after which he or she shifts to a chord-level percept.

To the musician wishing to create a sonority that is heard as a chord, these two scenarios suggest different but possibly complementary approaches. The first approach involves "brute force" methods which promote the initial binding or grouping of the chord's tones: this is most easily achieved through onset synchrony, although there are other factors—both acoustic and contextual—that support this initial binding. The second approach requires the use of music-specific cues that direct the listener's attention to the chord-level percept. As the relative "success" of these cues is likely to depend upon the listener's previous experience in hearing chords, it is more difficult to itemize them in a list of compositional strategies; they are taken up more extensively in section 2.5.[5]

2.4.1 Promote initial object binding

The theory of temporal coherence (Elhilali et al. 2009) states that objects are initially formed along salient boundaries in both time and frequency dimensions. Where differences in time or frequency are too close to be effectively perceived, the result is known as energetic masking; the auditory chimera, discussed in section 2.2.3 above, is a musical exemplar of energetic masking. Since musical chords most often include tones with different fundamental frequencies, the control of the time dimension—through on-

---

[5] Erickson (1975, chapter 2) provides a similar list of strategies for the composer wishing to create sonorities that are heard integrally; Erickson uses the term "a sound" (with quotes) to indicate an impermeable, holistic percept, whereas a sonority that more readily affords perception of its components is called a chord.

set synchrony of the sonority's tones—is the most effective way to promote an initially holistic percept.

Although we are capable of detecting very minute differences in onset timing—on the order of several milliseconds—this distinction is more of sound quality than of source determination: a listener might hear a pair of sounds with slightly staggered onsets as somewhat longer or thicker, but he or she would likely be unable to segregate the sounds based solely onset differences. Thresholds for segregation are likely to vary with differences in context and dynamic envelope (attack quality): however, a trained musical ensemble is able to coordinate its tone onsets within 30–50 milliseconds, which is considered under the energetic masking threshold for a typical musical listening environment (Rasch 1979). If we wish to define a chord based on the temporal coherence suggested by its musical notation, it is therefore reasonable to assume that multiple tones that are intended to be sounded simultaneously will be heard as such, even allowing for the onset differences common within an musical ensemble setting.

2.4.2 Limit use of domain-general processes

Although the two factors discussed in the following—limitations on salience and duration—are applicable to "real world" as well as musical listening, their utility in music is rather circumscribed: restricting the audibility of a musical object is more often than not undesirable in a musical setting. Nevertheless, these factors are effective in promoting the chord percept: a sonority that is barely audible, or audible for a brief time, will be especially resistant to division into even less salient components.

Limit salience. A salient stimulus is one that is available for perception; it "stands out" from its background. Conversely, a sound with limited salience will be barely discernable from other stimuli present. Limiting the salience of a sonority may be achieved by two means. First, a sonority may be performed at a low **sound pressure level**—the acoustic correlate of loudness or "volume"—such that the sonority is barely audible: as just noted, this has the corresponding effect of reducing the sonority's musical utility. Second, the sonority may be simultaneously accompanied by other sounds, either musical or "background," such that the chord is energetically masked by these sounds; this practice of limiting a chord's signal-to-noise ratio has a more obvious musical correlate in the use of simultaneous musical streams, such as a melody with chordal accompaniment—a practice discussed in greater detail below.

Limit duration. As noted by Erickson (1975), producing a chord with a very short duration is likely to preserve its initial holistic state: a composite sound that is perceived as too short in relation to its musical context is likely to be heard as a single object. This effect is even more prominent in sounds that have a salient attack in their dynamic envelope, such that the sound of the attack occupies most of the audible (above-threshold) duration of the sound: the attack serves as a "common fate" dynamic characteristic, providing further cues to holistic perception (see section 2.4.4 below). Although minimizing the duration of a chord limits the listener's ability to segregate its tones, sounds with brief durations tend to be exceptional in many musical cultures, in which the continuity of a melodic line is highly valued. This suggests that highly "punctuated" chords are not well-suited for melodic purposes—which in turn acts as an additional deterrent to segregation, as discussed in the following section.

2.4.3 Limit use of music-specific processes

The higher-level processes that give rise to musical features do not happen automatically or instantaneously: if a listener initially perceives a chord, then proceeds to "hear out" a note, he or she must direct his or her attention to the available note-level cues. We may therefore consider two factors that limit the listener's use of attention in perceiving possible segregation cues within a sonority: contextual cues that direct the listener's attention elsewhere, and the use of **informational masking**—presenting the listener's auditory system with more potential objects than it can process—that renders attention less effective.

Direct attention elsewhere. Providing the listener with a concurrent musical object in addition to a sonority may be an effective strategy for limiting attention to the sonority. When a melody is coordinated with a harmonic accompaniment—as happened historically with the development of monody in the late Renaissance, or as happens today when a singer accompanies himself or herself with guitar or piano—the listener's attention is most often directed toward the melody. Compared to a series of accompanying sonorities, a melodic line is distinguished by greater activity in pitch and rhythm, and as such it carries more informational "bits" for the listener to attend to and follow (Madsen and Widmer 2006; Duane 2010). The sonority, as an unattended or background object, is less able to be parsed into its component tones.

Promote informational masking. Whereas energetic masking describes the obscuring of time and frequency boundaries, informational masking concerns higher-level processing: it is a failure to correctly parse the auditory scene, and it is a common result of "flooding" the listener's perceptual processes with more stimuli that the listener can effectively process (Shinn-Cunningham 2008). Informational masking in auditory perception is often evoked by presenting the listener with many simultaneous sounds, without providing any contextual indicator that may suggest to the listener where to begin identifying the features that lead to formed objects.

Perceptual studies of listening to polyphonic music (Brochard et al. 1999; Huron 2001) suggest that four voices is the typical threshold of informational masking: when a listener is asked to enumerate the voices of a musical passage, he or she is significantly more likely to make errors when the number of voices is greater than three. The potential for informational masking in musical listening can nonetheless be mitigated in a number of ways: musically trained listeners often possess increased cognitive skills for processing multiple tones (Strait et al. 2010), and musically experienced listeners intending to segregate notes from the sonority are likely to begin at the highest tone of the sonority, where a melodic line is most likely to occur.

2.4.4 Limit segregation cues

The musician may prolong the initial holistic percept of a chord by controlling the availability of segregation cues within the chord's component tones. Apart from onset difference, the most commonly used segregation cue—in music as well as speech—is fundamental frequency (F0), the acoustic correlate of pitch. Two complex tones are most difficult to segregate when they have the same F0, so limiting F0 cues implies that all of a chord's tones have the same pitch—creating the unison, the trivial case of the chord. But as our definition of a chord includes the use of different pitches, preserving the chord's holistic state may be accomplished by controlling the non-pitch features of the chord's components, such that no single tone "pops out" into the listener's attention.

There are a number of cues that may be used for segregation; a few of these were introduced in our earlier discussion of sequential streaming, and a comprehensive review of these cues may be found in Christopher Darwin's summary of auditory grouping factors (Darwin and Carlyon 1995).

- *Frequency*: Two tones may have identical fundamental frequencies and still differ in frequency: the partials that make up each tone's harmonic series may vary in amplitude, creating a difference in the tones' **frequency spectrums**. A familiar example of spectral difference may be found in spoken vowels: we may sing /i/ and /u/ vowels with the same fundamental, and yet hear them as different due to their their different frequency spectra.
- *Loudness*: A chord tone that is significantly louder than the others is more likely to "pop out" and be heard as a separate sound.
- *Spatial location*: Although a tone's source location may be used for segregation, this cue is often not sufficient to elicit segregation by itself, without the presence of other segregation cues.
- *Dynamics*: As suggested by the Gestalt principle of "common fate," stimuli that change in a consistent manner are heard as a distinct object. The most most musically common form of dynamic cue is modulation, the rapid periodic change of amplitude or frequency; common musical forms are vibrato (frequency) and tremolo (amplitude).

2.4.5 Provide a salient pitch percept

Lastly, it is possible to limit the listener's musical motivation for tone segregation—the perception of a pitch-bearing entity—by imbuing the sonority with its own musical pitch. As discussed in section 2.1.2 above, two or more tones with fundamental frequencies forming (harmonic) ratios, such as the octave (2:1) or perfect fifth (3:2), are likely to be fused into a single percept. Pairs of tones in such harmonic relationships offer another acoustic cue to chordal integrity: the simplicity of the ratio between fundamentals is proportionate to the degree of energetic masking or "overlap" among the tones' combined partials. The harmonics of a tone at 200Hz, for instance, are matched by the even-numbered harmonics of a tone at 100Hz, so that their independence is compromised: the two tones are subject to perceptual fusion.

This maneuver obviously restricts the possible combination of tones within a chord, such that only those tones that "line up" within a single harmonic series may be used. Although such a restricted chordal palette is not often used within the triadic practice of tonal harmony, we may note the potential for pitch-bearing chords in earlier Western practices—such as the high-harmonicity open-fifth sonorities used at phrase boundaries in medieval counterpoint. Consequently, the use of such fusion-prone intervals is tightly controlled in later Western polyphonic practices, in which the perceptual independence of each voice is valued (Huron 2001).

2.5 The chord in musical culture

The strategies outlined in the preceding section permit us to speculate on the historical lag discussed in the opening to this chapter. As the mere presence of a notated sonority does little to ensure that it is heard as a chord, describing the establishment of chords within a musical practice requires us to posit agents or events—Helmholtz's "external circumstances"—that motivate the adoption of chordal practice. Although we may point to the advent of the early Lutheran chorale or the introduction of *basso continuo* practice (as Helmholtz does) as early chord-bearing practices, these bits of evidence merely narrow the question temporally and geographically: we are still left to wonder what events might have led to the use of chords in these musical cultures.

Alternatively, we might propose that the experience of hearing notes in harmonic ratios leads the listener to a more holistic level of sonority perception: the listener who is conditioned to hear open-fifth chords as unitary might not allow the presence of a major third (for instance) to disrupt this holistic percept. This is the proposal taken up by Dahlhaus, who sees the emergence of the chord as a consequence of two factors:

> In the idea of the chord as a given entity, it is necessary to distinguish between two aspects: that of psychology and that of musical logic. Stumpf defined or characterized the psychological entity as a "fusion" of the notes in a consonant triad (and to a lesser extent in the chord of a 7th too). The logical factor, however, is to a large extent independent of the psychological, although the conception of a chord as a logical entity could not have arisen in the first place without the psychological phenomenon of fusion. (Dahlhaus 1980)

Although this account of "fusion" as chordal prerequisite is a plausible explanation for the harmonically simpler chords of medieval practice (as noted earlier), Dahlhaus's inclusion of the triad in this category appears to misrepresent Stumpf's definition. While Stumpf did in fact propose *Verschmelzung* (fusion) as the basis for Western musical consonance in his early research (Stumpf 1890), his results suggested that there was little potential for fusion beyond the octave and fifth,[6] and he modified this proposal in his later writings: the consonance of the harmonic triad was more properly considered element of musical logic or "concordance" (Stumpf 1911).

We may supplement this pitch-based theory of chordal origins by accounting for the non-pitch factors that were outlined in the chord-production strategies in section 2.4. A sonority produced with simultaneously struck tones will be initially bound as a chord; the more precisely coordinated the onsets, the fewer cues toward the presence of multiple tones within the chord. A chord with a brief duration or reduced loudness will limit the perception of segregation cues, whereas a more audibly salient chord may control the availability of segregation cues by standardizing the non-pitch features of the chord, such as timbre, modulation, and loudness. Musical contexts that direct attention and higher-level processing away from the chord are more likely to preserve the chord's integral percept: this may be achieved by introducing a separate simultaneous musical stream, such as a melody, or it may be achieved by informational masking, making segregation of individual tones more difficult and dependent upon contextual cues.

With these chord-producing strategies in mind, I propose that the families of polyphonic or chord-producing instruments—the keyboard and the early guitar in particular—that rose to prominence in late Renaissance musical practice were essential to the

---

[6] A follow-up study by DeWitt and Crowder (DeWitt and Crowder 1987) produced similar results, with only octaves and perfect fifths eliciting a listener response of "single tone"—the indication of *Verschmelzung*—instead of "two tones."

cultural formation of the chord as a musical object. We may compare the relevant features of the sounds of these instruments to those of a vocal ensemble, the default performance medium of polyphony in the early modern period. For the most part, polyphonic instruments make use of struck or strummed strings, providing a defined attack, and yet they are able to be dampened to keep duration brief.[7] Whereas a vocal ensemble must coordinate both the primary (pitch, rhythm) and secondary (timbre, loudness) features that could lead to segregation, the instrumental performer retains control of the features; he or she may coordinate with relative ease the onsets and loudness of the multiple tones of the sonority, whereas segregation cues such as modulation and timbre are controlled through the instruments' design. Lastly, polyphonic instruments were often used in ensemble situations, simultaneous with one or more melodic components, such that listeners had no musical goal to challenge the integral nature of the instruments' chords.

This hypothesis—that instrument design contributed to the percept of chords—may be bolstered with the historical evidence of instrument construction. The emergence of chords as musical entities follows not long after the introduction of the keyboards and guitars that enabled the performer to more fully implement the chord-producing strategies discussed above. Although the keyboard's layout enables multiple keys to be pressed simultaneously, early keyboards had widely spaced keys that made chord play-

---

[7] The obvious exception to this qualification is the organ, which is capable of producing tones of indefinite duration. The sustained tones of the organ need not, however, work against the chord percept: during the Renaissance period, organs were constructed so that each key on its keyboard controlled a series of individual pipes. These pipes were typically tuned at harmonic intervals, such that a pipe at the so-called 16′ rank might be reinforced by pipes sounding at its octave (8′) and twelfth (5 1/3′) (*Grove music online,* s.v. "Organ"). In this respect, each tone of the organ was already designed to be perceptually fused; the use of multiple "notes" merely adds more pipes—which are similar in timbre, dynamics, and loudness—to the mixture.

ing difficult.[8] Narrower key spacing is found in harpsichords and organs of the 16th century; the span of an octave averaged under 17 centimeters, which approaches the modern piano's octave span of 16.5 centimeters (*Grove music online*, s.v. "Keyboard").

Similarly, the early guitar may be understood as a chordophone that is optimized for the performance of chords (Tyler and Sparks 2002). In comparison with the lute—the most commonly used chordophone of Renaissance music—the guitar offers a number of features that make chord playing easier: fewer strings (typically four or five courses) allow the left hand to completely fret a chord, permitting a strummed (*rasgueado* or *golpeado*) chord-playing technique that was impractical on the lute. Similarly, the early guitar's re-entrant tuning and use of double-strung courses, occasionally tuned in octaves (*bordón* tuning), allowed the guitarist to strum chords built of many closely-spaced simultaneous tones that promoted both informational and energetic masking of individual tones.

Historical support for this hypothesis may be found in the musical notation of many popular or dance pieces of the Renaissance period. Although the notation itself does not indicate how sonorities would be perceived, we may reconstruct the sounds of the sonorities through our knowledge of instrumental design and practice. Figure 2.5 is an excerpt of an anonymous *pavana* (dance piece) for keyboard dating from the second quarter of the sixteenth century, from the Castell'Arquato collection of manuscripts (Slim 1975). The open-fifth chords that dominate the left-hand texture may be understood as a chordal decoration of a familiar bass ground pattern known today as the

---

[8] Indeed, it is likely that early keyboard instruments—primarily organs—were not used to perform more than one or two simultaneous voices; the earliest written evidence of organ music requiring two independent hands is found in the Robertsbridge Codex (c. 1325), which does include occasional and sustained three-part chords.

romanesca; these chords serve to amplify the bass pitch, bringing the added "secondary" qualities of loudness and modulation, yet without compromising the availability of pitch percepts. The increased loudness and salient duration of these chords is a significant factor when considering the likely instruments on which this dance piece would be performed: as a "home edition" of dance tunes popularized by the court (Apel 1972, chapter 10), the keyboard used to perform these works was likely quite small—typically a miniature, 4′ version of the clavichord known as a *spinettina* (Judd 1995). The need for increased loudness is made apparent by the musical texture of the piece: in comparison with shorter metric values of the melodic stream, the chords' relatively greater inter-onset spacing is partially offset by their greater loudness, providing a longer-lasting sound.
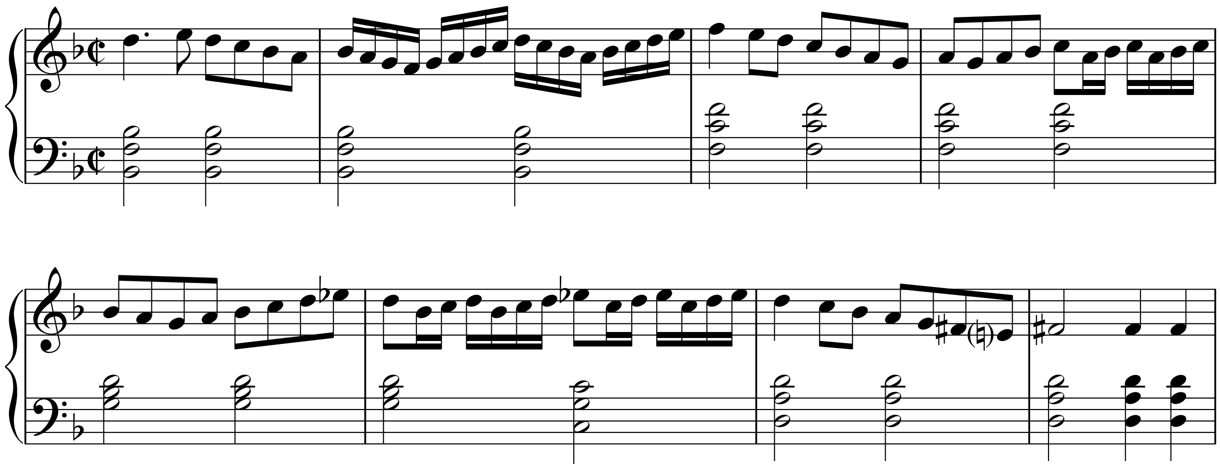
Figure 2.5: *Pavana* (No. 4), from Castell'Arquato collection (Slim 1975), measures 1–8

Comparing a work like this to the more contrapuntal keyboard works found in contemporary German sources, the historian of triadic harmony may bemoan "the obstinate retention of the favorite fingering" 1–5–8 in the left hand, which may be related to "primitive improvised polyphony, which one can still come across in the Italian country.” (Jeppesen 1962, x) Yet if we detach chordal practice from triadic harmonic practice, we may understand the chords in this piece to be an instrumental codification of the more culturally established use of perfect (octave and fifth) sonorities. And with this mindset, we may note the two-handed “major triads” spanning both staves of the grand staff in the final bar of this excerpt, and the “root-position minor triad” in the bass staff in bars five and six—part concession to triadic practice, part necessity of keeping the performer’s two hands separated—as recorded instances of the integration of chordal sonorities and triadic harmony.

Having identified the “fusion” of harmonic sonorities and the introduction of instrumental technology as separate yet coeval factors, we are still left with the task of considering how these two largely physical phenomena combined with Western polyphonic culture to initiate the production of chords. To model the influence and interaction of these factors, I turn to Marc Leman’s ecologically based theory of cultural resonance (Leman 2008, chapter 3). Leman distinguishes two categories of restraints—cultural and natural—that dynamically converge within the subject (musician) to form a cultural “resonance system,” as modeled in Figure 2.6; it is through this resonance system that “the interaction between natural and cultural restraints can lead to higher levels of accumulated abstraction and complexity” (Leman 2008, 68). As this dynamic process unfolds, the resulting cultural principles or schemata may reach a level of abstraction such that they become separated from their physical influences; the natural constraints at the base of the are transformed into cultural concepts.

This model may be readily adapted to explain the chord influences outlined above, which I propose form two significant resonance systems, dynamically proceeding in

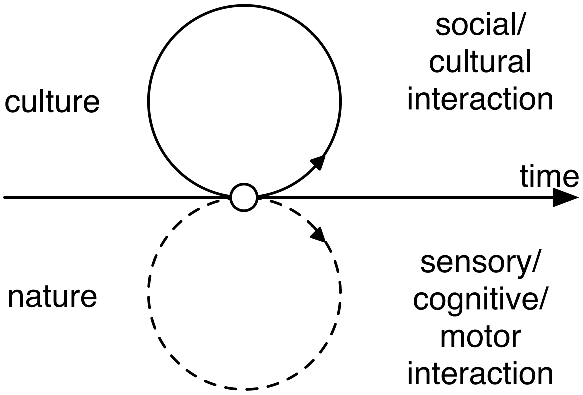

Figure 2.6: Model of an individual's interaction within a resonance system, adapted from Leman (2008, chapter 3)

what Leman describes as the "ratchet effect" (Tomasello, Kruger, and Ratner 1993); this proposal is illustrated in Figure 2.7. The first resonance system represents the musician's interaction with the natural constraints of harmonicity and the cultural constraints of medieval polyphonic practice. Of the different sonorities employed in medieval polyphonic practice, only the perfect consonances (that is, the octave and the fifth) conform with the natural constraints of harmonicity needed to promote perceptual fusion; in this manner, polyphonic culture expands to represent highly harmonic sonorities as single musical entities.

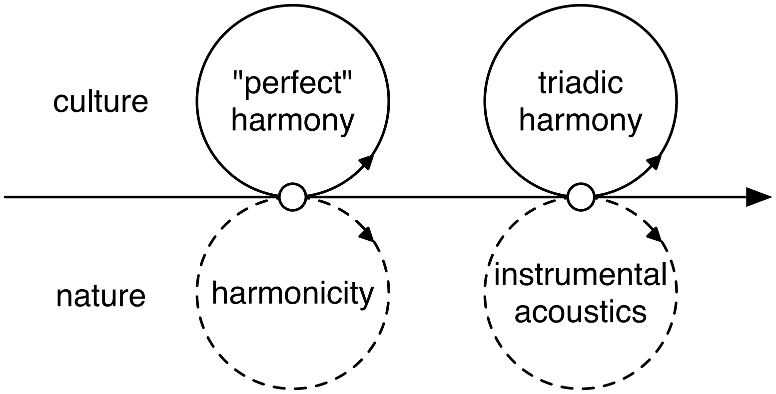

Figure 2.7: The introduction of triadic chords through multiple resonance systems

The second resonance system models the musician's interaction with triadic harmony and instrumental technology. The musician's creation and use of polyphonic instruments within this system constitute a kind of loosening of natural constraints: the perceptual fusion of sonorities may be achieved by means other than harmonicity.[9] As the musician introduces these instruments within polyphonic practice, he or she adopts the harmonic language of the polyphonic culture: the harmonic triad, only occasionally fusion-prone in the practice of vocal polyphony, is taken as the structural syntax of polyphonic instruments. In this regard, we may understand an early instrumental piece such as the *Pavana* not as bearing "primitive" harmonies, but as predating the complete convergence of instrumental technology and triadic syntax: it is an earlier artifact of the cultural resonance system that eventually produced triadic chords.

It is impossible, of course, to model within a single graph the complex interaction of acoustic and cultural conditions that engendered the practice of chords. For any given period or area, we may identify conditions that suggest greater or lesser influences of the constraints modeled above: the southern European *falsobordone* practice—the "thickening" of a psalm tone with static triadic sonorities (*Grove music online,* s.v. "Falsobordone")—and the homophonic Lutheran chorale both suggest that triadic "fusion" was also an active element in the development of chordal culture. But it is my claim that this harmony-based component was necessary but not sufficient for the development of chordal music: it operated in tandem with instrumental design and practice.

---

[9] Labeling the influence of musical instruments as a "natural" constraint is somewhat misleading: instruments are made by humans within a musical culture. Rather, it is the acoustic capabilities of these instruments that alters the natural constraints of sonorities. The impetus behind the instruments' design—likely spurred by a desire to more easily produce chordal sonorities—forms part of the cultural constraints that interact with the instruments' sounds.

2.6 Conclusion

The perceptual account of chords given in this chapter adds an experiential element to our historical account of triadic harmony. We understand our domain-general processes to operate in relatively predictable fashion given a particular sound stimulus, and we may identify the traits of musical listening that lead to specifically musical processes of object-formation. This combination of domain-general and musical factors allows us to outline performance strategies for creating sonorities that are heard as chords—and, in turn, to identify the musical practices that adopt these strategies to the greatest degree.

As noted in section 2.4, the strategies provided in this chapter tend to favor domain-general, "brute force" methods that manipulate the acoustic composition of sonorities and their accompanying (non-chord) musical objects. This approach is largely pragmatic: an emphasis on acoustics allows these strategies to be used effectively in a wide range of musical practices. An account of the strategies that draw upon musical context and listening experience requires a more contextual approach, in which the controlled manipulation of musical elements leads to observable differences in percept formation. This approach is taken up in the experiments that are presented in the following chapter.

CHAPTER 3

EFFECTS OF EXPERIENCE AND CONTEXT ON CHORD PERCEPTION

The compositional strategies of chord formation presented in chapter 2 provide a way to assess the likeliness of hearing a sonority as a chord. Although there are several strategies available to the musician to promote a chordal percept, the most effective is the use of onset synchrony—the temporal alignment of the sonority's tones. Onset synchrony is not, however, a guarantee of chordal listening: the listener's understanding of the musical significance of a sonority—as an integrated whole, or as a single focal part within a harmonic "background"—in turn shapes his or her listening task as either **holistic** (hearing the chord) or **analytic** (hearing its components) in nature. Additionally, musical context may promote the segregation of one of a sonority's tones within a melodic stream; the voice-leading strategies of Western polyphony have been shown to promote such streaming, such that the listener is better able to follow one or more voice parts in a harmonic passage (Huron 2001). It is therefore not sufficient to rely upon the acoustic structure of a sonority to assess its perceptual state; where possible, we must also examine the influence of experience and context on perceiving the sonority.

The assessment of cognitive influences in listening is often done through controlled listening experiments; a listener is asked to make decisions or judgments about a heard musical stimulus, and the facility with which they perform these tasks—measured in response times or error rates—can be used to support a hypothesis about our perceptual processes. Accordingly, in this chapter I present two experiments that examine the

integrality of a heard sonority—the extent to which the features or components of the sonority are perceptually inseparable. By proposing the null hypothesis that, all other things being held constant, experience and context do *not* affect the listener's adoption of chord-level or tone-level perception, I look to measure differences in listener responses that suggest specific conditions in which these cognitive influences might be present.

In the case of the musical chord, we may draw upon the listening strategies set out in section 1.3 to propose a general framework for assessing the integrality of a heard sonority. If a listener is asked to make decisions about a single tone within a sonority, he or she will often perform this task more effectively—with faster reaction times and lower error rates—when he or she perceives the tone as a separate musical object, rather than perceiving the chord. Similarly, decisions about emergent properties of a sonority will in general be faster and more correct when the listener perceives the chord instead of component tones. In the experiments that follow, I have chosen the tone as the object of the listener's tasks; this choice permits the use of pitch—an established musical feature of the tone—as the feature to be assessed and judged by the listener.

There are several studies that have asked listeners to make decisions about sonorities at either the holistic or analytic level; these studies have provided valuable insight into how our perceptual level influences our ability to detect emergent features such as pitch distance (Borchert, Micheyl, and Oxenham 2011) and harmonic (simultaneous) dissonance (Bigand et al. 2003). However, these studies commonly manipulate the onset synchrony of the sonority as a domain-general means of directing the listener's percept; these tasks are not suited to measure the specific influence of music-cognitive factors, such as context and experience, independent of the sonority's onset alignment.

The experiments presented in this chapter do not use onset alignment as a means of manipulating the sonority's integrality: the first experiment involves arrangement of the component tones' frequencies to form familiar or less-familiar sonorities, and the second experiment measures the influence of a variable pre-sonority musical context on how listeners hear a single tone within the sonority that follows.

3.1 Experience and chord formation

Research in the field of categorization—how we recognize and differentiate objects—has played an important role in our understanding of humans' cognitive capacities cognition (Bechtel, Graham, and Balota 1998). A series of experiments by Eleanor Rosch and colleagues (Rosch et al. 1976) showed that an object is most readily categorized at what has come to be called the *basic level*, which reflects our most common interactions with that object. For instance, when subjects are asked to describe the features of an *apple*, they respond more efficiently than if asked to describe the features of *fruit* (a superordinate level) or a *Mackintosh apple* (a subordinate level). The basic level has also been shown to be dependent upon experience, such that trained or "expert" participants may have a basic level that different from untrained participants (Tanaka and Taylor 1991).

Although the results of categorization research suggest that familiarity with an object leads to more optimal decisions and judgments about the object, they do not directly address the issue of object formation. Where we might speak of "concrete objects" or "natural objects" (Rosch et al. 1976) with a moderate amount of certainty in visual perception, our formation of auditory objects is less well understood (Adams and

Janata 2002). This is particularly true for our understanding of musical objects, which are formed according to both domain-general and music-specific principles. Recent research has shown that listeners asked to make decisions about sequences of tones tend to form objects of brief segments of tones, or "tone words" (Saffran et al. 1999), to facilitate these decisions; this ability to form sequential objects supports a model of musical listening that takes motives (Zbikowski 2002, chapter 1) or melodic "chunks" (Godøy 2009) as the primary objects of musical experience.

Similarly, listeners familiar with Western musical practice are likely to adopt strategies that take commonly occurring sonorities as musical objects. Highly familiar sonorities are likely to form a kind of model or exemplar by which other, less familiar sonorities are judged. One common strategy for examining this familiarity is the use of triadic sonorities composed with a variably-tuned triadic third, along a continuum from minor (3 semitones) to major (4 semitones) in pitch distance above the triadic root (Locke and Kellar 1973; Howard, Rosen, and Broad 1992; Klein and Zatorre 2011). As major and minor triads are much more familiar to Western listeners than sonorities with a neutral or "in-between" third, listeners were better able to make decisions about sonorities with thirds tuned toward the major or minor extremes of this continuum; notably, this difference was most significant in listeners with musical training (Howard, Rosen, and Broad 1992).

With respect to perceiving chords, the studies just listed suggest that familiar sonorities are more likely to be perceived holistically: the listener uses their prior experience to inform their formation of objects (Kersten, Mamassian, and Yuille 2004). Evidence for this suggestion is found in a series of studies (Pastore et al. 1983; Collins 1985) that use triadic sonorities composed of binaural stimuli presented over headphones: the triadic third of the sonority was presented to (for instance) the listener's left ear, while the remaining "frame" of root and fifth was presented simultaneously to the right ear. The results of these studies show that listeners commonly perceive the complete triadic chord to be sounding in one ear, with the isolated tone in the other; this phenomenon, known as *duplex perception* (Rand 1974), attests to the perceptual "goodness of fit" that leads to hearing the complete chord in a single ear. Significantly, a variant duplex perception study (Hall and Pastore 1992) found this effect to hold even in cases where the isolated tone was presented at levels below the listener's audibility threshold; listeners could still identify the sonority as major or minor, but operated at chance levels when asked to detect which one of two sonorities heard in succession contained a third (either major or minor).

More direct evidence of the integrality of major-triadic sonorities comes from a study by Acker and Pastore (1996), which is apparently the only study published in peer-reviewed journals that takes the integrality of sonorities as its primary research question. In this study, listeners heard a sonority composed of the tones C4–E4–G4—a root-position, close-position major triad—in which both the triadic third and fifth were subject to slight variations in tuning. To examine the integrality of these sonorities, the authors adopted a paradigm known as the *Garner interference test* (Garner 1974), which is designed to isolate the perceptual influence of one dimension on decisions made about the other dimension. In this experiment, listeners heard two sonorities in succession, and they were asked to decide if a *target tone* (for instance, the E4) did or did not vary in tuning across the two sonorities. The authors found that tuning differences in the non-target tone (in this instance, the G4) affected listener's judgments of the target tone, even though the non-target tone offered no contextual "assistance" in making this judgment; the results suggest that the E4 and the G4 were perceived integrally, such that adjustments in either tone affected the percept of both.

The experiment presented in the following section expands the work of Acker and Pastore by examining the perceptual integrality of different sonorities deemed to be either familiar or unfamiliar to listeners; these sonorities were presented to listeners in a Garner interference test, such that perceptual integrality may be examined with respect to the listener's familiarity with the stimuli.

3.2 Experiment 1: Garner interference test for sonority integrality

3.2.1 Design

The following experiment adopts the Garner interference paradigm, in which stimuli are organized along two different dimensions; each of these dimensions is represented by two discrete values, providing a total of four different stimuli. Figure 3.1 shows the musical notation[1] for the four stimuli composed for this experiment: stimuli were organized along two dimensions, consisting of a single target tone (labeled X or Y) and a two-tone frame (labeled A or B). Listeners were asked to make decisions about the target tone. The null hypothesis was that listeners' performance in this task would

---

[1] The diamond-shaped noteheads used in Figure 3.1 indicate the use of pure tones, rather than the more musically common harmonic complex tones; a fuller description of the stimuli used in this experiment is given in section 3.2.2.

be unaffected by the use of different frames, suggesting that the tone and frame were perceptually separable.

The frequencies of the tones used to create these stimuli were chosen to form sonorities that are either common or uncommon within Western musical practice. Two of the sonorities formed, labeled AX and AY, belong to the class of major and minor triads, generally described as "consonant." The other two, labeled BX and BY, are drawn from the class of three-note sonorities that do not form triads in the music-theoretic sense, and are generally described as "dissonant."

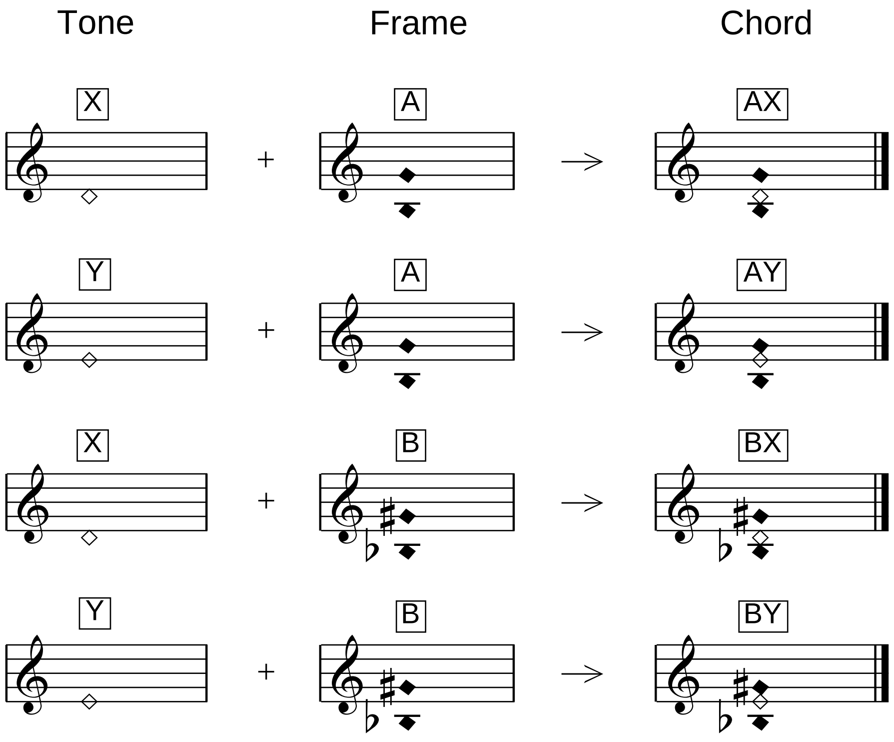

Figure 3.1: Target tones and sonority frames used to compose stimuli

The stimuli were arranged and combined to form five separate *conditions*, shown in Figure 3.2, that are designed to measure the influence of the frame upon perceiving the target tone; within the experiment, each listener hears a block of stimuli from a single condition before moving on to the next condition. Two *control* conditions are composed of sonorities that use the same frame, such that only the target tone changes within each condition. Two *redundancy* conditions use sonorities that differ in both frame and target tone: in these conditions, a change in target tone is correlated with a change in frame, giving the listener additional (redundant) cues to recognizing the target tone. A *filtering* condition is composed of all four sonorities, in which both frame and target tone vary independently; in this condition, the listener must "filter out" the variations in the task-irrelevant frame.

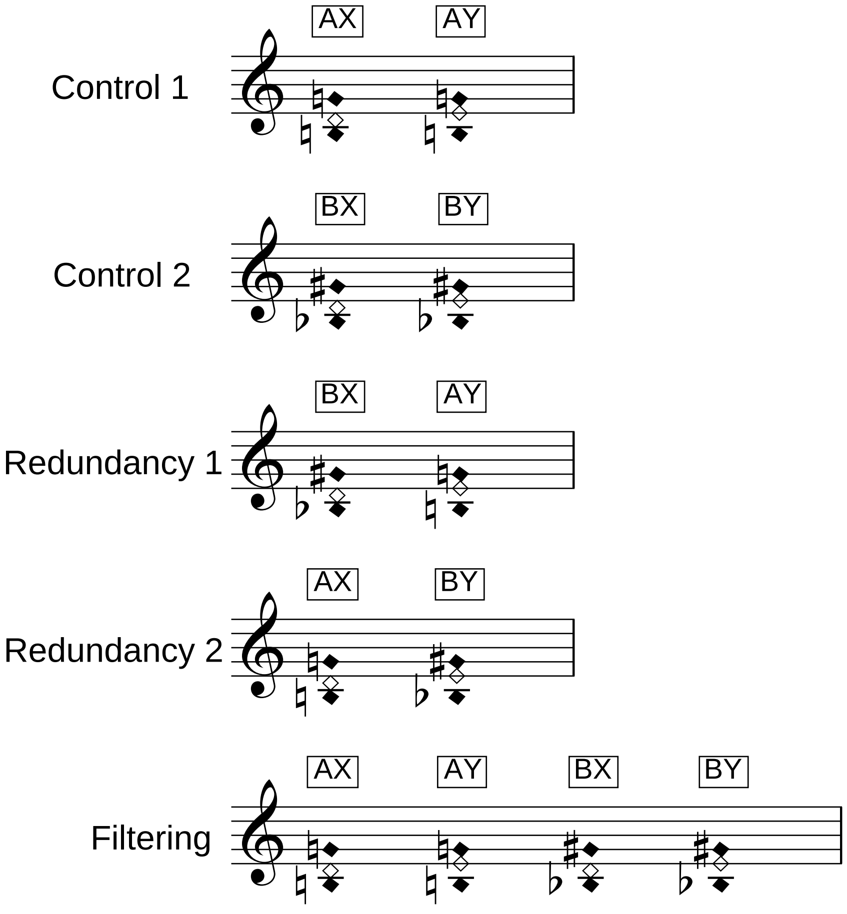

Figure 3.2: Conditions used in Garner interference test

3.2.2 Method

Participants

Participants were 66 undergraduate students from the University of Chicago, with an average of 5.7 years of musical training. None reported abnormal hearing; three participants reported possessing absolute pitch. All participants gave written informed consent prior to the experiment. Participants were naïve to the purposes of the experiment, and they were given course credit for participation.

Materials

All stimuli were composed of sine tones of 1 second duration. The amplitude envelope of each tone was a 15ms linear attack, followed by 185ms at peak amplitude, and a linear decay for the remainder of the duration. The onset times of the three tones within a sonority were adjusted so that the middle tone was delayed by 50ms; this was done to facilitate detection of the middle tone. The last 50ms of this middle tone was truncated so that all tones of the sonority ended synchronously. All tones were created using Amadeus Pro software using equal-temperament frequencies.

Procedure

The experiment presentation was created using E-Prime software, version 2.0. Sounds were generated on a PC using a Creative Labs SoundBlaster sound card, presented over Sennheiser HD–570 headphones and a Realistic HD–150 headphone amplifier; the amplified was adjusted so that a sonority was produced at approximately 70dB sound pressure level (A-weighted). Participants were seated at a computer terminal and were asked to follow the instructions given on the display; visual feedback ("correct" or "incorrect") was provided after each response in the training and trial phases. Response times and error rates were recorded in E-Prime.

Each participant began with two separate training phases before proceeding to the trial phase. The first training phase began with the presentation of the two target tones (X and Y) used in the stimuli; these tones correspond to the musical pitches D4 and E4, respectively, and they were labeled "low tone" and "high tone." Each tone label was associated with a key on the computer keyboard; participants indicated which tone they heard by pressing one of the two keys. In this tone training phase, the target tones of the stimuli were heard in isolation, as single tones; participants were asked to classify the tone as either "high tone" or "low tone."

A second training phase presented the sonority stimuli taken from the filtering condition—that is, any one of the four sonorities used in this experiemnt—to familiarize participants with the trial phase task; participants were asked to classify the middle tone of a heard sonority either the high tone or low tone. Both training phases consisted of 16 stimuli; at least 12 of 16 correct responses were required before before proceeding to the next phase. Less than 12 correct responses in either training phase led to a repeat of the phase, and more than six attempts in either phase led to the termination of the experiment.

The trial phase consisted of each of the five conditions (two control, two redundancy, and filtering) presented in Figure 3.2. There were 24 stimuli in each condition, resulting in 120 total trials, and the order of conditions was varied for each participant using a Latin square design. As in the second training phase, participants were asked to classify the tone within a heard sonority as either the high tone or low tone. Participants were asked to respond as quickly and accurately as possible.

3.2.3 Results and discussion

Twenty participants did not progress past the training portion of the experiment, leaving 46 participants for the trials. Of these remaining participants, six performed at near-chance levels in the trial phase; these participants were excluded from the analysis, leaving 40 participants' results to be considered.

Both error rates and reaction times were analyzed using a one-way repeated measures analysis of variance (ANOVA), with condition as a five-level, within-subject factor (Snedecor and Cochran 1989). Two populations were analyzed: all participants (n=40), and those with overall error rates (across all five conditions) below 10% (n=20). Results are shown in Figure 3.3 for both populations.

Response times and error rates were averaged for the two control and redundancy conditions and compared to the Filtering condition: this comparison did not reveal significant differences in response time, though error rates for the filtering condition were significantly higher than all other conditions, $p \leq .05$. Comparing individual conditions separately—that is, a five-way comparison—revealed significant differences in response time: participants were faster in their responses to the Control 1 (sonorities AX and AY) and Redundancy 1 (sonorities AY and BX) conditions when compared with the Filtering condition (all four sonorities), $p \leq .01$. These differences were magnified in the pool of participants with overall error rates less than 10%; in this group, response times in

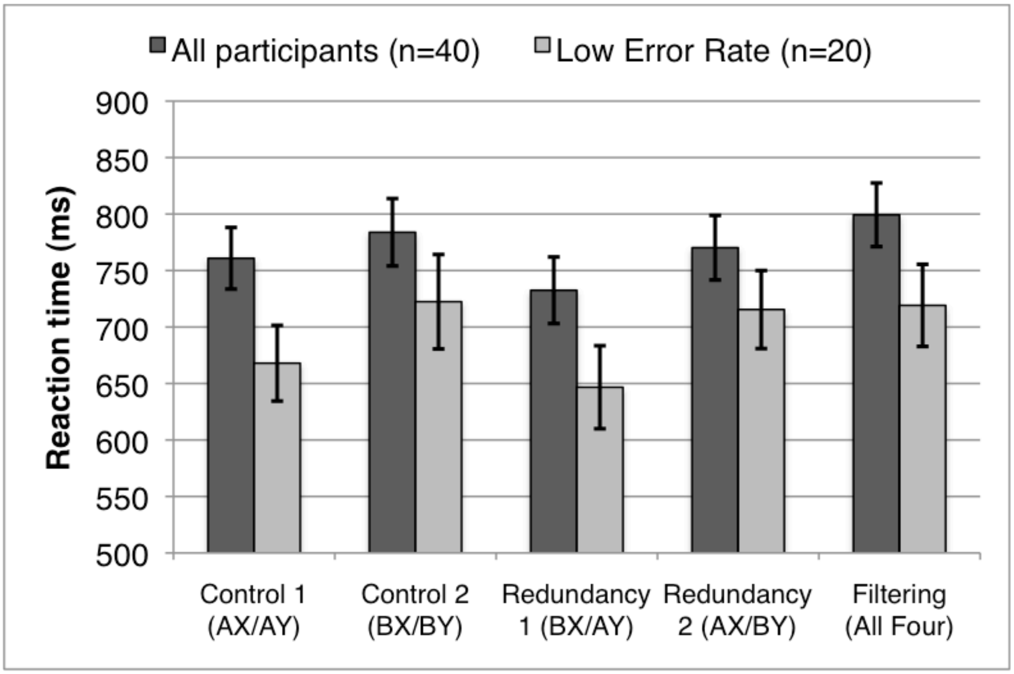

Figure 3.3: Reaction time by condition for all (n=40) and low error rate (n=20) populations; error bars represent standard error

both Control 1 and Redundancy 1 conditions were also faster than in the Redundancy 2 condition (sonorities AX and BY), $p \leq .01$.

As results were inconsistent across the control and redundancy conditions, they do not suggest either separability or integrality within the Garner interference paradigm; this is likely due to the use of "dimensions" that are not typically perceived as features. In particular, these results suggest that absolute frequency was difficult to perceive as a feature of these sonorities; given the rareness of absolute pitch abilities in listeners, this is not surprising, and the comparatively brief training phases used in this experiment were not sufficient to orient listeners toward this "feature."

Despite the lack of integrality or separability results, the differences in response time suggest that less familiar sonorities inhibit the ability to make decisions about one of its component tones. This phenomenon is especially observable in the low-error-rate group, in which response times for the two conditions lacking the "whole-tone" sonority (BY) were significantly faster than two of the three other conditions. This suggests that participants relied upon perceiving the chord, rather than the tone, in completing the task, and that the chord provided a stimulus context effect that facilitated the task (Ashby and Maddox 1994). For example, participants could have adopted a strategy that associated familiar sonorities with the position of the target tone. The low tone would be correlated with the major triad (AX), as well as an "augmented sixth" sonority (BX) that could be enharmonically interpreted as an incomplete "dominant seventh"; similarly, the minor triad (AY) would be associated with the high tone. In this scenario, the "whole-tone" sonority, offering no familiar percept, would more resistant to being used in a contextual strategy.

3.3 Context and chord perception

Although it seems likely that our perception of musical features would be influenced by the musical context in which they appear, the empirical study of contextual influence in music has only recently been taken up extensively. One notable exception is the research of Robert Francès, who demonstrated that that listeners hearing a leading tone proceed to the tonic—such as the succession B3–C4 in a C-major context—were able to detect small deviations in the leading tone's F0, but only when they conflicted with its perceived tendency to resolve upward. When listeners were asked to detect mistunings of the B3 leading tone, listeners were more able to detect lower tunings—away from the tone's musical tendency—than they were higher tunings of an equivalent musical (logarithmic) distance (Francès 1958/1988, chapter 3).

Similarly, the musical context surrounding a sonority may influence our abilities to discriminate its emergent musical features. In many sonority-discrimination studies, the listener hears a harmonic progression that both defines a tonality and suggests an appropriate closing or cadential harmony—for example, a context ending with a dominant harmony that suggests a concluding tonic harmony. A target sonority that follows this context may either conform to or violate this harmonic implication: in the given example, a tonic harmony is syntactically appropriate, whereas a subdominant harmony is not. When listeners are asked to make decisions about the target sonority, they routinely perform with greater accuracy and faster response times when its harmony is expected (tonic) than when it is not (subdominant). Such context effects have been observed for tasks of detecting mistunings within the sonority (Bharucha and Stoeckig 1986; Warrier and Zatorre 2002), detecting added "dissonant" tones in the sonority (Bigand et al. 2003), and categorizing the sonority as major or minor (McMurray, Dennhardt, and Struck-Marcell 2008).

Although the studies outlined suggest that our perception of sonorities is subject to the contexts in which they appear, these studies do not specifically address how the sonority is perceived—as a chord, or as multiple tones. Our knowledge of auditory scene analysis suggests that different pre-sonority contexts would have different effects on the listener's percept of a target sonority. For instance, a single melodic stream that precedes a sonority gives the listener a contextual cue to perceptually capture one of the tones of the sonority within this stream; in contrast, a homophonic context would condition the listener to form a “chord stream” that promotes hearing the target sonority as a chord (Cambouropoulos 2008). The experiment in the following section uses contexts that vary in harmonic texture to assess their potential effect on the listener’s ability to “hear out” a single tone from a three-tone target sonority; the null hypothesis states that performance in this task will not vary significantly by the type of context heard.

3.4 Experiment 2: effects of pre-sonority context

3.4.1 Design

This experiment uses a *probe tone* paradigm that has been used in chord perception studies by DeWitt and Samuel (1990) and Hubbard and Datteri (2001). In this paradigm, the listener heard a sonority followed by a probe tone, and he or she was asked to determine if the probe tone was heard in the preceding chord. Although this paradigm does not offer direct insight into the perceptual integrality of the chord—the listener makes a decision *after* hearing the chord, not *while*—it is effective for measuring the separability of a tone from a sonority, and we may interpret the results of the experiment below to have some bearing on how the listener hears the sonority in question.

Stimuli for this experiment were composed of three variable components: a pre-sonority context, a target sonority, and a probe tone, as shown in Figure 3.4. Each stimulus contained one of three contexts, which lasted for 2 seconds or four “beats” in duration:

- a Chord context, consisting of three-tone sonorities with simultaneous tone onsets;

- a Melodic context, consisting of the highest "voice" from the Chord context;
- a Disjoined context, consisting of octave transpositions of the Melodic context.

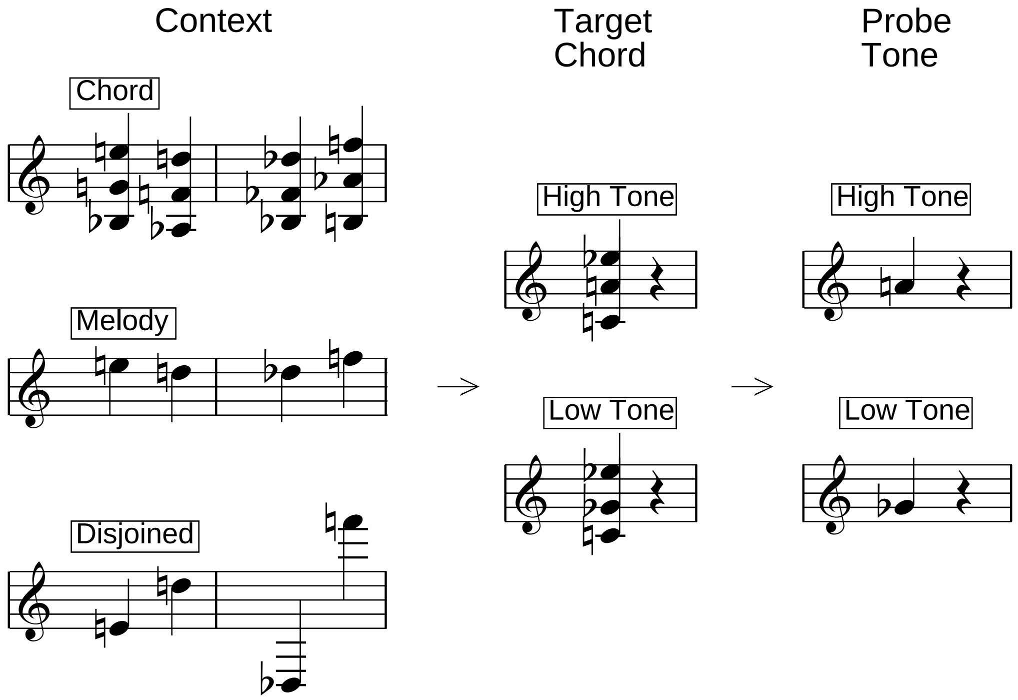

Figure 3.4: Contexts, target chords, and probe tones used in stimuli

Following the context, one of two different target sonorities was presented: the middle tone of the target sonority was one of two varieties, "high" or "low," as shown in Figure 3.4, whereas the outer tones remained in the same position. A probe tone followed the target sonority; the tone was identical to either the high or low tone of the target sonority, with an equal chance of being present or absent within the target sonority.

Context, sonority, and probe tone were assembled to form a musical stimulus, as shown in Figure 3.5; rests were added between context, sonority, and probe tone, producing a duple-meter passage with a tempo of quarter note = 120.

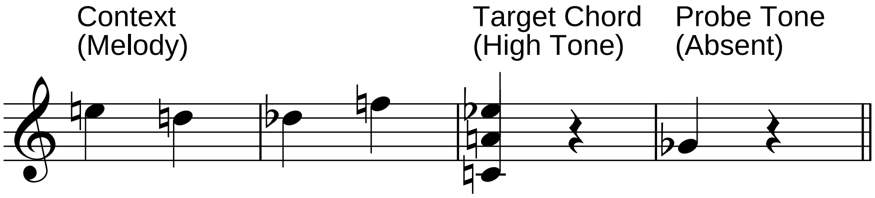

Figure 3.5: Example context, target chord, and probe tone presented in sequence

All contexts and target sonorities were composed of non-triadic (“atonal”) pitch content to minimize the influence of pitch center and harmonic syntax on perceiving the target sonority, and the pitch classes of the target sonority are not present within any of the three contexts.

3.4.2 Method

Participants

Participants were 72 undergraduate students from the University of Chicago, with an average of 5.6 years of musical training; 36 of these participants also participated in the trials for Experiment 1. None reported abnormal hearing; four participants reported possessing absolute pitch. All participants gave written informed consent prior to the experiment. Participants were naïve to the purposes of the experiment, and they were given course credit for participation.

Materials

All stimuli were composed of synthesized piano tones, using Sibelius 6 with Garritan Personal Orchestra Lite sound samples. Each tone used in compiling the stimuli was 500 milliseconds in duration. All 12 combinations of context, sonority, and probe tone were reproduced at three different transpositions, with the lowest tone of the target sonority at C4, B♭3, and A♭3, respectively, for a total of 36 stimuli.

Procedure

Experiment presentation was conducted as in Experiment 1 (section 3.2.2). Each participant started with a separate training phases before proceeding to the trial phase; to establish the meter of the stimuli, the training phase used a percussive (cowbell sound) pre-sonority context in place of the pitch-containing contexts used in the trials. The training phase consisted of 12 stimuli; at least 9 of 12 correct responses were required before before proceeding to the trial phase. Less than 12 correct responses led to a repeat of the phase, and more than six attempts at the training phase led to the termination of the experiment.

The trial phase consisted of a single session with 72 trials; each combination of context, sonority, and probe tone was presented six times, twice in each of the three pitch transpositions. The participant was informed that a musical fragment would be heard, followed by a chord and a tone; he or she was asked to ignore the fragment and indicate whether the tone was present or absent within the preceding chord. Participants were asked to respond as quickly and accurately as possible. Visual feedback was given after each response.

3.4.3 Results and discussion

Twelve participants did not progress past the training portion of the experiment, leaving 60 participants for the trials. A signal sensitivity (*d′*) score was calculated for each of the three context conditions; *d′* was calculated as the z-transforms (number of standard deviation units) of hits minus false alarms (Wickens 2002, chapter 2). *d′* scores were analyzed using a one-way repeated measures ANOVA, with context as a three-level, within-subject factor; results are shown in Figure 3.6. This analysis reveals that *d′* scores for the Melodic context condition were significantly higher than those in the Chord context condition, $p \leq .05$. No other significant results were observed.

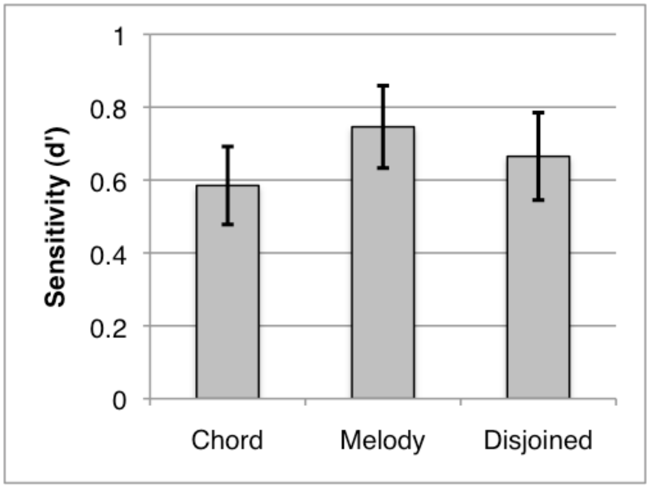

Figure 3.6: Signal sensitivity by context; error bars represent standard error

The results of this experiment suggest that the listener's ability to analytically perceive a sonority may be mediated by the context preceding the sonority; hearing a melodic context primes the listener for a single-tone percept, whereas hearing a homophonic context primes the listener to perceive a chord. These results appear to support the concept of "chord streaming" as described by Cambouropoulos (2008): in a highly homophonic or chordal context, the listener is more likely to hear the sonorities within a single auditory stream, rather than hearing and maintaining segregated streams of tones or voices.

3.5 General discussion

The task of hearing out the middle tone of a three-tone sonority was selected for its difficulty: the results of previous studies (DeWitt and Samuel 1990; Palmer and Holleran 1994) suggest that the middle tone is the most difficult to discriminate. Whereas the use of middle-tone comparison tasks discourages segregation of sonorities' upper or lower tones to form a melodic stream, it is possible that only musically trained listeners will be able to perform these tasks consistently at above-chance levels; future experiments using such tasks will draw upon a participant pool of trained musicians, either as a whole or in comparison with subjects without formal musical training.

Although the Garner interference test is a promising paradigm for measuring sonority integrality, the use of absolute frequency of tones as a perceptual dimension is likely without basis, in that nearly every musical culture makes use of relative, not absolute, pitch (Huron 2006, chapter 7; Bispham 2009). Other perceptual dimensions of the middle tone, such as instrumental timbre, might be used; however, it is likely that changes in the middle tone's timbre will have an effect on the holistic percept of the sonority, such that notationally identical sonorities with differing middle-tone timbres might be categorized differently (Kendall and Vassilakis 2010). Alternately, the tuning paradigm from Acker and Pastore (1996) could be expanded to vary the tuning of tones in both familiar and unfamiliar sonorities within Western music.

The context paradigm from Experiment 2 could be modified to present a musical context as a form of probe that precedes the target sonority. The results of Hubbard and Datteri (2001) suggest that using a pre-target probe, instead of the post-target probe used in Experiment 2, will lead to more accurate participant results. One possible form of this pre-target paradigm involves a higher/lower comparison task, asking the listener to compare the last tone of the melodic context with the middle tone of the sonority; increased response time or error rate would suggest that the sonority was more resistant to segregation.

Lastly, the tones used to form stimuli in these experiments included both sine tones (Experiment 1) and complex tones sampled from a musical instrument (Experiment 2). Although it is desirable to use stimuli that are as similar as possible to the sounds created in musical practice, the use of sine tones has the advantage of controlling for energetic masking; the lack of upper partials eliminates the possibility of conflicting (rough) or aligning (harmonic) partials within the stimulus. In order to more closely approximate musical sounds, Experiment 1 used a "piano-like" dynamic envelope, with a short attack and a length decay, to simulate the dynamics of piano tones. Controlling for the effects of upper partials may also be accomplished by using both sine tones and complex tones; comparing results across similar participant pools would indicate if these stimuli were processed differently.

Taken together, these experiments suggest that hearing a sonority either holistically or analytically may be subject to the same top-down influences that have been observed in perceiving the features of a single tone or melody. As both experiments examine the listener's ability to discriminate features of the tone, the logical complement of this strategy would be to assess the listener's perception of emergent features of the chord. The following chapter explores the possibility of musical pitch—the feature most commonly associated with the tone—as an emergent feature of the chord.

CHAPTER 4

PERCEIVING AND PRODUCING CHORD PITCH

The "missing fundamental" effect—hearing a pitch from a complex tone that contains no sound energy at its fundamental frequency—may be observed in our everyday listening activities. We understand that the "tinny" sounds heard over a copper-wire ("land line") telephone connection are due to the limitations of the phone system's transmission technology, which is unable to reproduce sounds below about 300 Hz; even so, we may hear pitches below this threshold—from the voiced sounds of a male speaker, for instance—by hearing the complex tone's upper harmonics.[1]

This ability to perceptually reconstruct the pitches of missing-fundamental tones has been used to suggest a similar effect in hearing missing fundamentals for musical chords: the tones of a holistically perceived sonority may form a harmonic series and produce a pitch sensation at a fundamental that may or may not be present within the chord (Terhardt 1984). As this pitch percept may be interpreted as the fundamental or "root" of the sonority, an antecedent of this theory can be found in the writings of Jean-Phillipe Rameau, who derives the *son fondamentale* of a sonority from the harmonic series formed by its tones (Duchez 1986). The following passage from Rameau describes how this fundamental may be heard; of particular note is Rameau's emphasis on the role of *sous-entendre*—translated as "understand," but also forming a compound word meaning "hear under"—in perceiving the fundamental sound of a sonority:

---

[1] Although the missing fundamental effect was first demonstrated by Seebeck (1841), its validity was still in question over a century later; J. C. R. Licklider's "duplex pitch" theory (Licklider 1951) is acknowledged as the first pitch theory to fully account for this effect (Plomp 1991).

> By the word *sous-entendre* one must be made aware that the sounds to which it is applied can be heard in chords in which they are not in fact present; and with regard to the fundamental sound, it is even necessary to imagine that [this sound] must be heard below the other sounds, when one says that it is *sous-entendu*. (Rameau 1722, Table of Terms; translated in Cohen 2001b)

What makes this passage even more remarkable is that it appears in Rameau's *Traité de l'harmonie* of 1722, a work which predates his adoption of the harmonic series—the basis of his later concept of the *corps sonore* (Christensen 1987). In this regard, Rameau's claim is an impressive assertion of the capabilities of the listener: he or she is asked to "imagine" hearing a sound that is not within the chord.

This chapter looks at the perceptual basis of Rameau's rather forward-looking claim—that we may hear a fundamental pitch from a sonority, derived from the harmonic series formed from its partials. I begin with a summary of our current understanding of the pitch perception process, focusing on the acoustic and cognitive factors that are most significant to this process. I then turn to an analysis of pitch models that attempt to capture these factors. The **virtual pitch** model of Terhardt (1974) is adapted to analyze a range of musical sounds—including chords—in an effort to pin down the nature of chord pitch as it relates to the more musically common phenomenon of tone pitch. Lastly, the results of these analyses suggest musical scenarios or environments in which chord pitch may be particularly relevant to musical listening.

4.1 Defining pitch

Although it may seem surprising, the phenomenon of pitch is still a subject of current research and discussion; in this section I review some of the known aspects of pitch perception, with an emphasis on how these might apply to musical chords. A comprehensive review of pitch perception may be found in Plack et al. (2005); the following sections highlight the topics from this review, as well as present newer research that is particularly relevant for the perception of chords.

Whereas pitch is commonly associated with a tone's fundamental frequency (F0)—as when one tunes a violin's A string to 440Hz—pitch is not a physical property of sound stimuli: it is a perceptual feature that is associated with a sound's spectral content. As such, it is difficult to provide a definition of pitch that is both objective and functional: the American National Standards Institute (ANSI) defines pitch as "the attribute of auditory sensation in terms of which sounds may be ordered on a scale extending from low to high" (ANSI 1994). Significantly, the "scale" in this definition need not be the pitch scales common within many musical cultures: we may conceive of several perceptual dimensions of sounds, such as brightness (spectral centroid) or loudness, that permit organization from low to high, and it is possible to perceive and recognize brightness or loudness "melodies" in the same manner that we might hear a melody using pitches from a musical scale (McDermott and Oxenham 2008). It is therefore desirable to further restrict the definition of pitch to the perceptual feature derived primarily from the common frequency attributes of a tone's partials: this form of pitch is known as **complex pitch**.[2] Complex pitch is the percept derived from the common frequency attributes of the multiple partials of a sound stimulus; it may be reckoned in terms of periodicity, as the common period (rate of repetition) of the partials, or

---

[2] Other terms for complex pitch include periodicity pitch or "low" pitch; it is not to be confused with other combinatorial pitch percepts, such as "residue" pitch or "difference tones"; although all of these percepts involve perceiving frequency patterns among multiple partials, they need not be equivalent for a given stimulus.

in terms of harmonicity, as the common fundamental or denominator of the partials' frequencies. Hereafter, the more musically common term "pitch" is understood to represent complex pitch unless otherwise specified.

4.1.1 Why do we perceive pitch?

Many environmentally common sounds, including the vowels of human speech, are harmonic complex tones: they consist of several partials that vibrate at multiples of a common F0. The ubiquity of harmonic sounds, combined with our ability to detect minute differences in pitch, makes pitch a useful percept in many listening environments; accordingly, humans demonstrate pitch-perception abilities from an early age (Trainor and Corrigall 2010). Pitch provides ecologically meaningful information about the source of complex sounds: for instance, we may use pitch to determine the size, gender, and age of a human speaker (He and Trainor 2009). As a testament to its environmental usefulness, pitch perception has also been observed in other species, including birds, fish, and a range of mammals.

Pitch may also by used by the top-down perceptual processes that segregate and stream harmonic sounds: a listener may follow the speech of one talker by following the pitch of his or her speech. As the human auditory system is able to distinguish very small differences in F0 among simultaneous sounds, pitch is particularly useful in noisy auditory environments: for instance, listeners may use pitch differences as small as a semitone to segregate simultaneous vowel sounds (Assmann and Paschall 1998; Micheyl, Hunter, and Oxenham 2010).

4.1.2 Pitch and domain

Pitch takes on additional significance in a musical domain. Nearly every musical culture uses some form of pitch scale (Nettl 2000), in which musically produced tones are perceived within a culturally defined scale of pitch distances or intervals (Peretz 2006). Accordingly, musical listening implies the categorization of pitch within the discrete elements of a musical scale—scalar pitch—such that slight discrepancies in F0 relations do not disrupt musical organization: a listener may hear a number of melodic intervals with differing F0 distances as all belonging to the category "minor third," for instance (Burns and Ward 1978).

The effect of musical scales on pitch perception also extends to perception of a tone's **chroma**—the perceptual correlate of pitch class. Most musical scales (including the scales of Western music) treat tones that are octave multiples, such as tones at 110 Hz (A2), 220 Hz (A3), and 440 Hz (A4), as having the same musical function; the scale is said to possess *octave equivalence*. In musical practices that include octave equivalence, perceiving the chroma of a tone allows the listener to assess the tone's function within the musical scale. The ability to perceive chroma as a distinct feature has led to two-dimensional helical models of pitch (Révész 1953; Shepard 1964), such that chroma and octave (or "pitch height") provide two separate dimensions for perception and categorization of musical tones.[3]

---

[3] Although there is some neurological evidence for the separate representations of chroma and octave (Warren et al. 2003), further research on chroma perception (Kadosh et al. 2008) suggests that these representations are strongly connected with the assigned listening task, such that the significance of chroma is maximized.

4.2 The mechanics of pitch perception

As with object perception, pitch perception may be modeled through a combination of acoustic (bottom-up) and cognitive (top-down) factors. Patterns in the frequency dimension of a stimulus provide the acoustic cues to pitch, while musical context and experience in perceiving pitch affect how these patterns are converted to a musical percept. The following sections outline the elements of each factor, such that they may be used to put forth a model pitch perception that attempts to predict a listener's pitch percept of a given stimulus.

4.2.1 Acoustic factors

Harmonicity is an essential component of complex pitch perception: partials that contribute to a complex pitch percept must have frequencies that form harmonic ratios.[4] These ratios need not be exact: although slight "mistunings" of harmonics may nudge the resulting pitch percept in the direction of the mistuning, they will not in most cases inhibit the perception of pitch. The ability to accommodate mistuned harmonics is ecologically useful for determining the pitch of sounds that have undergone a systematic or regular shift in frequency—such as when sounds are reflected, or when the sound source is moving with respect to the listener (the "Doppler effect"). In musical practice, systematic inharmonicity may be introduced in the physical structure of instruments: the endpoints of a string, such as the bridge and nut of a guitar, slightly alter the pe-

---

4. This is not the case with other forms of pitch discussed above, such as brightness or spectral centroid: we may hear a "wave" of noise with a rising spectral centroid as ascending in pitch, for instance. Similarly, a partial that is perceived in isolation, as a pure tone, will bear a pure-tone pitch that is not dependent upon the frequencies of surrounding sounds.

riodicity of the string passing over them, such that higher harmonics do not vibrate at perfect multiples of the fundamental, and yet our pitch percept of the sounding string remains relatively unaltered.

The *number of harmonics* also affects pitch perception, with a greater number of harmonic partials leading to a more salient pitch percept. One study of infants' pitch perception (Clarkson, Martin, and Miciek 1995) found that 7-month-old infants were able to perceive the pitch of complex tones made up of as few as three harmonics. Among musical professionals, this ability is even more refined: a recent study by Seither-Preisler et al. (2007) showed that professional musicians were largely able to discern the pitch of tones made up of only two harmonics.

The *harmonic number* of a partial (in relation to the F0) affects its ability to contribute to a pitch percept. Pitch appears to be most affected by the presence of the lowest five or six harmonics of a harmonic series; this factor is known as *spectral dominance*.[5] One significant implication of spectral dominance is that higher harmonics may deviate more greatly from the harmonic series without greatly affecting the pitch percept of the complex tone. It is nonetheless possible to perceive the pitch of a tone made up of only of higher harmonics; although such sounds are rare in real-world listening environments, the ability to perceive pitch from such sounds has interesting implications for modeling our pitch perception mechanisms (as discussed below).

The degree to which a pitch percept may be heard as representing a complex sound is known as **pitch salience** (Cariani and Delgutte 1996), alternately as pitch

---

[5] Helmholtz notes the role of spectral dominance in tones that are considered appropriate for music: "For a good musical effect we require a certain moderate degree of force in the five or six lowest partial tones, and a low degree of force in the higher partial tones." (Helmholtz 1877, chapter 19)

strength (Fastl and Zwicker 2007) or pitch weight (Terhardt 1979). The concept of pitch salience provides a means of distinguishing clear or unambiguous pitch percepts, such as those from harmonic complex tones, from pitch percepts of inharmonic sounds that are "weaker" or more ambiguous. Beyond this distinction, it is unclear if pitch salience is a representation of likeliness (the chance of hearing a pitch), or of robustness (the degree to which the pitch is represented in the auditory system). Whereas it is possible to quantitatively measure listeners' assessments of the pitch salience of various sounds, as has been done by Fastl and Zwicker (2007, chapter 5), a more recent neurological study by Barker, Plack, and Hall (2011) found that brain activity correlated with pitch perception showed no representation of salience: the authors suggest that the pitch-processing areas of the auditory cortex are responsive only to the presence or absence of pitch, and not to salience or degree. Accordingly, the analyses in this dissertation adopt this "all or none" approach to salience and define it as a function of certainty of pitch: the greater a sound's pitch salience, the more likely the sound is perceived by multiple listeners as having the same pitch.[6]

4.2.2 *Cognitive factors*

As with object-formation processes, top-down processes of pitch perception may be identified with two factors, context and experience. These factors are mutually dependent, of course: a given context may only have meaning to a listener who has learned what to find within that context. But this distinction remains useful at an empirical level,

---

[6] McLachlan (2009) provides a variant of this salience-as-certainty definition: the authors define pitch strength is defined as the certainty of a sound's pitch height within a specific octave.

as we wish to distinguish whether the cues toward pitch perception are found primarily in the sound environment (context) or in the listener (experience).

Evidence of *experience* in perceiving pitch may be observed in the "number of harmonics" studies mentioned above: trained musicians are able to hear the F0 of a two-partial tone, whereas this ability is largely absent in untrained populations. Learning is not only a factor in perceiving such low-salience tones, however: trained listeners are better at perceiving pitches in information-filled environments—that is to say, they have a higher threshold for information masking (Strait et al. 2010)—and both lifelong and short-term training has been shown to enhance listeners' neural representations of musical features such as pitch (Kraus et al. 2009).

The influence of *context* in pitch perception is often revealed in how a listener organizes pitches: for instance, the context of Western tonal music may privilege the perception of tones that are more common or significant, such as the tonic (Bigand and Tillmann 2005). More broadly, context may influence our ability to detect slight adjustments in pitch, as suggested by the research of Robert Francès discussed previously in section 3.3.

Context plays an even greater role in the pitch perception of complex tones with low harmonicity: when hearing such tones, the listener may rely upon his or her knowledge of musical behavior to apply the most likely pitch percept to the stimulus. A study by Schulte et al. (2002) used a series of tones with three harmonics each, in which the frequencies of harmonics were manipulated such that their frequencies decreased as the F0 increased—that is, the partials went down in frequency while the F0 went up. These tones were placed in a sequence in which their F0's formed the first four notes of the folk melody “Frere Jacques.” Listeners were asked to listen to the stimulus for an unusually extended period of time—one hour per day, for up to one week—for the presence of a familiar melody; the majority of listeners reported a sudden change during this listening period that enabled their perception of the melody, and they were consequently unable to return to their previous hearing.[7]

4.3 Modeling pitch perception

The perceptual factors reviewed above outline the difficulties faced by attempts to empirically model pitch perception: it is a process that is influenced by acoustic patterns, contextual information, and the listener’s skill level. That said, many of these influences are controlled within the musical environment: for instance, a single musical complex tone will present little in the way of masking information, which increases the relevance of bottom-up factors and limits the need to call upon context or learning to clarify the percept. It is therefore possible to model the pitch percepts of typical musical sounds without needing to control for cognitive processes.

As stimuli become more complex—less harmonicity, fewer partials, or more noise—top-down processes become correspondingly more involved in forming the listener’s pitch percept, and it becomes respectively more difficult to include these processes within a pitch model. This is especially true with stimuli that may be perceived either holistically or analytically, such as musical sonorities: as pitch is a feature at-

---

[7] This inability to revert to a more analytic hearing mode mimics the phenomenon of hearing sine-wave speech (Remez et al. 1994); a listener often initially hears separate streams of sine-tones, but upon “discovering” the speech content of the combined (grouped) stimulus, he or she is unable to revert to hearing separate streams.

tributed to a single perceived object, changing the listener's object-formation processes will affect the availability of pitch percepts. It follows that the most useful model of pitch perception may in fact be an incomplete model, one that attempts to capture the physiological part of pitch hearing while acknowledging the variable influences of top-down processes. Alternatively, a pitch model may be designed to focus on achieving the most likely human results, without specifically representing the combination of processes that contribute to the percept. In the following section, I will look at both kinds of models—the process-agnostic "black box" approach and the psychophysical approach—and weigh their benefits for modeling the pitch perception of chords.

4.3.1 Black box models

The simplest form of a black box pitch model is the analysis of a tone's spectrogram: we may take the common fundamental of the tone's partials as the pitch percept of the listener, with the assumption that any cognitive processes involved serve only to reinforce the perception of the observable harmonic series. As the acoustic complexity of tones is increased, this "eyeball" method of pitch analysis becomes less and less informative, and more complicated algorithms are needed to extract the acoustic attributes that contribute to pitch perception. Black box pitch models that operate on recorded sound use a variety of techniques to extract salient harmonics from the sound signal. A summary of these techniques is found in Camacho and Harris (2008), in which the methods applied in the authors' SWIPE pitch model are compared to other prominent approaches and models.

A significant limitation of black box models is the need to make assumptions about the nature of the stimulus being analyzed. As a model is optimized to work with a particular form of pitch-bearing sound, such as voiced speech or musical instrument tones, the application of this model to different kinds of sounds may produce results that are not typical of human pitch perception. This limitation is most significant in the pitch perception of multiple simultaneous tones, as in a musical sonority: a model designed to estimate a single pitch will treat the sonority as a single tone with conflicting harmonic series. Consequently, many algorithms for the musical transcription of multi-part or harmonic music have been designed to seek and remove pitches from the signal, using iterative processes to identify the multiple pitches available at a given time (Ryynänen and Klapuri 2008).[8]

In examining a chord's potential to afford pitch percepts, it may be productive to examine the performance of black box models, in which the stimulus analyzed is assumed to be a single musical object; to this extent, the pitch analyses in this chapter will occasionally use the SWIPE algorithm of Camacho and Harris to examine the possible pitch percepts of a variety of musical sounds. However, as monophonic pitch models have not been specifically designed to model the pitch perception of chords, the results of these models may not reflect the perception of a chord's pitch. It is therefore desirable to compare and possibly combine these results with a model that attempts to more closely represent the human auditory system and its perceptual processes.

---

[8] Musical transcription models that are designed to detect chords, rather than discrete pitches, need not use the iterative "seek and remove" process; one recent model by Mauch (2010) takes a Bayesian approach, comparing the overall frequency spectrum of the sound with a knowledge bank of chords, and choosing the most likely pitch. This approach is more closely in line with the chord-level mode of perception outlined in section 1.3.2.

4.3.2 Physiologically based pitch models

The auditory object dimensions of time and frequency appear to be reflected in the physical capabilities of the human hearing system. The inner ear or cochlea is capable of responding to specific frequencies of incoming stimuli; the cochlea's basilar membrane acts as a series of auditory filters that passes frequency information to the central auditory system. Similarly, temporal information is derived from the time intervals between the peaks of a sound wave; the rate of auditory nerve activity effectively reflects the periodicity of the stimulus. It is therefore possible that pitch perception may occur via two distinct methods: the spectral method—based on the spatial arrangement of frequency receptors in the inner ear—matches perceived frequency patterns with learned templates for a given F0, and the temporal method correlates time-adjacent "snapshots" of neuronal activity—known as *autocorrelation*—to derive the rate or period (the inverse of frequency) at which the greatest activity occurs (Plack et al. 2005).

The existence of two modes of pitch perception has led to two corresponding theories of pitch perception: spectral (pattern-matching) theories (Goldstein 1973; Terhardt 1974), based on the matching of perceived frequency patterns with a series of learned templates (Shamma and Klein 2000); and temporal (autocorrelation) theories (Schouten 1970; Meddis and Hewitt 1991), based on the detection and correlation of waveform periodicity. Accordingly, both spectral and temporal methods have been adapted to form pitch models (de Cheveigné 2005): spectral models use a harmonic "comb" filter to match harmonics, whereas temporal models seek the most salient periodicity across all partials.

Despite advances in the development of both types of models, it appears that neither spectral nor temporal models are able to completely account for both kinds of pitch perception methods (Hartmann 1996; Carlyon 1998); each method has shortcomings that may be exposed with certain stimulus conditions. Accordingly, while either method may be used to form a reasonably accurate model of pitch perception, it is likely that the human auditory system may derive pitch information from *both* temporal and spectral methods in combination. Indeed, this possibility was first suggested by Licklider (1951), who proposed a "duplex model" of pitch perception that makes use of both frequency and temporal information.

Continued research within this duplex model suggests that spectral and temporal methods are particularly suited to operate upon low-frequency and high-frequency partials, respectively.[9] Low-frequency partials, such as the lower harmonics of a complex tone, are said to be *resolvable* partials: as each partial is received by a separate auditory "filter" along the basilar membrane, without interference from adjacent filters. Partials that are higher in frequency are more likely to "flood" the wider bandwidths of the high-frequency auditory filters, and are therefore considered *unresolvable*. However, a flooded filter area still excites the auditory nerve associated with the filter, and the rate of nerve activity may be used in an autocorrelation approach to periodicity detection.

The recognition of two components of pitch perception has not led to the development of a corresponding algorithmic model; de Cheveigné notes in his review of pitch perception models that, while the development of such a two-component pitch model

---

[9] There is no absolute cutoff between low and high partials: various experiments have suggested that partials up to the eighth or tenth harmonic are resolvable. It is harmonic number, not frequency, that appears to determine the cutoff, since spacing of harmonic partials and bandwidth of auditory filters both tend to increase at higher frequencies (Plack et al. 2005).

would potentially bring increased accuracy, it would also increase complexity and the possibility of using "free parameters" to model pitch perception results (de Cheveigné 2005). Accordingly, the choice of a pitch model—either temporal or spectral—is a choice to gain precision with one type of stimulus at the cost of accuracy with another.

4.3.3 Virtual pitch model

As the pitch analyses in this dissertation are mostly directed toward musical sounds, which are typically rich in low-numbered harmonics, the pitch model that I adopt is a spectral model, based on the auditory system's ability to record frequency information and detect harmonic patterns. The spectral model that I use in this dissertation is the virtual pitch model of Terhardt (1974), which is particularly well-suited for examining stimuli that contain multiple harmonic complex tones, such as chords (Terhardt 1984).

Terhardt's virtual pitch model model takes as its input the frequency and intensity (in decibels of sound pressure level, or dB SPL) of a sound's partials. Each partial is assigned an audibility level in decibels above the auditory threshold, with higher audibility correlated with greater pitch weight (salience) in the partial's contribution to a pitch percept. Partials occurring at nearly or exactly the same frequencies can combine to form a more audible partial, while those that are a bit further apart may be subject to energetic masking—the mutual inhibition of perception that occurs at the basilar membrane, the cochlea's set of auditory filters. The resulting palette of partials is then matched to a series of harmonic templates, one for each **pitch candidate**: in other words, for each audible partial, the auditory system asks the question "Which harmonic of what fundamental may this be?" (Terhardt 1978). In the case where one or more partials fit a particular harmonic template (meaning that the pure tones "line up" to suggest a virtual pitch), this candidate is assigned a pitch weight based on the audibility and number of partials that match its harmonic template. The results of this model are an array of pitch weights and frequencies, with the "heaviest" frequencies representing the most likely pitch percepts for the complex sound.

Many recent pitch-perception applications (such as SWIPE) take as their input a recorded sound file, from which the application performs frequency or temporal analysis to determine pitch. In contrast, the computer implementation of Terhardt's virtual pitch model (Terhardt 2004) requires the frequency and amplitude parameters of individual partials to be entered by the user; it is not able to derive these parameters from a recorded sound. In order to compare the results of the Terhardt model with other pitch model applications, it is necessary to capture the frequencies and amplitudes of the salient partials within a sound; accordingly, I have developed a program in the MATLAB programming environment that analyzes the sound file for salient partials to be submitted to the Terhardt procedure. There are two stages to this procedure: first, it analyzes the sound's **power spectrum**—the SPL of the sound, distributed across its frequency components—and seek out peaks in the spectrum that represent partials; second, it submits the frequencies and SPLs of the detected partials to the Terhardt virtual pitch model to identify pitch candidates. The MATLAB program that is used to generate these results, as well as their accompanying figures, is discussed in appendix A of this dissertation.

4.4 Pitch analysis of musical sounds

4.4.1 *Analysis of a single harmonic complex tone*

Before proceeding to the pitch analysis of musical chords, we will examine the performance and results of the two pitch models discussed in this chapter—SWIPE and the Terhardt model—when applied to a single complex tone produced by a piano at a nominal F0 of A3 (220 Hz). Not surprisingly, the SWIPE pitch model identifies the pitch of the nominal fundamental, A3. The Terhardt model offers the same A3 pitch candidate, as shown in Figure 4.1: A3 has the largest pitch weight of any virtual pitch candidate. More importantly, the results of the Terhardt model are notable for the other, less salient candidates that appear: the octave-related candidates at A4 (440 Hz) and A2 (110 Hz), and the candidates at D3 (∼147 Hz) and D2 (∼73 Hz), a fifth and twelfth below A3, respectively. Another less salient candidate appears at the double octave, A5 (880 Hz); this candidate may be heard as a **spectral pitch**, in that it (and others like it) may be heard as a pure-tone pitch percept, without the reinforcement of a supporting harmonic series.

The relatively high pitch weights of the octave-related candidates, A4 and A2, suggests that the partials of this complex tone resemble the harmonic series of these two candidates as well. For instance, harmonics 2 (A4), 4 (A5), and 6 (E5) of the fundamental at A3 form the first three harmonics of an F0 that occurs an octave higher, at A4; similarly, harmonics 1 (A3), 2 (A4), and 3 (E4) form the even-numbered harmonics of an F0 an octave lower, at A2. This potential for multiple pitch interpretations is the basis of *tonal affinity* among harmonic complex tones (Terhardt 2000): two tones that share one or

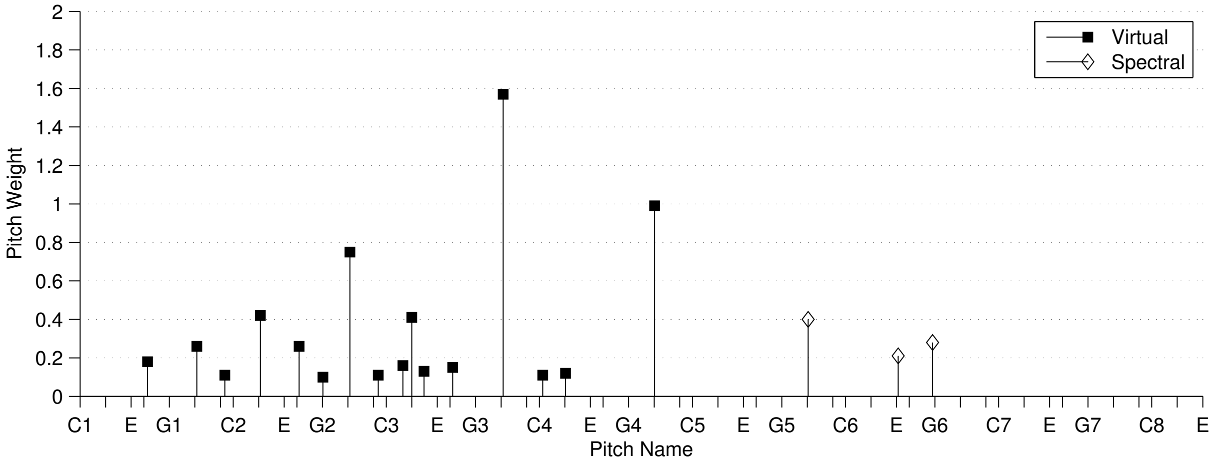

Figure 4.1: Terhardt virtual pitch analysis for piano tone at A3

more partials in their respective harmonic series may be perceived as similar, to the extent that one tone may be substituted or mistakenly perceived as the other. Tonal affinity is strongest among octave-related tones,[10] as shown in the analysis above; although this octave affinity is not equivalent with the phenomenon of octave "confusion," in which a complex tone is heard as bearing a pitch an octave higher or lower than its F0, it does serve as the acoustic basis for this phenomenon.

The presence of multiple pitch candidates in the above analysis does not, however, imply that all of these candidates are likely to be heard as pitch percepts. When hearing the piano tone analyzed above, most listeners will be able to identify the pitch as A3; previous research suggests that, if asked to adjust the frequency of a synthesized sine tone to match the pitch of the piano tone, most would adjust the sine tone in the close

---

[10] The authors of the SWIPE pitch model (Camacho and Harris 2008) note that octave errors are the most common sort of error in their pitch results. Accordingly, a modified model known as SWIPE' ("swipe-prime") uses a harmonic filter that contains only the first and prime-numbered harmonics of a potential F0; this effectively limits the affinity of octave-related tones by ignoring the harmonics (mostly even-numbered) that are shared between them.

vicinity of 220 Hz.[11] The listener who hears the piano tone analyzed above is not usually faced with a choice of pitches; he or she has learned through exposure to recognize the pattern of partials that indicate the A3 fundamental (Terhardt 1974; Shamma and Klein 2000).

Figure 4.2 contains a power spectrum analysis of the A3 piano tone examined above: the partials of the tone appear at regularly spaced intervals, 220 Hz apart, which form the harmonic series of the tone—the first six of which are labeled at the top of the graph. Within this power spectrum we may also find support for the high pitch weight of the A4 candidate identified above; the comparatively low amplitudes of harmonics 1, 3, and 5 places a relative emphasis on the even-numbered partials, which may be taken as harmonics one through three of an A4 fundamental.

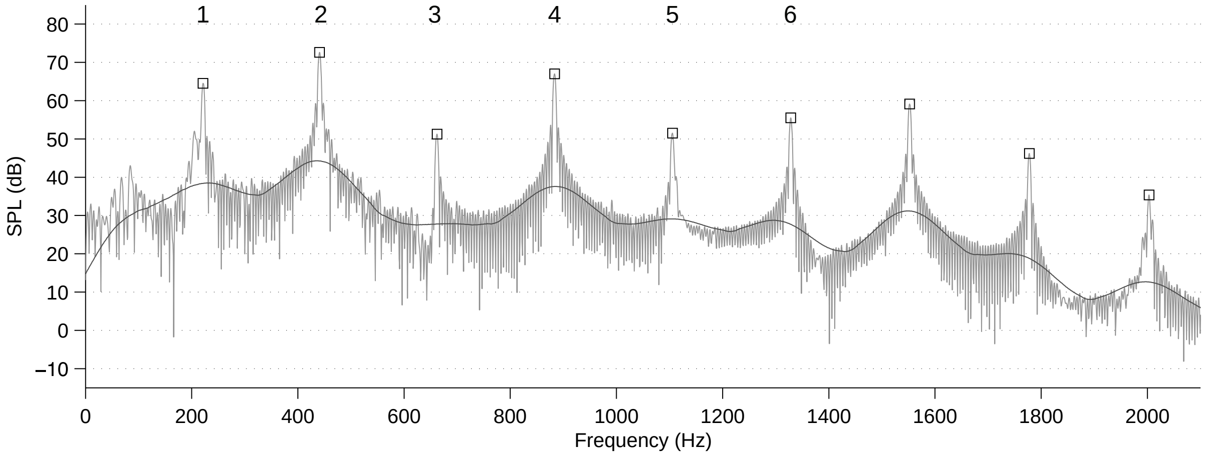

Figure 4.2: Power spectrum of the A3 piano tone analyzed in Figure 4.1.

---

[11] This sine-tone adjustment paradigm is given by William Hartmann as a working definition of pitch: "A sound can be said to have a certain pitch if it can be reliably matched by adjusting the frequency of a pure tone of arbitrary amplitude." (Hartmann 1997, chapter 12)

4.4.2 *Analysis of a single inharmonic complex tone*

Further imbalance of the harmonics' amplitudes would lead to a greater possibility of perceiving a pitch other than the fundamental as the pitch of the complex sound; this possibility is significant in perceiving pitch from an instrument known for the complexity of its sound—the carillon. A typical carillon bell is designed to produce a harmonic series that affords a virtual pitch percept at the nominal fundamental of the bell, known within the carillon community as the *prime* (Hibbert 2008). However, carillon bells also produce a partial an octave below the prime, called the *hum*, which could also be heard as the bell's fundamental. When a bell's sound is heard in isolation, the percept of either hum or prime as the bell's "strike note" (main pitch) varies with the details of the bell's construction; a listening experiment by Terhardt and Seewann (1984) revealed that about one-third of the bells heard by a group of listeners afforded multiple salient pitch candidates, such that some listeners heard the prime as the strike note, while other heard the hum.

The bell's spectrum contains additional salient partials that do not participate in the harmonic series of these pitch candidates. Partials such as the *tierce* or *quint*—at frequencies roughly a minor third and perfect fifth above the prime, respectively—contribute to the "rich" sound associated with carillon bells;[12] the salience of these partials may be so great that they may in rare instances compete with the prime or hum to be heard as the strike note of the bell (Schneider and Leman 2002).

---

12. See Fletcher and Rossing (1998) for a summary of bell design and construction, particularly with regard to the tuning of a bell's partials.

This possibility of alternative pitch percepts, as noted in the experiment by Terhardt and Seewann mentioned above, may be captured through a virtual pitch analysis of a bell's sound. Figure 4.3 shows the virtual pitch analysis of a recording of a carillon bell housed in Rockefeller Chapel at the University of Chicago; the prime of the bell is E♭3.[13] The analysis suggests that E♭3 is the most likely candidate to be heard as the bell's pitch,[14] but only marginally so: two octave-related pitch candidates appear at the octave (E♭4) and double octave (E♭5).[15] Other salient alternatives appear at the pitches of the hum (E♭2), tierce (G♭3), twelfth (B♭4), and triple octave (E♭6); the twelfth and triple octave are spectral pitch candidates, by virtue of their audibility as pure tones. The pitch candidate at G♭2, an octave below the tierce, provides an unusual alternative that corresponds to a "missing fundamental" an octave below the tierce.

A power spectrum analysis reveals that the pitch percepts observed above are enabled by the unusual collection of partials that constitute the carillon bell's sound. Figure 4.4 shows the power spectrum of the carillon bell analyzed above. In this figure, the partials labeled 1 through 4 form an approximate harmonic series for the bell's prime (fundamental), an approximate equal-tempered E♭3 (155 Hz). The relatively low amplitude of the hum (H) and quint partials, which would be harmonics 1 and 3 of a virtual

---

[13] The virtual pitch analysis and corresponding power spectrum of this carillon bell takes a time window of 0.2 seconds in length, centered at 0.8 seconds after the onset of the bell; see appendix A for a discussion of using time windows in virtual pitch analysis.

[14] A SWIPE analysis of the bell's sound identifies the prime as the pitch of the bell, although with a much lower pitch strength than was attributed to the piano tone analyzed earlier.

[15] In the analysis in Figure 4.3, many of the E♭-chroma candidates appear as tightly-grouped sets of two pitch possibilities within a semitone. These paired candidates are due to the relative inharmonicity of the E♭ harmonic series—the partials form a somewhat "stretched" series—which fools the virtual pitch algorithm into detecting two distinct, slightly offset harmonic series. We may assume that these are not really separate choices in a musical listening context, in which tones are perceived categorically within a scale, and that a listener hearing a pitch of E♭4, for instance—an octave above the prime, called the *nominal*—would not be faced with a perceptual choice between a slightly lower or higher E♭4.

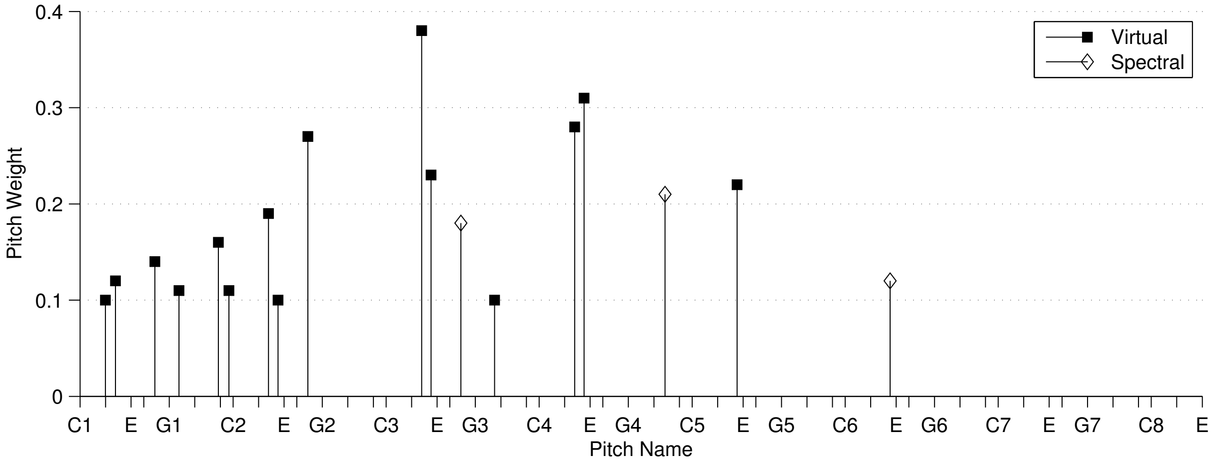

Figure 4.3: Virtual pitch analysis of a carillon bell (E♭3) in Rockefeller Chapel, University of Chicago

pitch percept at the hum, accounts for the low pitch weight assigned to this possibility. And the highly salient tierce (T) partial affords pitch percepts at the tierce and at an octave below: the tierce combines with the partial labeled 3, at approximately 460 Hz (the *superquint*, a perfect twelfth above the prime) to form the second and fifth harmonics, respectively, of a pitch at about 92 Hz, near an equal-temperament G♭2.

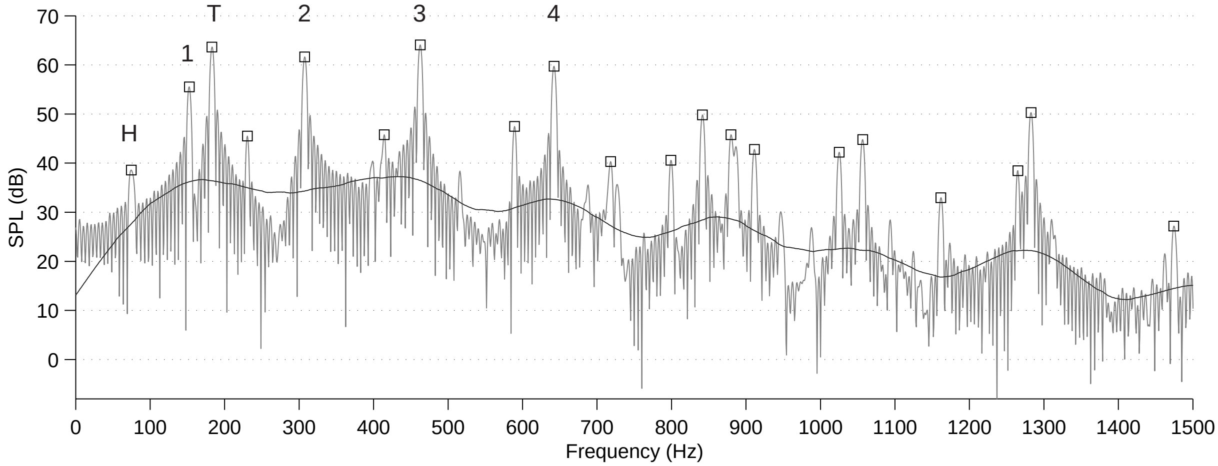

Figure 4.4: Power spectrum of the carillon bell in Figure 4.3, 0–1500 Hz

Whereas both carillon bell and piano produce single tones with multiple pitch candidates, the bell tone's candidates are more evenly weighted; as a consequence, it is much more likely for listeners to perceive different pitches from the same bell. A sound that permits multiple pitch percepts under musical listening conditions, as with the bell above, is said to possess **pitch ambiguity**: the stimulus provides mixed or weak cues to our pitch perception processes. While defining pitch ambiguity as an acoustic condition is potentially misleading—pitch is a perceptual feature, not located within the stimulus—use of the term is consistent with descriptions of similarly ambiguous stimuli in the visual domain (Leopold and Logothetis 1999).

A classic visual example of ambiguity, shown in Figure 4.5, is the so-called "Boring figure," named after psychologist Edwin Boring.[16] An observer viewing this figure may see either an old woman in three-quarters view or a young woman with her head turned, but not both at once. If he or she is given a verbal cue that favors one interpretation over the other—such as a story about an old woman, for instance—then his or her percept is likely to be skewed in favor of this verbal context. Similarly, a listener hearing the sound of a carillon bell in isolation, outside of a musical context, may identify one of the several salient pitch candidates as the "main pitch" of the bell. But when hearing the same bell sound within a melodic context, he or she may use this context to perceive the intended pitch of the bell, with a fundamental corresponding to the bell's prime. In both auditory and visual contexts, then, competing bottom-up information may be resolved by top-down processes that use contextual cues to favor one of several perceptual alternatives.

---

[16] Although experimental psychologist Edwin Boring is credited with presenting this figure, the figure itself is an adaptation of an earlier anonymous painting.

Figure 4.5: The "Boring figure"

We may also observe the influence of experience in the perceptual resolution of pitch ambiguity. Experiments using computer-generated "Shepard tones" or octave-ambiguous tones, built solely of partials belonging to a single pitch class, suggest that listeners show a tendency to identify pitches that cluster around a central region of 260 to 300 Hz, approximately C4 to D4 in equal temperament (Repp and Thompson 2009; Terhardt et al. 1986). Consequently, this frequency range is one of the most commonly used in musical tones: one survey of cross-culture melodies by Huron and Parncutt (cited in Huron 2001) revealed that the mean pitch height of a melodic tone was near D♯4, just a semitone above the region identified in the research cited above. This tendency toward a central range is captured in the virtual pitch model behind the above analysis: the E♭4 pitch candidate, which is enabled by the highly salient octave and double-octave partials within the bell's spectrum, is also given a higher weight because of its centrality within the frequency range of musical tones. Even in the absence of a contextual cue to pitch, the listener uses his or her experience with musical tones to select the pitch candidate that is most musically common.

As pitch is derived primarily from a sound's salient harmonic series, inharmonicity is strongly associated with pitch ambiguity. Yet not all inharmonic sounds will contribute to pitch ambiguity. Many sounds without a discernable harmonic series, such as white noise, afford no particular pitch percept at all. Given this consideration, the carillon bell's sound is perhaps best categorized as being in between harmonic and inharmonic—not restricted to a singular pitch percept, yet affording a a number of different pitch percepts; we may call the sound **multiharmonic**.

4.4.3 Analysis of chords

A chord may afford a pitch percept in the same manner as does a complex tone: a salient partial or harmonic series provides a pitch candidate that may be heard as representing the holistically perceived sound. But the arrangement of the constituent tones of a chord—which (if any) of its tones are replicated in other octaves, and how the tones are arranged relative to one another—plays a significant role in both the number and weights of available pitch candidates. A chord may be highly harmonic, resembling the single piano tone examined earlier; it may be highly inharmonic like white noise, affording no pitch percept; or it may be multiharmonic like a carillon bell, offering multiple harmonic series from which one or more pitch percepts may be formed.

Chords that most approximate the harmonicity of a single tone are built from tones that bear simple harmonic ratios to the fundamental; the "open-fifth" chords discussed in section 2.4.5 best exemplify this class of chords. The tones of open-fifth chords are aligned such that deviation from the fundamental's harmonic series is minimal: the combined harmonic series of the octave differs from the unison only in terms of amplitude, not frequency, whereas the fifth introduces an additional harmonic series that is partially subsumed by that of the fundamental. Figure 4.6 and Figure 4.7 show the power spectrum and virtual pitch analysis, respectively, of one of the left-hand chords from the *Pavana* introduced in Figure 2.5, which uses the notes D3, A3, and D4. The virtual pitch analysis reveals that the F0 of the chord's lowest tone, D3, affords the most salient pitch candidate for the chord, with the F0's of the remaining chord tones, at D4 and A3, having slightly lower pitch weights. Other pitch candidates arise from the salient partials of the chord tones' harmonic series: a virtual pitch candidate at A4 (440 Hz) is built form the high-amplitude partials at 440 and 880 Hz, as shown in the power spectrum, and the candidate at D2 borrows the fundamentals of each chord tone to act as harmonics 2, 3, and 4 of its own harmonic series.[17]

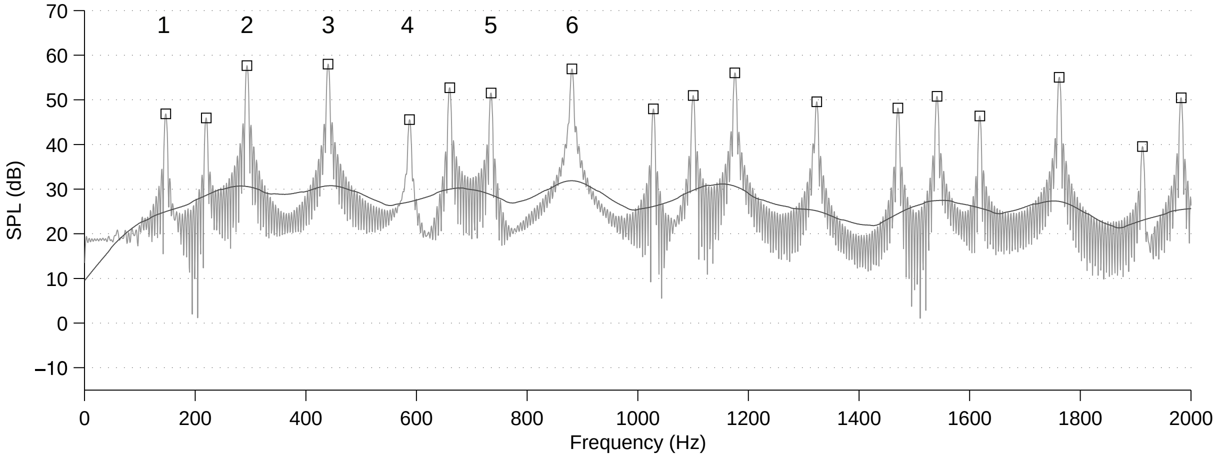

Figure 4.6: Power spectrum of chord from *Pavana*, D3–A3–D4

---

[17] A SWIPE analysis of this chord returns a fundamental of D2, taking the chord tones' F0's as harmonics two through four of this fundamental. These results suggest that SWIPE may not give as much weight to SPL differences as does the Terhardt model; we may interpret this difference as part of the "black box" model's assumption that analyzed sounds are single tones—not auditory chimerae—as noted in section 4.3.1.

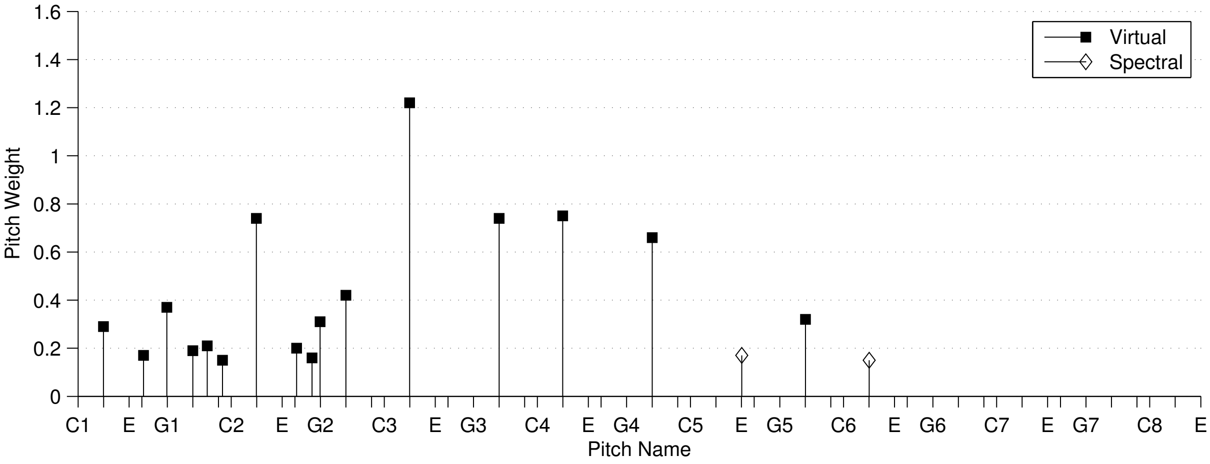

Figure 4.7: Virtual pitch analysis of chord from *Pavana*, D3–A3–D4

If we take the lowest tone's F0 (D3) as the nominal F0 of the chord, we find that the virtual pitch analysis of the chord shares some characteristics with that of the single piano tone analyzed earlier. Both graphs show a highly salient candidate at the nominal fundamental, with weaker alternatives an octave above, and weaker still an octave below. Within this single chroma, the difference between chord and single tone is largely one of degree: the pitch weight differences are greater within the single tone. Additionally, the fifth of the chord imbues the chord with additional pitch candidates at A3 and A4. Although these candidates are significantly less likely to promote an A chroma percept, their presence is registered as part of the timbre or spectral "remainder" of the chord's sound—much as the tierce partial contributes to the overall sound of a carillon bell.

Adding a major third above the open-fifth chord—creating a "major triad in root position", as is done in measure 8 of the *Pavana*—does little to alter the chord's pitch candidates, as shown in Figure 4.8. The pitch weight of the D3 candidate is lowered, reflecting the increased inharmonicity of adding a F♯ harmonic series to the mix, and

D2 becomes a stronger candidate, as it is supported by salient harmonics two through five—the fundamentals of the chord's component tones. Of course, this does not mean that the tone at F♯4 is itself only barely perceivable: as the F♯4 is part of a right-hand melody, it is likely that the listener hearing this piece will perceive this tone as part of a melodic stream, separate from the accompanying left-hand chords. But the listener may also relax his or her attention at this cadential measure, and form a holistic percept that encompasses all four tones; in this listening mode, he or she is likely to take one of the chord's salient pitch candidates as the pitch of the chord.

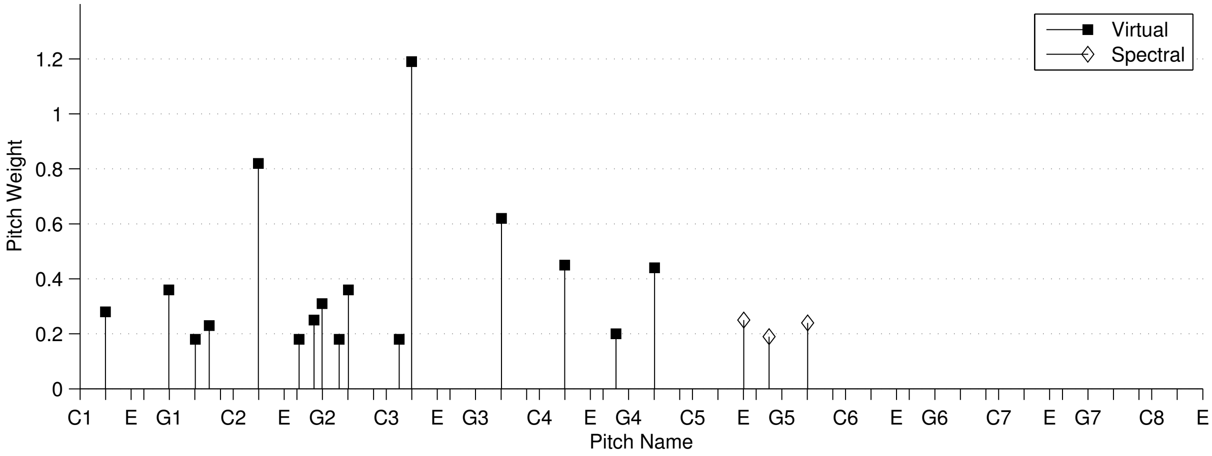

Figure 4.8: Virtual pitch analysis of chord, D3–A3–D4–F♯4

Were we to remove the bass tone at D3 from this major triad, we would create the chord analyzed in Figure 4.9; this chord can be characterized as a "major triad in second inversion" or a "six-four chord." The removal of D3 affects the pitch weight of the fundamental at the newly-anointed bass note A3: without any partials below it to create upward energetic masking, the A3 candidate is now more salient than any of the D-chroma candidates. Yet the moderate difference in pitch weights between A3 and a cluster of secondary candidates—at D2, A2, D3, D4, and F♯4—suggests that the chord bears some amount of pitch ambiguity; the pitch percept of this chord will be somewhat dependent upon both the context in which it is used and the listener's experience in hearing chords with simular patterns of partials.

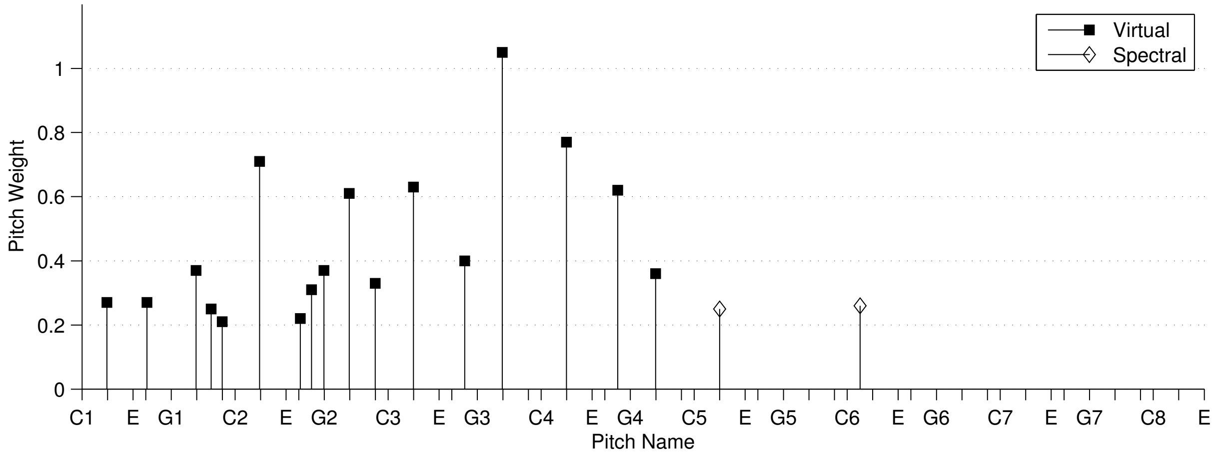

Figure 4.9: Virtual pitch analysis of chord, A3–D4–F♯4

Significantly, both context and learning are often invoked in discussing this particular permutation of the major triad. If we take the listener's chordal pitch percept to correlate to the music-theoretical concept of chord root—a potentially hazardous move in terms of theoretical foundations, despite its psychological plausibility (Thomson 1993)—then we find a similar emphasis on context and learning in the realm of scale-step or *Stufen* harmonic theory taught in the classrooms of North America. If this chord is placed in a musical context that suggests a D-major tonality, the listener may use this context—to the extent that he or she has learned to hear it as informative—to direct his or her pitch percept toward either an A-chroma candidate (likely A3) or a D-chroma candidate (such as D3). Within the common-practice period of tonal harmony, hearing A as the chord pitch implies that the chord is perceived as a musical dissonance or "suspension" on a dominant (V) harmony; hearing a D chord pitch implies a musical consonance or tonic (I) six-four chord. Although it is unlikely that this perceptual distinction is the only significant factor in the lengthy historical debate of the six-four chord's musical function (Beach 1967), we may take this music-theoretical distinction, and its perceptual correlate, as a point of entry into the study of chord perception and categorization—a topic that is more fully addressed in the following chapters.

4.5 Chord pitch and musical listening

As Rameau's *son fondamental* is modeled in the *Traité* after the undivided string of the monochord, and not on the acoustic principles that would inform his later writing, it is not surprising that Rameau describes hearing the fundamental sound as an act of imagination, of hearing a sound that is not present. But it may also be argued that Rameau is describing a sensation of virtual pitch (Duchez 1986), in which a pitch percept is formed from a harmonic series within the holistically perceived chord. In particular, Rameau's description of the capacity of "the ear" to hear this *sous-entendu* fundamental sound appears similar to the learning hypothesis put forth by virtual pitch theory (Terhardt 1974), and this passage points more generally to the role of exposure and context in learning to perceive the pitch of complex sounds.[18]

Although I present the pitch analyses above in support of the perceptual phenomenon of chord pitch, I have not yet established a basis for a *musical* hearing of chord pitch. Despite the possibility of a chord bearing an emergent pitch feature, only certain combinations of perceptual cues—combinations of acoustics, context, and listener

---

[18] Cohen (2001a) notes that Rameau himself would not have regarded this ability to hear the fundamental sound as being "learned"; this ability is product of the passively-trained ear (*l'oreille*), and not the rational mind (*l'esprit*) that understands the music-theoretical logic behind the phenomenon.

experience—are likely to promote the perception of this feature. A consideration of the musical use of chord pitch, therefore, must take into account the conditions under which it may or may not be applicable to musical listening; the three conditions outlined below encompass the listening approaches most common to hearing sonorities in the context of a harmonic tonality.

First, as pitch is a feature of perceived objects, the perceptual form taken by a chord—as single object or as simultaneity of tones—will affect the availability of pitch.[19] As explored in chapter 2, holistic perception of a chord may be thwarted by melodic streaming of one or more chord tones; the use of "passing tones" in a salient outer voice, found in many Western musical practices, permits the more active voice to be segregated from the chord. Similarly, sonorities with one or more tones separated by large distances in frequency, such as the separation of bass and upper voices common to keyboard chords, may encourage the segregation of a separate note or stream.

If a listener hears a chord analytically, and is asked to identify or match "the pitch" of a chord, he or she may instead perceive the pitch correlated with the F0 of one of the chord's more salient or more easily segregated members. Because sonorities within a musical setting often provide cues for both holistic (chord-level) and analytic (tone-level) modes of listening, efforts to promote a holistic mode of perception must either limit the availability of segregation cues, as outlined in section 2.4.4, or provide task-based incentives toward hearing holistically, as with the experiment by Schulte and his colleages discussed in section 4.2.2 above. It is therefore difficult to assess the pitch po-

---

[19] A similar situation exists with the perception of dissonance, as pointed out by Wright and Bregman (1987): simultaneous tones that are heard within separate streams are less likely to be experienced as dissonant than tones that perceptually fuse to form a single object.

tential of a chord taken out of musical context; as the virtual pitch analyses above assume holistic perception, this assumption must be borne out in the musical environment in which it is heard.

Second, not all chords will provide a single salient harmonic series from which pitch may be derived. Only the chords whose partials most resemble a single harmonic series, such as the open-fifth chords discussed above and in chapter 2, will afford a pitch percept with the pitch salience or strength of a typical harmonic complex tone. More complex chords, such as the "major six-four" chord analyzed above, are more likely to offer multiple pitch candidates with roughly equal pitch weights, though registral separation of bass and upper voices may promote the pitch of the bass tone more strongly. The listener presented with a chord's multiple pitch candidates may be able to select one of the more salient candidates as the chord's pitch, though this ability typically requires musical training on the part of the listener.

Just as hearing a fundamental from a limited number of harmonics is challenging to listeners without musical training (Seither-Preisler et al. 2007), hearing a pitch from a chord's multiharmonic array of partials is likely to be at least as challenging, if not more. Indeed, not all musicians are equally well-equipped to perceive holistic pitch: some tend to hear analytically—focusing on one or more partials within an aggregate sound—according to the performance and listening tasks that are required of the performer (Schneider and Wengenroth 2009). The ability to detect an out-of-tune tone within a chord, for instance, relies on the ability to segregate slight deviations in frequency; perception of a chord's fundamental pitch is of no use in this task. Similarly, it is possible that listeners without musical training may not derive any pitch percept from a complex chord. A musical sound may be heard as having a particular timbre or sound, but without a sensation of complex or periodicity pitch: this is our common experience of inharmonic percussion sounds in a musical environment, and this mode of listening may be extended to chords that present multiple salient harmonic series.

Third, the role of experience makes it possible to develop a hearing for chord pitches that differs from hearing complex tones. If we consider the chord as simply "a sound," outside of musical context, then hearing chord pitch is similar to hearing the pitch of any other sound within our environment: we detect a salient harmonic series and perceive its fundamental as the pitch of the sound. Yet we may also consider that a listener's experience hearing multiharmonic sounds could lead to his or her development of a different pitch "template"—conditioned not only by the sound's salient patterns, but by the way in which these sounds are organized in musical practice. The listener who has experience with the sounds of carillon bells, for instance, learns the musical roles of the hum and prime partials; he or she learns to hear the prime as bearing "the pitch" of the bell, despite the availability of a pitch percept at the fundamental an octave below—the hum partial.

Similarly, we may consider that the harmonic template underlying virtual pitch perception need not be the only conceivable manner of hearing chord pitch; it is possible to learn to hear chord pitch as a separate procedure from hearing the pitch of a single tone. This possibility allows us to propose a perceptual basis for alternate theories of harmonic structure. Theoretical systems that are not grounded in the concept of an underlying fundamental sound, such as the polar-opposite orientation of major and minor triadic chords in "harmonic dualism" theories (Vogel 1993; Harrison 1994), could be applied to train the listener's pitch processes, such that his or her pitch percept would be based on the partials or harmonic series that correspond with this theoretical approach.[20]

4.6 Conclusion

As outlined in section 4.4.3 above, many chords typical to Western music afford a high degree of pitch ambiguity. Only those chords whose arrangement of component tones most resemble the harmonic series of a single harmonic complex tone—such as the open-fifth chords analyzed above—will produce an unambiguous pitch percept. The musician wishing to elicit a specific chord pitch percept, therefore, must rely upon both acoustic and contextual cues to influence the listener's hearing of the chord as bearing a given pitch; ideally, the combined effects of acoustic and contextual cues would be so great that it may shape the percepts of listeners with little or no experience in hearing a chord as bearing pitch.

The acoustic factors leading to chord pitch are located within the chord's frequency spectrum. Although the spectrum of a chord may be altered by shaping the timbres of one or more component tones, a more common and pragmatic approach is the arrangement of chord tones to either promote or obscure the presence of a harmonic series. As suggested by the chord analyses presented above, change in chord structure—such as doubling, inversion, register and spacing—may significantly alter the chord's pitch candidates in both weight and number. With experience, the musician who is sensitive to

---

[20] Schneider (2010) briefly discusses developing such a "dualist" hearing approach over the course of a seminar led by Martin Vogel. "As with many things," notes Schneider, "this is a matter of training and experience."

holistic chord pitch may recognize the particular permutations of sonorities that produce an ideal amount of pitch ambiguity, somewhere between between the high harmonicity of open-fifth chords and the aperiodicity of “rootless” sonorities, such as a cluster of minor seconds.

Placing a pitch-ambiguous chord within a controlled musical context may also provide cues toward a desired pitch candidate. The increased role of cognitive factors in the presence of ambiguous bottom-up cues affords the musician a potentially powerful tool for shaping the listener’s percept, such that a carefully crafted sonority—one that is constructed to emphasize holistic pitch perception—may become in itself a musical object that merits extended exploration on the part of the musician and the listener. The following chapter suggests how the manipulation of a sonority’s acoustic factors to promote an emergent pitch percept may be augmented by the skilled use of musical context, such that the chord’s pitch percept is made both salient and informative in the course of a musical passage.

CHAPTER 5

CHORD PITCH IN MUSICAL CONTEXT

In the previous chapter I established the concept of chord pitch as an emergent feature of a chord: the chord, like the complex tone, may contain one or more salient harmonic series that promote the perception of a pitch. Although the virtual pitch model introduced in section 4.3.3 provides a means of examining a chord's pitch candidates, the model alone is not sufficient to describe the pitch perception process: it measures the acoustic cues within the chord, while leaving the contextual and experiential cues to be determined by the analyst.

Accordingly, this chapter shows how a musical analysis may be used to account for the cognitive factors that can promote and influence the pitch percept of a chord. The chord chosen for this analysis is one of the least well-understood chords in the Western art music repertory: it is the striking sonority that opens Igor Stravinsky's *Symphony of Psalms*. My analysis of the "Psalms chord" is built upon the premise that chord pitch may be used as a compositional resource: Stravinsky's manipulation of the acoustic and contextual factors of the Psalms chord leads the listener to experience its pitch as a salient musical entity. Following this analysis I propose that, whereas chord pitch may constitute a familiar and readily analyzable emergent feature of the chord, a full account of chord perception requires us to propose and examine other emergent features that may shape our experience of hearing a chord.

5.1 Prerequisites for a chord pitch analysis

In order for a chord pitch analysis to be musically informative, such that it describes a listening experience that is both plausible and worthy of investigation, a number of prerequisites must be met. First, the listener must be likely to hear the sonority in question as a chord. The acoustic composition of a sonority provides domain-general cues toward holistic perception, as outlined in chapter 2; although a glance at the sonority's notation will tell us if the sonority is intended to be onset-synchronous, a more thorough assessment of acoustic cues requires an understanding of "secondary" features such as instrumental timbre and dynamics.

Second, the chord must afford at least one pitch candidate. Although most musical chords will meet this requirement, it is possible to compose chords that use energetic masking to obscure the component tones' harmonic series. Complex tones that form small musical intervals, such as major and minor seconds, may mutually cancel out their harmonic series. The mechanism behind this phenomenon is concerned with **critical bands** within our hearing systems: two or more pure tones within a critical band are unable to be resolved (have their frequencies perceived) by the basilar membrane, the organ responsible for transforming perceived partials into frequency content.[1] While we may use the virtual pitch model to examine a chord for its pitch-producing potential, we may initially make a rough initial assessment of the masking effects of critical bands by looking at the chord's notation: the use of close intervals, particularly within the

---

[1] Huron (2001) provides a more extensive summary of critical band research within general and musical domains.

lower register—D3 and below provides a good rule of thumb—is likely to lead to mutual energetic masking.

Last, the listener must have a musical reason to perceive the pitch of the chord in question. We must be cautious in assuming that that the listener has sufficient experience in perceiving chord pitch, other than perhaps hearing the fundamental or "root" of a typical major- or minor-triadic chord. If chord pitch is to be used as a compositional resource, the chord bearing this pitch should be placed in a context that draws the listener's attentional focus: it should be foregrounded, such that the listener may perceive features of the chord (such as pitch) in greater detail than if he or she heard the chord as accompaniment or background material.

It is in this last criterion—the construction and deployment of attention-grabbing sonorities—that the music of Stravinsky is revealed as an ideal subject for chord pitch analysis. Whereas Stravinsky's early works, and his Russian ballets in particular, are characterized by his occasional use of unconventional chords, his compositional attention to chords is not simply restricted to the dissonant and metrically jarring sonorities found in works like *Rite of Spring*. Even the most apparently simple and common chords could take on an air of significance and uniqueness in Stravinsky's descriptions, as can be seen in this recollection from Walter Piston:

> In the course of a talk about *Oedipus Rex*, […] an observation that he [Stravinsky] made threw a bright light on a most important aspect of his artistic ideals. He said, "How happy I was when I discovered that chord!" Some of us were puzzled, because the chord, known in common harmonic terms as a D-major triad, appeared neither new nor complex. But it became evident that Strawinsky regarded every chord as an individual sonority, having many attributes above and beyond the tones selected from a scale or altered this way and that. The particular and marvelous combination of tones in question

> owed its unique character to the exact distribution of the tones in relation to the spaces between them, to the exact placing of the instrumental voices in reference to the special sound of a given note on a given instrument, to the dynamic level indicated, and to the precise moment of sounding of the chord. (Piston 1947, 256–257)

We may share Piston's puzzlement concerning Stravinsky's attention to this seemingly ordinary chord, shown in Figure 5.1[2]—that the creator of such controversial chords as the "Petrushka chord" and the "Augurs chord" could lavish such praise on a D-major triad, no matter how carefully it were orchestrated. And yet to regard the chord as simply a D-major triad is to ignore the "attributes above and beyond the tones"—the elements of registration, instrumentation, timing and dynamics—that Piston identifies as significant to Stravinsky's compositional technique. In the analysis that follows, I claim that one of these emergent attributes is chord pitch, and that Stravinsky's attention to both the acoustic details of the chord and its function within a musical context lead the listener to perceive the Psalms chord's pitch as an essential component of the passage in which it appears.

5.2 Analysis: Igor Stravinsky's *Symphony of Psalms*, first movement

5.2.1 Introducing the chord

Figure 5.2 shows a reduction of the opening measures of the *Symphony of Psalms*, in which the Psalms chord appears a total of four times. From a music-theoretical standpoint, the Psalms chord is a root-position E-minor triad, though its unusual distribution of pitches—notably the octave Gs in the middle—is often cited as the cause of its unusual

---

[2] All scores are shown at concert pitch.

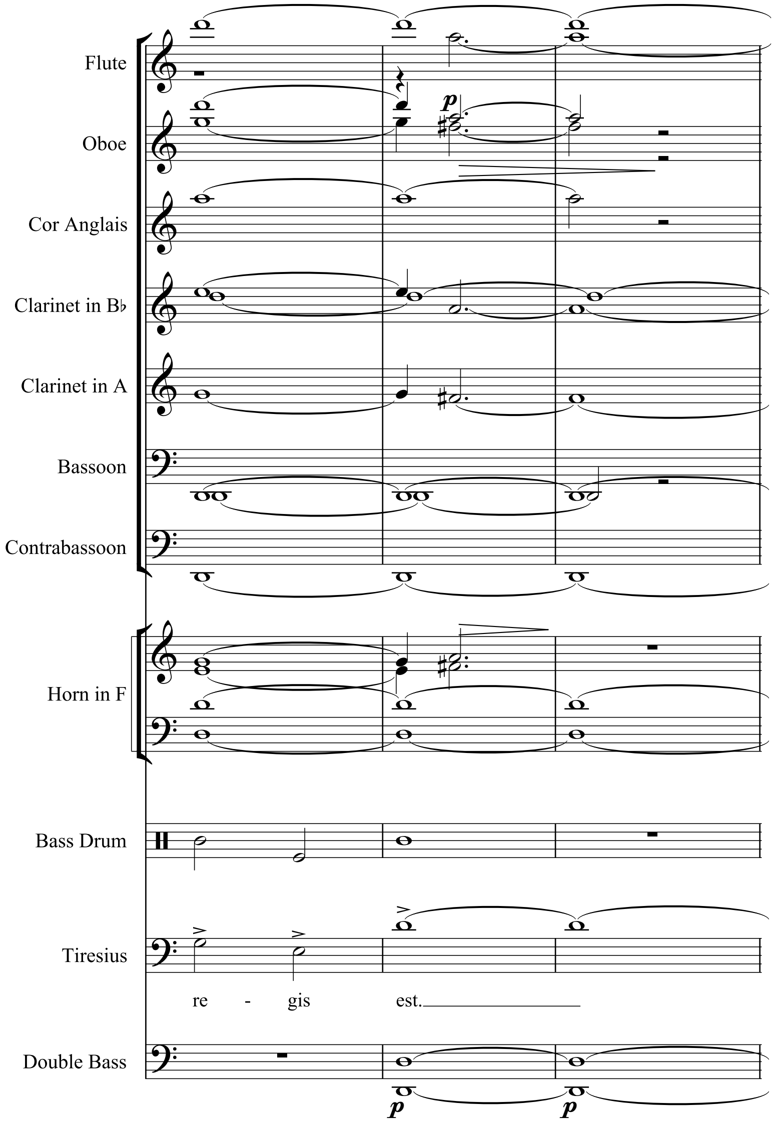

Figure 5.1: "That chord!" from Stravinsky, *Oedipus Rex*, Act III, rehearsal number 82

and memorable sound; Joseph Straus notes that this chord "is fraught with a musical tension not normally associated with a simple triad" (Straus 1982, 268). Indeed, Stravinsky subjects this "simple triad" to such unusual acoustic and contextual conditions that it bears little resemblance to the foundational harmony from our textbooks.

Although the Psalms chord contains two significant acoustic factors—onset synchrony and brief duration—that promote its perception as a chord, Stravinsky's orches-
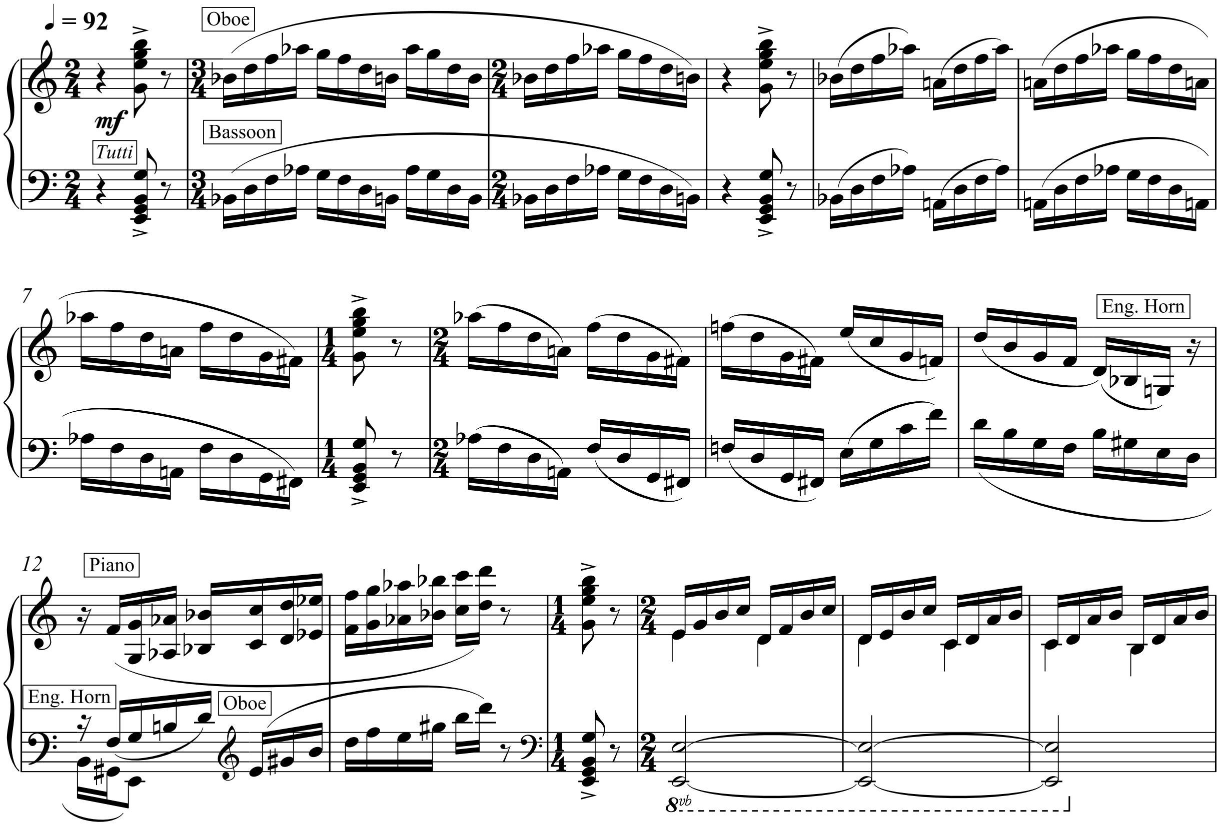
Figure 5.2: Igor Stravinsky, *Symphony of Psalms*, first movement, mm. 1–17

tration of the chord (shown in Figure 5.3) provides other acoustic cues that strengthen this initial binding. Stravinsky's typically wind- and brass-heavy orchestration affords a consistent timbre among many of the chord's tones, meaning that a single tone is unlikely to "pop out" to the listener based on its timbre alone. And the use of sharp, percussive onsets in all instruments creates a unified attack (onset) profile, much like that of a sharply struck piano chord. This combination of onset synchrony, onset quality, duration, and timbral similarity obscures all but the most subtle cues to the perception of the individual tones that make up this chord. Although the listener who is familiar with orchestral music knows that this chord is composed of many individual instrumental sounds, he or she is unable to act upon this intuition during the brief time while the chord is sounding.

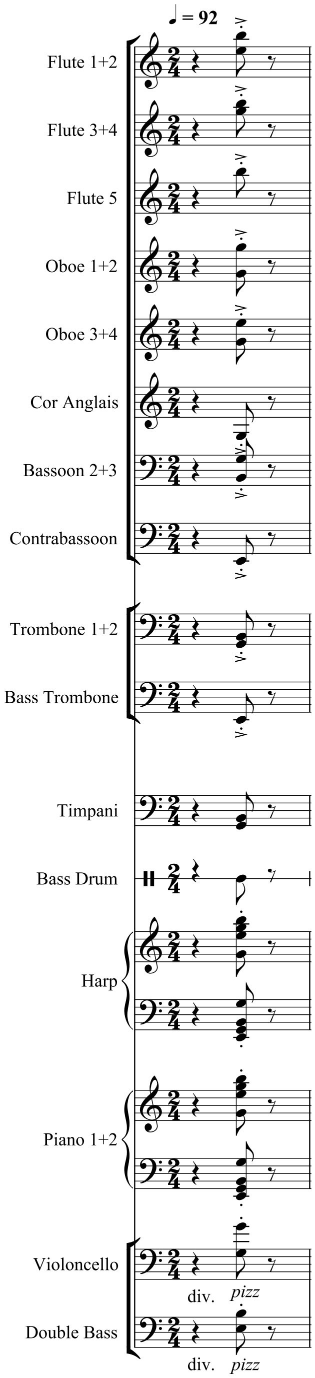

Figure 5.3: Initial occurrence of the Psalms chord, orchestral score

The Psalms chord's rather impermeable acoustic front is reinforced by Stravinsky's use of musical rests to physically isolate the chord from its musical context: of the four instances of this chord in the introductory passage, three of them are preceded and followed by rests.[3] By presenting the chord in isolation, Stravinsky reduces the effectiveness of any contextual cues the listener might bring to hearing the chord: the melodic passages in between instances of the chord are less able to capture one of the chord tones within a melodic stream. With the acoustic and contextual fortification that Stravinsky applies to the Psalms chord, we may reasonably assume that the "Psalms chord" is indeed heard as a chord, rather than a collection of tones.

5.2.2 *Virtual pitch analysis of the chord*

Having established the Psalms chord as likely to be perceived holistically, we may now apply a virtual pitch analysis and determine the salient pitch candidates for the chord. Figure 5.4 shows the results of a virtual pitch analysis of the Psalms chord; using Sibelius 6 and the Garritan Personal Orchestra Lite package to generate the sound file for analysis.[4] As noted in section 5.2.1, the presence of the octave G's (specifically, G3 and G4) in the middle of the chord has been identified as one of the factors in the chord's unusual sound. We might expect that the relative frequency isolation of the G's would

---

[3] The exception to this isolation strategy occurs in bars 7 and 8, which removes the rest between melody and chord; in this instance, Stravinsky trades isolation for ambiguity, as will be further discussed in section 5.2.4 below.

[4] Although analysis of sound files taken from recordings of one or more performances of *Symphony of Psalms* would be useful in gauging virtual pitch percepts of hearing the chord "live," the typically reverberant nature of recordings makes virtual pitch analysis particularly difficult; using a sound file generated from samples permits a "cleaner" sound that improves the signal-to-noise ratio of the partials of the Psalms chord.

promote one or both of these pitches as the most salient candidates of the chord. As it turns out, this intuition is correct: as shown in Figure 5.4, the G3 candidate is the most salient. However, the pitch weight of this G3 candidate is not solely attributable to its frequency isolation. The triad of tones that lie in the lowest register of the chord (E2, G2, and B2) are effectively positioned too close together to be perceived as separate pitch candidates: as they all fall within the same critical band, they are subject to the mutual interference of energetic masking. As a result of this energetic masking in the bass register, the G3 provides the lowest salient pitch candidate of the chord.

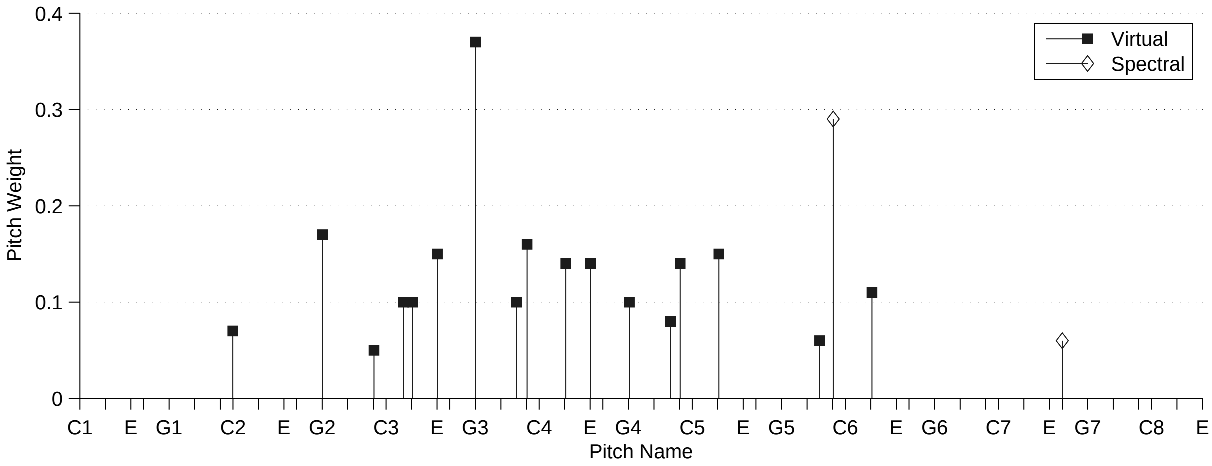

Figure 5.4: Virtual pitch analysis of the Psalms chord

The availability of a B5 spectral pitch candidate, only slightly less salient than the G3 candidate, may be attributed to both orchestration and registral position: three of the five flutes performing the Psalms chord play a B5 tone, and its status as the highest tone of the chord allows it to be more easily identified and segregated. We may therefore suppose that a listener with experience focusing on the highest pitch of a sonority as the most musically informative (Madsen and Widmer 2006; Duane 2010) may hear the integral chord as having B5 as its "main pitch." The listener who takes B5 as the chord's pitch percept, however, is left out of the listening journey in which Stravinsky places the chord: he or she is unable to follow the chord's contextual metamorphosis, as I describe in section 5.2.4 below.

5.2.3 Chord pitch and chord quality

As much as the G3 pitch candidate is a prominent feature of the chord, its lack of a salient E—particularly in the bass register—is even more significant in shaping the percept of this chord. Although the Psalms chord is nominally in root position, Stravinsky suppresses any acoustic quality of "rootedness" by flooding the lowest register of this chord, masking the bass E with interfering G and B tones that all fall within the same critical band. And it is this rather unusual feature—a root-position triad that doesn't sound like one—that Milton Babbitt identifies as a key to its musical utility within this passage:

> It is not merely what would be termed conventionally a root position, E-minor triad, but is so uniquely a specific representation of such a triad that it is possible for Stravinsky, throughout the entire movement, until the final sound, to use no root-position triads other than E-minor triads. (Babbitt 1964, 38)

Babbitt's comments suggest that, since the root-position triad in Western music is commonly associated with rest or repose, the repeated use of a saliently root-position chord would effectively remove the sense of "musical tension" that Stravinsky is apparently interested in. But while the E2 tone may not contribute a salient pitch percept, its presence is still manifest in the upper harmonics it contributes to the frequency spectrum of the chord. With this in mind, we must allow for the listener to perceive the chord as being some kind of E-minor triad, despite the lack of a salient E pitch percept. And although this **chord quality** percept is not captured in a virtual pitch analysis, I claim that it constitutes a separate emergent feature of the chord. As we lack an equivalent perceptual model of chord quality—an issue that I take up further in chapter 6—for the purposes of this analysis I propose that the combination of harmonics comprising the chord provide an acoustic cue to the chord's E-minor quality.

In this regard, the Psalms chord may be described as bearing an internal conflict—a salient G pitch percept versus a latent E-minor chord quality. This conflict places even more significance on the chord's musical context; Stravinsky's melodic interludes may be understood not just as motivic material, but as a means of guiding the listener toward or away from a particular interpretation of the Psalms chord. If an E pitch is suggested too strongly by the surrounding melodic material, the listener may perceive a E chord pitch and consequently hear the chord as authentically root-position, signaling a stopping point in the passage. Stravinsky's response to this challenge is a gradual shift in melodic material that subtly prepares the listener to hear the E-minor quality of this chord, in anticipation of the E tonality that begins in measure 15.

5.2.4 The Psalms chord in context

The initial occurrence of the Psalms chord is heard outside of any particular musical context: it is the first sound that the listener hears in this piece. It is therefore likely that the listener may adopt a more domain-general approach to its perception. As the listener has no pitch or tonality reference with which to organize the chord, he or she may hear it as a complex sound—an auditory chimera—bearing the domain-general qualities of loudness ("quite loud") and duration ("quite short"). In addition, the orchestration of the chord gives it a particularly distinctive timbre that may be perceived as an emergent feature of the complex sound. In subsequent occurrences of the chord, in which the chord is situated within a melodic context, the listener may begin to appraise the chord's musical features (pitch and chord quality) while retaining a memory of the chord's timbre.

As shown in Figure 5.2, the arpeggiated melody of measures two and three, cycling through B♭7 and G7 pitch collections, offers little clue to the chord's true nature as an E-minor triad; if we may posit any contextual inclination toward a pitch class in this passage, it is toward the G, as the A♭–G descending half-step in the middle of both measures provides both registral and motivic emphasis for the pitch class G. While bars five through seven continue the previous measures' arpeggiated motive, including the A♭–G motion in measure six, the end of measure seven contains new material that introduces another melodic half-step, G–F♯. Even more significantly, within this passage Stravinsky does away with the rest or silence that precedes every other occurrence of the Psalms chord: the descending G–F♯ melodic motion may permit the listener to hear its continuation within the chord—but to what concluding pitch? Is this a diatonic descent to an implied E, or a neighboring figure to the highly salient G? Compared to the G-centered content of the previous melodic passage, this two-note figure provides a more ambiguous listening scenario, in which both G and E become possible points of contextual emphasis.

The passage following Psalms chord number three, spanning measures 9 through 13, presents more support for E as a perceptual focus: the melody at last introduces the pitch class E in the latter half of bar 10, and arpeggios on an E-dominant harmony appear in the bassoon and oboe in measures 11 through 13. Since Stravinsky still wishes to avoid the stasis of a root-position chord at this point, the influence of these melodic maneuvers must be subtle—the arpeggiated E-dominant figures are counterbalanced by the harmonically unrelated material in the rising piano figure. More significantly, Stravinsky's melodic material has at last moved away from promoting G as a pitch reference, and the E pedal in measure 15 signals the end of the perceptual journey from G to E.

The observations I have made above suggest that Stravinsky was aware of an emergent G pitch percept associated with the Psalms chord—that virtual pitch was one of the "attributes above and beyond the tones" with which Stravinsky was concerned. Although Stravinsky may not have conceived of the chord's G percept as "the pitch" of the chord, he does use the concept of *polarity* to refer to a form of pitch hierarchy within his compositions: he is concerned less with "tonality, properly so called, than what might be described as the polarity of a sound, of an interval, or even of a sonic complex [*complexe sonore*]" (Stravinsky 1942/1947, 153). We may therefore suggest that the Psalms chord bears a specific polarity, pairing the acoustic emphasis of the inner-octave G's with the chord's tonal implication of an E fundamental; the contextual transition from G to E is an expedition from one pole to the other.[5]

---

[5] Several scholars have offered interpretations of Stravinsky's use of the term polarity—a term that appears in writing only in his *Poetics of Music*; for a summary of this research, see Palmer (1998, chapter 1).

5.3 Conclusion

The above interpretation of the introduction to the *Symphony of Psalms* is built upon an appealing musical use of chord pitch: the chord's virtual pitch profile, with its emphasis on the G chroma, obscures its tonal function as an E minor chord, and Stravinsky uses the material in between chord instances to guide the listener's attention toward the chord's E-minor quality. When the Psalms chord returns later in the movement, once each before rehearsal numbers 9 and 10, it is no longer perceived as ambiguous; this game is effectively ended by the firmly established E tonality that follows the introductory passage. In these instances, the chord may be heard as a complex sound, drawing the listener back to the opening measures of the piece. And in this manner, Stravinsky extracts maximum musical utility from the humble minor triad: the Psalms chord acts as a bearer of a G pitch percept in its initial instances, as a tonic triad within the E-minor context that follows the introductory passage, and a signature complex sound throughout. Each attribute of the chord serves a different structural purpose within the movement.

The analysis above proposes that the percept of chord pitch may be used as a musical device: we may use a chord's pitch within the organizational means we apply to the more common phenomenon of scalar or tone pitch. In this regard, perceiving the pitch of the Psalms chord is much like perceiving the pitch of a multiharmonic complex tone, such as the carillon bell (as discussed in section 4.4.2): both bell and chord afford sufficient acoustic cues to be regarded as integral sounds that afford multiple pitch candidates, and the listener draws upon his or her experience, and the musical context in which the sound occurs, to perceive a pitch.

CHAPTER 6

CHORD PITCH AND CHORD QUALITY

In the analysis of the Psalms chord in chapter 5, I claimed that the listener could experience a contextually guided transition in pitch percept in the opening measures of Stravinsky's *Symphony of Psalms*, such that the melodic context leads the listener from a G pitch percept to E. As an aide to this transition toward E, I posited a second musical feature of the chord—chord quality—as potentially assisting this guided listening: the listener may perceive a E-minor quality by virtue of patterns within the chord's frequency spectrum. Unlike chord pitch, this proposed feature of chord quality offers no analogous feature within the realm of the single musical tone: to account for the "minor" part of "E minor," we must propose a musical feature that is apparently unique to chords. Yet given the foundational role of chords within harmonic tonality, it is perhaps surprising that the perception of chord quality is relatively unexplored in empirical research. In this concluding chapter I consider why this percept is not better understood, and I take the chord quality feature proposed in the Psalms chord analysis as a starting point for future research on chord perception.

6.1 Tone and chord

In chapter 1, I introduced the terms "sound" and "tone," as proposed by Scruton (1997), as an essential part of musical listening: the listener forms a musical object by perceiving the sound as a tone. To this operation we may attribute an equivalent transformation to the performer of the sound: he or she conceives a "note" and produces a sound. In many musical listening situations, this transition from note to sound to tone operates without raising awareness of itself, so that neither performer nor listener is aware of any transition at all: the performer thinks of a note and produces a sound, which is heard as a tone. And yet this distinction is valuable, in that it allows us to account for the perceptual experiences of tone that are not readily apparent when examining the sound: two sounds that differ slightly in fundamental frequency may still be heard as having the identical tone features of scalar pitch.

I have argued throughout this dissertation for a similar distinction to be made with respect to the chord—a term that is presently defined only by the music-theoretical description of its compositional technique. As we have no acknowledged term for "chord-as-perceived," the definition that describes the chord's performance is occasionally applied to account for the acoustic and perceptual realms as well: as noted by Eric Scheirer, "the term *chord* is used variously to refer to the notation on the page, the action by the performer, the analytic object of artistic interest, the acoustical signal, and the perception that corresponds to this acoustical signal" (Scheirer 2000, 60). By defining chord as a specifically perceptual object, rather than a compositional technique, I have attempted to bring the same distinction of performative, acoustic and perceptual facets of the chord: two or more *notes* form a *sound* that is perceived as a *chord*.

By framing tone and chord within the same perceptual processes, I offer a corollary claim that the difference in perceiving a tone or a chord is a difference in how the listener conceives of the musical object. Both tone and chord are formed by the perception of patterns in the sound stimulus, and whereas the acoustic composition of chords in general affords a greater opportunity for analytic listening—that is, hearing the chord's component tones—this is largely a matter of context: chords such as the Psalms chord offer the listener relatively few cues for segregation, whereas tones may be placed in a musical context that encourages segregation of one or more partials, as in the "overtone singing" practice found in central Asia (*Grove music online*, s.v. "Overtone-singing"). We may also point to the concept of note as a significant difference between tone and chord, in that a tone correlates with a single note, whereas a chord correlates with two or more. However, we may identify liminal cases of musical sounds that may be interpreted as either one note or two: in describing a melodic passage that is doubled at the octave, the question of one note or two is largely an issue of notation—the presence or absence of written octaves in the score—and performance technique—how many "things" the performer must do to produce the sounding octaves.

With these acoustic and perceptual similarities in mind, the exploration of pitch in chapter 4 provided a fruitful path into describing the perceptual nature of chords. As pitch is understood to have a separate acoustic correlate—fundamental frequency—examining pitch perception allows us to identify further aspects of the distinction between sound and tone. By examining the potential for chords to bear a pitch percept, we used a familiar and relatively well-understood musical feature to introduce this same distinction with respect to chords. A fuller understanding of perceiving chords, however, requires us to account for the properties or features that correlate with chord quality—the percepts that differentiate the many types of chords that happen to share the same pitch percepts.

6.2 Toward a perceptual account of chord quality

As the quality of a chord is defined purely within the dimension of frequency—that is, we don't use temporal information to categorize chord quality—we may find an analogue of chord quality within the musical features of a tone: a chord's quality is similar to the spectral content of a tone's timbre, known as **spectral timbre**. Both spectral timbre and chord quality may be used to characterize differences in frequency content that are not manifest in pitch: two identically pitched tones may differ in timbre, just as two chords with the same fundamental may have different qualities. As both quality and timbre are derived from patterns within a sound's spectral content, the difference between perceiving quality or timbre is one of classification (Pressnitzer and McAdams 2000)—it is a consequence of hearing the sound as either a chord or a tone, respectively.[1]

In the following sections, I attempt to further define chord quality as an emergent feature by situating chord quality perception within the same three perceptual factors—acoustics, context, and experience—that I have used to describe the processes of object formation and pitch perception. Exploring these perceptual factors enables us to make the distinction between the perception of chord quality—how we categorize the chord—and the concept of **chord type**—the music-theoretic description of a chord, derived from its notation.

---

[1] This similarity between chord quality and timbre has also been acknowledged in a recent pair of journal volumes (*Analyze musicale* 47–48, 2003) on the topic of "harmony-timbre," as a memorial to Hector Berlioz; articles within this two-volume set discuss the intersection of harmony and timbre concepts in the works of Webern, Varèse, Messiaen, Stravinsky, Risset, and Murail, as well as the works of the dedicatee of these volumes.

6.2.1 Acoustics

The acoustic parameter of chord quality is similar to that of pitch: the chord's frequency spectrum affords a salient pattern that is recognized and associated with a perceptual feature, much as a salient harmonic series leads to a pitch percept. The most significant implication of this statement is that chord quality need not be derived exclusively from the F0's of its component tones; in this regard, it obviously differs from chord type, which is defined in terms of "notes" and intervals. Although the difference between F0-derived and a spectrum-derived definitions may not be significant in many listening situations, this distinction does have important ramifications for our understanding of how we form a chord quality percept: perceiving chord quality is not necessarily dependent upon perceiving the pitches of a chord's component tones.

As support for a spectrum-based definition of chord quality, Kendall and Vassilakis (2006; 2010) found that when listeners were asked to judge the "consonance" of a given major triadic chord set to different instrumentations, their ratings varied with different permutations of instruments—that is, chords that were identical in pitch content took on different emergent features with different instrumentation. Although the "consonance" studied in these experiments is not equivalent to chord quality, the studies do confirm that the features we associate with chords are emergent features, formed at the level of the holistic percept. Further evidence for an effect of spectral timbre on chord quality comes from a study by McMurray, Dennhardt, and Struck-Marcell (2008): in a major/minor triad identification task for sonorities with ambiguous thirds, the authors found a bias toward major when listening to stimuli built with complex tones comprised of the first six harmonics. The authors attributed this bias toward the presence of the fifth harmonic, which forms a "just-tuned" compound major third with the fundamental. Subsequent tests that used pure (sine-tone) stimuli were comparatively unbiased, even though the F0's used in both sets of tests were the same; listeners' tendencies to hear a chord as major or minor were influenced solely by the spectral timbres of the chord stimuli.

6.2.2 Context

The examination of chord pitch in section 4.4.3 found that many of the chords typical to Western music are pitch-ambiguous; in the presence of ambiguous acoustic information, the role of cognitive processes becomes more significant. We may therefore expect context to play a significant role in hearing chord quality, particularly with chords that contain atypical registration or doubling. But there is also a more musically significant reason for an increased role of context in particular: different musical passages employ chords in different ways. The concepts of harmonic tonality permit a single sonority to take on different harmonic roles or "functions" depending upon the tonality in which it appears: for example, a C-major triadic chord serves as the tonic in a C tonality, but acts as a subdominant in a G tonality. Similarly, we may conceive of musical passages in which a given chord's quality might also vary with its context, though such a chord would need to possess a high degree of "quality ambiguity" to permit such a contextual transformation. One chord common to Western harmonic music that provides such ambiguity is the "minor seventh chord," shown in musical notation in Figure 6.1.

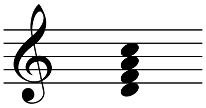

Figure 6.1: Minor seventh chord

The quality ambiguity of this chord in particular is suggested by Rameau's use of context in assigning this chord type a fundamental bass (Christensen 1993, chapter 5): in the key of C major, the chord is understood as a major (subdominant) sonority, with an F fundamental bass, when followed by the tonic chord on C, but it acts as a minor (supertonic) sonority, with a D fundamental, when followed by the dominant chord on G. Although we need not take Rameau's taxonomy as directly suggesting different chord quality percepts, Rameau does outline the essential role of context in shaping our understanding of a sonority: given an ambiguous pattern, we are inclined to hear the chord quality that best fits within our understanding of its musical context.

6.2.3 *Experience*

The role of experience in chord quality perception may be revealed by presenting the chord without a musical context, as an isolated chord. Trained listeners are usually able to identify the chord qualities most common to Western music, and can (for instance) distinguish a major-triadic or minor-triadic chord when heard in isolation; however, this ability to perceive chord quality may not be as developed in listeners without musical training. One study (Heinlein 1928) presented pairs of four-part triadic chords (major and minor) in a same-different task; the author found that even among some musically trained subjects, performance on this task was poor when the third of the triad was presented in an inner voice. Although further research in this paradigm would produce more generalizable results, this apparent difference in ability suggests that hearing chord quality may not result from simple exposure to music with triadic chords.

We would be likely to observe further differences in chord quality categorization based on the listener's knowledge of the harmonic conventions of a musical practice. A listener who has learned to distinguish the sounds and musical functions of major-triadic chords in different inversions, for instance, may choose to hear the two chords shown in Figure 6.2 as belonging to different categories, even though they share the same music-theoretic definition as first-inversion major triads. One perceptual interpretation of these chords might take the first chord as a major triad in first inversion (a "six-three chord"), such as would be found in a passage with a rising bass line; in contrast, the second might be heard as an off-beat component of a pianistic stride or "oom-pah" chordal accompaniment, in which the root-position bass would be understood as occurring on the preceding beat. Of course, listeners with different musical backgrounds may categorize these two chords in alternative manners: perhaps both would be heard as first-inversion major triads, or as belonging to a larger category of major-triadic chords, without specifying inversion.

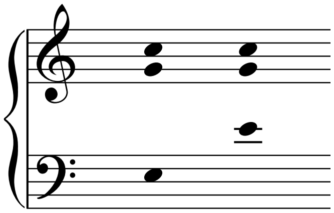

Figure 6.2: Two major-triadic chords in "first inversion"

6.3 A model for future research

We may take the affinities between timbre and chord quality noted above as a suggestion to ground the study of chord quality perception within a growing body of research on musical timbre (Caclin et al. 2007, 2008; McAdams and Giordano 2009). Although the empirical methods of timbre research provide a promising set of paradigms and with which we may study chord quality, I propose that a more closely aligned research goal may be found in the study of phoneme perception, and in particular the perception and categorization of vowels. Unlike instrumental timbre, vowels and chords both use frequency content to present organizational features within their respective domains: just as understanding speech requires hearing vowels, understanding a chordal musical passage requires hearing chord qualities.

For the purpose of comparison with chord quality perception, a very brief history of vowel perception is offered below. Modern phoneme research may be said to begin with the introduction of the spectrogram, a tool used to reveal the frequency and amplitude content that makes up speech stimuli (Potter, Kopp, and Kopp 1947; Jakobson, Fant, and Halle 1952). Working with these spectrograms, researchers identified particularly salient clusters of partials, known as *formants*, that could be observed to vary with vowel quality, and subsequent research identified a number of formant patterns that could be associated with vowels in the English language (Peterson and Barney 1952). These associations between formant patterns and vowel percepts continued to be refined in later research, as our understanding of the effects of speech context (Ladefoged and Broadbent 1957) and the listener's language experience—that is, his or her "native" language (Werker and Tees 1984)—provide a fuller account of vowel perception in a typical speech environment.

An empirical approach to chord quality perception may draw upon the same analytical tools developed for speech research: the spectrogram provides an effective means of capturing the patterns that we associate with acoustic features. One form of pattern that has been used to model chord quality perception is the **chroma profile**—a mapping of a chord's frequency spectrum onto the 12 chroma values of the chromatic scale (Parncutt 2011).[2] A chord rich in partials at the E, G, and B chroma, for instance, contains a chroma profile that affords hearing an E-minor triad. The use of chroma profiles or "chromagrams" in computer transcription tasks for identifying chords (Mauch 2010) suggests a significant parallel with early phonetics research, which was also motivated by the desire to automate the transcription of acoustic information.

Given both a means of recording spectral content and a working definition of an acoustic correlate of chord quality, we may propose empirical paradigms that highlight acoustic, contextual, and experience-based influences upon chord quality perception. The paradigm introduced by Kendall and Vassilakis (2006), which uses variations in timbre instead of F0 to affect the perception of emergent chord features, may serve as a template for further research. Within this paradigm of timbral adjustments, we may examine the perception of chords that have been regarded in music-theoretical research as having ambiguous or debatable chord types. Two such examples have been intro-

---

[2] Although Parncutt (2011) provides empirical evidence for the use of chroma profiles in perceiving harmony, the most significant supporting evidence comes from experiments that use octave-generalized tones; as such tones contain only a single chroma each, they are in essence optimized for chroma perception tasks and models. A more thorough assessment of the perceptual utility of chroma profiles would therefore need to include listening to harmonic complex tones, such as the piano tone examined in section 4.4.1.

duced previously within this dissertation: the “major six-four” sonority analyzed in section 4.4.3, and the “minor seventh” sonority discussed above in section 6.2.2. Varying the amplitudes of the low harmonics in the chord’s tones may affect the perceived fundamental of the chord, such that it is heard as having a different quality.

This proposed study of the acoustic basis of chord quality would provide a base level of knowledge that permits further research on the cognitive factors of chord quality. The paradigms used for the experiments presented in chapter 3—the manipulation of context and familiarity—could then be used to examine the effects of these factors, much as the associations of vowels with patterns of formants was shown to vary with context and experience.

6.4 Conclusion

The exploration of chords in this dissertation began with an observation by Dahlhaus: the chord is located within perception as well as notation. I introduced the concept of the object—used in domain-general listening as well as musical listening—as a way of grounding chords within our knowledge of auditory perception. The listener begins to form musical objects by perceiving acoustic cues, and he or she acts upon these cues within the framework of his or her listening experience and knowledge of musical context.

A series of performance strategies for producing chords was derived from the acoustic and contextual cues that lead to chord perception. From these strategies, I hypothesized that musical instruments designed to promote chord-perception cues played a role in the emergence of chordal practice. An empirical exploration of context and listener experience was used to support this hypothesis by examining the effects of these top-down factors on how a sonority is perceived—as a perceptually integral chord, or as separable tones.

I introduced chord pitch as a familiar, emergent musical feature that may be derived from chord percepts. Because chords often afford pitch ambiguity, there is usually a greater role for top-down processes in perceiving chord pitch. I used the virtual pitch model of Terhardt (1974) to capture the acoustic cues to pitch within a chord, and I more fully explored the effects of context and experience on pitch perception in a musical analysis of a deceptively unusual chord: the opening sonority of Stravinsky's *Symphony of Psalms*.

Within this musical analysis, I suggested that listeners may also perceive chords as having another musically significant emergent feature: chord quality. As chord quality is under-represented in music perception research, I framed a proposal for chord quality perception within the same factors—acoustics, context, and experience—that were used in the examination of object formation and pitch perception. Finally, I outlined a research agenda for chord quality perception that, combined with our further understanding of object formation and pitch perception, may be used to develop a properly perceptual account of chords; the reward for generating such an account is a richer understanding of the musical and conceptual relationships that shape our experience of listening to chords.

APPENDIX A

A MATLAB PROGRAM FOR ANALYZING VIRTUAL PITCH

The program that I have used to perform virtual pitch analyses in this dissertation is provided as a supplemental file; in this appendix I describe its purpose, design, and use.

A.1 Purpose

The virtual pitch model, first proposed by Terhardt (1974), is the basis for a later algorithm designed to calculate virtual pitch percepts (Terhardt, Stoll, and Seewann 1982). This algorithm has been implemented as a program in the C programming language (Terhardt 2004), known as `ptp2svp`—pure-tone pattern to spectral and virtual pitch. The program takes as its input a list of numbers, representing the frequencies and sound pressure levels (dB SPL) of one or more partials, and it returns a list of virtual and spectral pitch candidates.

The MATLAB program created for this dissertation extracts the frequency and SPL information of partials from a recorded sound file, so that sound recordings may be analyzed for virtual pitch content. The program performs four main functions:

1. Prepare digitized sound file for use in the following functions.
2. Identify partials or "peaks" in the sound file, and extract their frequencies and SPLs.

3. Send this information to the `ptp2svp` program for virtual pitch analysis; capture its results, and display them in graphical form.

A.2 Design

A.2.1 Sound file preparation

The MATLAB Signal Processing Toolkit is able to extract data from sound recordings in the Waveform Audio File Format, commonly known as WAV. The process of converting a WAV sound file to a representation of its partials requires two stages of preparation, with an optional third step to parse the file into smaller time segments.

First, the voltage of the WAV file must be converted to sound pressure level, using a reference level of 20 micropascals (μPa). The resulting data may be calibrated to a specified SPL, such that the overall sound may be represented as louder or quieter. Because SPL affects the pitch salience of virtual pitch candidates—Terhardt notes that some virtual pitches are more easily perceived when the overall SPL of the sound is kept relatively low (Terhardt 1977)—higher or lower calibration levels will affect the virtual pitch analysis of a sound. For the analyses in this dissertation, all sounds were calibrated to 70 dB SPL (unweighted), which approximates a comfortable listening level.

Second, the time information in the sound file must be converted to frequency information; this is done using a fast Fourier transform (FFT), which produces a power spectrum—frequency plotted against dB SPL. In this program, the FFT is calculated using a large length—at least five times the length of the sound file—such that the frequency resolution of the resulting transform is maximized; this practice is known as *zero-padding*, as the sound data is "padded" with additional zeroes to increase its overall length (Abe and Smith 2004). Optimizing the FFT for increased frequency resolution requires a tradeoff in the time dimension, with larger FFT lengths requiring larger time windows. As musical chords are relatively static in frequency content, they are well-suited for this technique, and the increased frequency resolution of a large FFT permits the analysis of partials that differ by only a few Hertz in frequency content.

Third, a large sound file may be segmented into smaller time windows, which allows the analysis of each segment. This may be useful when analyzing a sound with dynamic attack and decay phases, such as a sonority played on a chordophone: as the SPLs of the sonority's partials are likely to be much higher during the initial attack than in the subsequent decay phase, the virtual pitch analysis of each time window is likely to differ by small but potentially significant amounts. Windowing is particularly useful when analyzing dynamically changing sounds, such as a carillon bell's tone; as the carillon bell's partials may individually rise and fall in intensity throughout its sounding duration, each window presents a slightly different power spectrum for virtual pitch analysis.

A.2.2 Identification of peaks

The bulk of the processing in this program is done by the "peak picker" routine, which analyzes the power spectrum of the sound data for the presence of partials. Because the FFT procedure is inherently "noisy," in that it identifies phantom activity at frequencies that are not contained in the sound, selecting the frequencies that represent sounding partials is not a straightforward task. The power spectrum shown in Figure A.1 below, taken from the same piano tone analyzed in section 4.4.1, reveals the challenge of this task: identifying the actual sounding partials within the power spectrum involves separating the partials—the "main lobes" of the power spectrum, marked in Figure A.1 with small black squares—from the "side lobes" that are artifacts of the FFT procedure.

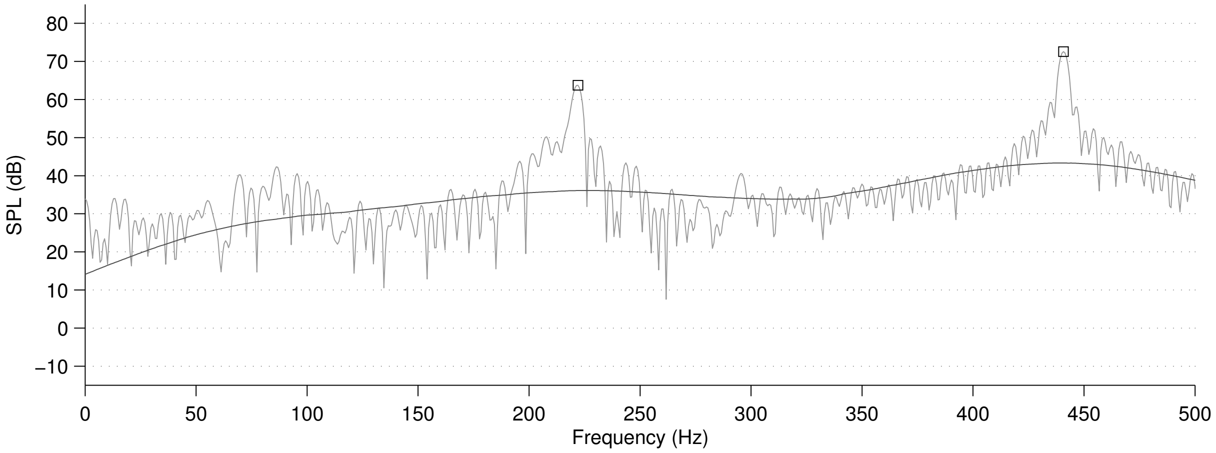

Figure A.1: Power spectrum of a piano tone at A3, 0–500Hz

In this program, identifying partials in a power spectrum is performed using two separate stages. The first stage seeks out local maxima or peaks in the power spectrum; a peak is considered to be any data point that has a larger dB SPL than its adjacent data points. Once the local maxima have been identified, the program performs a second scan for maxima among these peaks—that is, those peaks that are higher than their adjacent peaks. The result of this process is an array that likely includes all main lobes that represent sounding partials, but may also include some side lobes that are artifacts of the FFT process.

A second processing stage compares the dB SPL of a peak with a "running average" SPL, represented by the solid wavy line running transversely in Figure A.1. This average is modified by a windowing function, such that power spectrum data near the selected peak are weighted more heavily than those data further apart in frequency. The resulting vector provides a suitable threshold for measuring peak height; as it rises and falls with SPL, it may be used to measure the height of the peak relative to the overall average SPL at a given frequency. And as partial peaks (main lobes) typically rise significantly above this vector, the vector may be used as a threshold: only those peaks that are at least a specified amount of dB SPL above the vector are retained as representing partials. For the analyses in this dissertation, a typical starting threshold was 10 dB SPL; adjustments up or down could then be used to refine the peak-identification process, such that captured peaks correspond with our knowledge of the frequency content of the analyzed sound.

A.2.3 Virtual pitch analysis and graphs

Having identified the partials' frequencies and SPLs, the program sends this information to the Terhardt virtual pitch program, `ptp2svp`, for virtual pitch analysis. The virtual pitch program returns the frequencies of virtual and spectral pitch candidates, along with their respective pitch weights. After converting frequencies to musical pitch notation (using an equal-tempered scale), this data is plotted in a two-dimensional line plot, pitch name against pitch weight; these plots are used in the virtual pitch analyses appearing throughout this dissertation.

A.3 Use

The virtual pitch analyses used in this dissertation require the MATLAB programming environment (version r2009a or more recent), as well as the collection of files that make up the virtual pitch analysis program; these files are found in the supplemental files that accompany this dissertation.

The main program that launches the virtual pitch analysis, `myvp.m`, takes a number of optional variables; these variables control aspects of the pre-processing and peak detection routines, and they may be adjusted to provide optimal peak identification of a given sound file. The syntax for the program is shown below:

```
myvp(fname, cal, lowf, pscale, wtime, wover, A4)
```

- `fname`: the filename of the sound file, placed in single quotation marks
- `cal`: the dB SPL (unweighted) for calibration of the sound file (default = no calibration)
- `pscale`: the dB SPL threshold above the "running average" required for peak detection (default = 10)
- `lowf`: the mininum frequency, in Hertz, at which a peak will be detected (default = 32)
- `wtime`: the window size, in seconds (default = 0, no windowing)
- `wover`: if windows are used, the percent window overlap (default = 0.5)
- `A4`: the frequency in Hertz of the pitch at A4 (default = 440)

GLOSSARY

analytic Perceiving the part, rather than the whole. See also holistic. 51

attention A perceptual mechanism that permits a listener to perceive one object in the environment with greater detail. 8

auditory chimera An auditory object derived as a composition of other potential objects. 30

auditory object A perceptual object formed of auditory sensations. 4

auditory system The collection of mental and physical facilities, from the outer ear to the brain, that enable the perception of sounds. 6

bottom-up Stimulus-based; the class of processes that act upon stimuli. See also top-down. 25

chord A simultaneity of two or more notes that are heard as a single musical entity. 3

chord quality The perceptual correlate of chord type. 117

chord type The music-theoretic description of a chord, apart from its fundamental or root. 126

chroma The perceptual correlate of pitch class. 77

chroma profile A mapping of a stimulus's frequency content onto the 12 chroma values of the chromatic scale. 132

complex pitch A pitch percept derived from the common frequency attributes of the multiple partials of a sound stimulus; commonly known as pitch. 75

complex tone A sound made up of two or more partial or pure tones. 8

context An environment in which an object is perceived. See also domain. 7

critical band A range of frequencies in which two or more partials are unable to be resolved (have their frequencies perceived) by the basilar membrane, the organ responsible for transforming perceived partials into frequency content. 109

dimension The media in which our perceptual systems to organize and differentiate stimuli; the dimensions of auditory perception are frequency and time. 6

domain An environment that is associated with one or more perceptual tasks. See also context. 7

domain-general The class of processes that apply to all perceptual domains. 25

domain-specific The class of processes that appear to be developed for use in a specific domain. 25

emergent feature A feature of the whole that is not necessarily a feature of its parts. 11

energetic masking The inability of the auditory system to perceive differences in the time or frequency dimensions as segregation cues. See also informational masking. 31

experience The conscious and unconscious learning and conditioning that follows our perceptual actions. 7

F0 Abbreviation for fundamental frequency. 8

feature A stimulus pattern for which our perceptual systems have acquired distinct detection abilities. 6

frequency spectrum The overall frequency content of a sound. 40

fundamental frequency The frequency of the fundamental of a harmonic series; the acoustic correlate of pitch. Abbreviated as F0. 8

fusion The perceptual grouping of simultaneous sounds. 23

grouping The perceptual process of combining multiple partials to form an auditory object. 8

harmonic As a noun, a partial that belongs to a harmonic series; as an adjective, having partials that conform to a single harmonic series. 8

harmonic series Two or more partials that share a common F0. 14

harmonicity The condition in which two or more partials vibrate at frequencies that are multiples of a common F0. 8

holistic Perceiving the whole, rather than the part. See also analytic. 51

informational masking The inability of the auditory system to process numerous simultaneous cues to object formation. See also energetic masking. 37

inharmonic Having partials that do not form a single harmonic series. 33

integrality The extent to which the features or components of a stimulus are perceived as inseparable. 51

intensity The amplitude of a partial. 6

modulation A rapid periodic change of amplitude or frequency. 32

multiharmonic Affording multiple pitch percepts through the presence of multiple salient harmonic series. 96

musical object An auditory object formed within a musical domain. 9

onset synchrony The condition in which two or more partials begin to sound at the same time. 8

partial A single sine-wave component of a complex tone. See also pure tone. 8

perceptual object A mental representation or image of some part of the environment, composed of sensory information. 4

pitch ambiguity Affording multiple pitch percepts; characteristic of multiharmonic sounds. 94

pitch candidate A potential virtual pitch. 87

pitch salience The degree to which a pitch percept may be heard as representing a complex sound; also known as pitch weight or pitch strength. 79

power spectrum The energy of a sound, distributed across its frequency spectrum. 88

process A perceptual operation that contributes to forming an object. 19

pure tone A tone made up of a single partial. 21

scalar pitch A discrete, categorical pitch percept within a musical scale. 9

segregate Perceptually separate a subset of simultaneously occurring stimuli for object formation. 20

sequential A class of processes that seek out regularities or differences in sounds across a span of time. 20

simultaneous A class of processes that seek out regularities or differences among multiple sounds within a single "slice" of time. 20

sonority A simultaneity of two or more notes. 3

sound pressure level The effective sound pressure of a sound, relative to a reference level, commonly measured in decibels (dB SPL); it is the acoustic correlate of the perceptual feature of loudness. 36

spectral pitch A pitch percept that may be derived from a single partial, without the reinforcement of a supporting harmonic series. 88

spectral timbre The spectral content of a sound's timbre; a feature formed of a sound's frequency spectrum. 125

stimulus A unit of environmental information that may be perceived by a sensory system. 4

stream As a verb, the perceptual process of grouping of sounds across time; as a noun, an object formed from sequential grouping. 20

top-down Knowledge-based; the class of processes that act upon knowledge. See also bottom-up. 25

virtual pitch A spectral model of complex pitch perception, proposed by Terhardt (1974); a complex pitch percept formed according to this model. 74

REFERENCES

ANSI. 1994. *American National Standard Acoustical Terminology.* New York: American National Standards Institute.

Abe, Mototsugu, and Julius O. Smith. 2004. Design criteria for simple sinusoidal parameter estimation based on quadratic interpolation of FFT magnitude peaks. In *Proceedings of the 117th convention of the Audio Engineering Society.* San Francisco.

Acker, Barbara E., and Richard E. Pastore. 1996. Perceptual integrality of major chord components. *Perception & Psychophysics* 58 (5): 748–761.

Adams, Reginald B., and Petr Janata. 2002. A comparison of neural circuits underlying auditory and visual object categorization. *NeuroImage* 16 (2): 361–377.

Apel, Willi. 1972. *The history of keyboard music to 1700.* Bloomington: Indiana University Press.

Ashby, F. Gregory, and W. Todd Maddox. 1994. A response-time theory of separability and integrality in speeded classification. *Journal of Mathematical Psychology* 38 (4): 423–466.

Assmann, Peter F., and D. Dwayne Paschall. 1998. Pitches of concurrent vowels. *Journal of the Acoustical Society of America* 103 (2): 1150–1160.

Babbitt, Milton. 1964. Remarks on the recent Stravinsky. *Perspectives of New Music* 2 (2): 35–55.

Barker, Daphne, Christopher J. Plack, and Deborah A. Hall. 2011. Human auditory cortical responses to pitch and to pitch strength. *NeuroReport* 22 (3): 111–115.

Barsalou, Lawrence W. 1999. Perceptual symbol systems. *Behavioral and Brain Sciences* 22 (4): 577.

Beach, D. 1967. The functions of the six-four chord in tonal music. *Journal of Music Theory* 11 (1): 2–31.

Bechtel, William, George Graham, and D. A. Balota. 1998. *A companion to cognitive science.* Malden, MA: Blackwell.

Besseler, Heinrich. 1950. *Bourdon und Fauxbourdon: Studien zum Ursprung der niederländischen Musik.* Leipzig: Breitkopf & Härtel.

Bharucha, Jamshed J., and Keiko Stoeckig. 1986. Reaction time and musical expectancy: Priming of chords. *Journal of Experimental Psychology: Human Perception and Performance* 12 (4): 403–410.

Bigand, Emmanuel, and Barbara Tillmann. 2005. Effect of context on the perception of pitch structures. In *Pitch: Neural coding and perception,* ed. Christopher J. Plack, Richard R. Fay, Andrew J. Oxenham, and Arthur N. Popper, 306–351. New York: Springer Science+Business Media.

Bigand, Emmanuel, Bénédicte Poulin, Barbara Tillmann, François Madurell, and Daniel A. D'Adamo. 2003. Sensory versus cognitive components in harmonic priming. *Journal of Experimental Psychology: Human Perception and Performance* 29 (1): 159–171.

Bispham, John C. 2009. Music's "design features": Musical motivation, musical pulse, and musical pitch. *Musicae Scientiae* 13 (2 suppl): 41–61.

Borchert, Elizabeth M. O., Christophe Micheyl, and Andrew J. Oxenham. 2011. Perceptual grouping affects pitch judgments across time and frequency. *Journal of Experimental Psychology: Human Perception and Performance* 37 (1): 257–269.

Brattico, Elvira, Pauli Brattico, and Thomas Jacobsen. 2009. The origins of the aesthetic enjoyment of music—a review of the literature. *Musicae Scientiae* 13 (2 suppl): 15–39.

Bregman, Albert S. 1990. *Auditory scene analysis: The perceptual organization of sound.* Cambridge, MA: MIT Press.

———. 1993. Auditory scene analysis: Hearing in complex environments. In *Thinking in sound: The cognitive psychology of human audition,* ed. Stephen McAdams and Emmanuel Bigand, 10–36. Oxford: Oxford University Press.

Bregman, Albert S., and Stephen Pinker. 1978. Auditory streaming and building of timbre. *Canadian Journal of Psychology* 32 (1): 19–31.

Brochard, Renaud, Carolyn Drake, Marie-Claire Botte, and Stephen McAdams. 1999. Perceptual organization of complex auditory sequences: Effect of number of simultaneous subsequences and frequency separation. *Journal of Experimental Psychology: Human Perception and Performance* 25 (6): 1742–1759.

Bronkhorst, Adelbert W. 2000. The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions. *Acta Acustica united with Acustica* 86: 117–128.

Burns, Edward M., and W. Dixon Ward. 1978. Categorical perception—phenomenon or epiphenomenon: Evidence from experiments in the perception of melodic musical intervals. *Journal of the Acoustical Society of America* 63 (2): 456–468.

Caclin, Anne, Marie-Helene Giard, Bennett K. Smith, and Stephen McAdams. 2007. Interactive processing of timbre dimensions: A Garner interference study. *Brain Research* 1138: 159–170.

Caclin, Anne, Stephen McAdams, Bennett K. Smith, and Marie-Helene Giard. 2008. Interactive processing of timbre dimensions: An exploration with event-related potentials. *Journal of Cognitive Neuroscience* 20 (1): 49–64.

Camacho, Arturo, and John G. Harris. 2008. A sawtooth waveform inspired pitch estimator for speech and music. *Journal of the Acoustical Society of America* 124 (3): 1638–1652.

Cambouropoulos, Emilios. 2008. Voice and stream: Perceptual and computational modeling of voice separation. *Music Perception* 26 (1): 75–94.

Cariani, Peter A., and Bertrand Delgutte. 1996. Neural correlates of the pitch of complex tones. 1. Pitch and pitch salience. *Journal of Neurophysiology* 76 (3): 1698–1716.

Carlyon, Robert P. 1998. Comments on 'A unitary model of pitch perception'. *Journal of the Acoustical Society of America* 104 (2): 1118–1121.

Carlyon, Robert P., Rhodri Cusack, Jessica M. Foxton, and Ian H. Robertson. 2001. Effects of attention and unilateral neglect on auditory stream segregation. *Journal of Experimental Psychology: Human Perception and Performance* 27 (1): 115–127.

Cherry, E. Colin. 1953. Some experiments on the recognition of speech, with one and two ears. *Journal of the Acoustical Society of America* 25 (5): 975–979.

Christensen, Thomas. 1987. Eighteenth-century science and the corps sonore: The scientific background to Rameau's principle of harmony. *Journal of Music Theory* 31 (1): 23–50.

———. 1993. *Rameau and musical thought in the Enlightenment.* Cambridge, UK: Cambridge University Press.

Clarke, Eric F. 2005. *Ways of listening: An ecological approach to the perception of musical meaning.* New York: Oxford University Press.

Clarkson, Marsha G., Rhonda L. Martin, and Sheridan G. Miciek. 1995. Infants' perception of pitch: Number of harmonics. *Infant Behavior and Development* 19 (2): 191–197.

Cohen, David E. 2001a. The 'gift of nature': Musical 'instinct' and musical cognition in Rameau. In *Music theory and natural order from the Renaissance to the early twentieth century,* ed. Suzannah Clark and Alexander Rehding, 68–92. Cambridge, UK: Cambridge University Press.

———. 2001b. "The imperfect seeks its perfection": Harmonic progression, directed motion, and Aristotelian physics. *Music Theory Spectrum* 23 (2): 139–169.

Collins, Sylvie C. 1985. Duplex perception with musical stimuli: A further investigation. *Perception & Psychophysics* 38 (2): 172–177.

Cusack, Rhodri, and Robert P. Carlyon. 2004. Auditory perceptual organization inside and outside the laboratory. In *Ecological psychoacoustics,* ed. John G. Neuhoff, 15–48. San Diego: Elsevier Academic Press.

Dahlhaus, Carl. 1980. Harmony. In *The New Grove dictionary of music and musicians,* ed. Stanley Sadie. London: MacMillan.

———. 1968/1990. *Studies on the origin of harmonic tonality.* Trans. Robert O. Gjerdingen. Princeton, NJ: Princeton University Press.

Darwin, Christopher J., and Robert P. Carlyon. 1995. Auditory grouping. In *Hearing,* ed. Brian C. J. Moore, 387–424. San Diego: Academic Press.

de Cheveigné, Alain. 2005. Pitch perception models. In *Pitch: Neural coding and perception,* ed. Christopher J. Plack, Richard R. Fay, Andrew J. Oxenham, and Arthur N. Popper, 169–233. New York: Springer Science+Business Media.

DeWitt, Lucinda A., and Robert G. Crowder. 1987. Tonal fusion of consonant musical intervals: The oomph in Stumpf. *Perception & Psychophysics* 41 (1): 73–84.

DeWitt, Lucinda A., and Arthur G. Samuel. 1990. The role of knowledge-based expectations in music perception: Evidence from musical restoration. *Journal of Experimental Psychology: General* 119 (2): 123–144.

Deutsch, Diana, Rachael Lapidis, and Trevor Henthorn. 2008. The speech-to-song illusion. *Journal of the Acoustical Society of America* 124 (4): 2471.

Duane, Ben. 2010. Information content in melodic and non-melodic lines. In *Proceedings of the 11th International Conference on Music Perception and Cognition (ICMPC11).* Seattle.

Duchez, Marie-Elisabeth. 1986. Valeur épistémologique de la théorie de la basse fondamentale de Jean-Philippe Rameau: Connaissance scientifique et représentation de la musique. *Studies on Voltaire and the eighteenth century* 254: 91–130.

Elhilali, Mounya, Ling Ma, Christophe Micheyl, Andrew J. Oxenham, and Shihab A. Shamma. 2009. Temporal coherence in the perceptual organization and cortical representation of auditory scenes. *Neuron* 61 (2): 317–329.

Erickson, Robert. 1975. *Sound structure in music.* Berkley: University of California Press.

Falk, Simone, and Tamara Rathcke. 2010. On the speech-to-song illusion: Evidence from German. In *Proceedings of the 5th Conference on Speech Prosody.* Chicago.

Fastl, Hugo, and Eberhard Zwicker. 2007. *Psychoacoustics: Facts and models.* Berlin: Springer.

Fletcher, Neville H., and Thomas D. Rossing. 1998. *The physics of musical instruments.* 2nd ed. New York: Springer.

Francès, Robert. 1958/1988. *The perception of music.* Trans. W. Jay Dowling. Hillsdale, N. J.: L. Erlbaum.

Fritz, Jonathan B., Mounya Elhilali, Stephen V. David, and Shahib A. Shamma. 2007. Auditory attention—focusing the searchlight on sound. *Current Opinion in Neurobiology* 17 (4): 437–455.

Fuller, Sarah. 1992. Tendencies and resolutions: The directed progression in *ars nova* music. *Journal of Music Theory* 36 (2): 229–258.

Garner, Wendell R. 1974. *The processing of information and structure.* Potomac, MD: L. Erlbaum.

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

Gibson, James Jerome. 1966. *The senses considered as perceptual systems.* Boston: Houghton Mifflin.

Godøy, Rolf Inge. 2009. Chunking sound for musical analysis. In *Computer music modeling and retrieval. genesis of meaning in sound and music,* ed. Sølvi Ystad, Richard Kronland-Martinet, and Kristoffer Jensen, 67–80. Berlin: Springer.

Goldstein, Julius L. 1973. Optimum processor theory for central formation of pitch of complex tones. *Journal of the Acoustical Society of America* 54 (6): 1496–1516.

Griffiths, Timothy D., and Jason D. Warren. 2004. What is an auditory object? *Nature Reviews Neuroscience* 5 (11): 887–892.

Grove music online. Oxford Music Online, s.v. "Chord" (accessed Oct. 5, 2010).

———. Oxford Music Online, s.v. "Organ" (by Barbara Owen, Peter Williams, and Stephen Bicknell) (accessed Oct. 5, 2010).

———. Oxford Music Online, s.v. "Keyboard" (by Nicolas Meeùs) (accessed Oct. 5, 2010).

———. Oxford Music Online, s.v. "Falsobordone" (by Murray C. Bradshaw) (accessed Oct. 5, 2010).

———. Oxford Music Online, s.v. "Overtone-singing" (by Carole Pegg) (accessed Oct. 5, 2010).

Hall, Michael D., and Richard E. Pastore. 1992. Musical duplex perception: Perception of figurally good chords with subliminal distinguishing tones. *Journal of Experimental Psychology: Human Perception and Performance* 18 (3): 752–762.

Hannon, Erin E., and Laurel J. Trainor. 2007. Music acquisition: Effects of enculturation and formal training on development. *Trends in Cognitive Sciences* 11 (11): 466–472.

Harrison, Daniel. 1994. *Harmonic function in chromatic music: A renewed dualist theory and an account of its precedents.* Chicago: University of Chicago Press.

Harrison, Frank Ll., ed. 1980. *Motets of English provenance.* Polyphonic music of the fourteenth century. Monaco: Éditions de l'Oiseau-Lyre.

Hartmann, William M. 1996. Pitch, periodicity, and auditory organization. *Journal of the Acoustical Society of America* 100 (6): 3491–3502.

———. 1997. *Signals, sound, and sensation.* Woodbury, NY: American Institute of Physics.

He, Chao, and Laurel J. Trainor. 2009. Finding the pitch of the missing fundamental in infants. *Journal of Neuroscience* 29 (24): 7718–7722.

Heinlein, Christian P. 1928. The affective characters of the major and minor modes in music. *Journal of Comparative Psychology* 8 (2): 101–142.

Helmholtz, Hermann von. 1877. *On the sensations of tone as a physiological basis for the theory of music.* New York: Dover Publications.

Hibbert, William A. 2008. The quantification of strike pitch and pitch shifts in church bells. Ph.D. thesis, The Open University.

Howard, David, Stuart Rosen, and Victoria Broad. 1992. Major/minor triad identification and discrimination by musically trained and untrained listeners. *Music Perception* 10 (2): 205–220.

Hubbard, Timothy L., and Darcee L. Datteri. 2001. Recognizing the component tones of a major chord. *American Journal of Psychology* 114 (4): 569–589.

Huron, David. 2001. Tone and voice: A derivation of the rules of voice-leading from perceptual principles. *Music Perception* 19 (1): 1–64.

———. 2006. *Sweet anticipation: Music and the psychology of expectation.* Cambridge, MA: MIT Press.

Jakobson, Roman, Gunnar M. Fant, and Morris Halle. 1952. *Preliminaries to speech analysis.* Cambridge, MA: MIT Press.

James, William. 1890. *The principles of psychology.* Vol. 1. New York: Henry Holt.

Jeppesen, Knud, ed. 1962. *Balli antichi veneziani per cembalo.* Copenhagen: W. Hansen.

Judd, Robert. 1995. Italy. In *Keyboard music before 1700,* ed. Alexander Silbiger, 235–311. New York: Schirmer.

Kadosh, Roi Cohen, Warren Brodsky, Michal Levin, and Avishai Henik. 2008. Mental representation: What can pitch tell us about the distance effect? *Cortex* 44 (4): 470–477.

Kendall, Roger A, and Pantelis N. Vassilakis. 2010. Perception and acoustical analyzes of traditionally orchestrated musical structures versus non-traditional counterparts. *Journal of the Acoustical Society of America* 128 (4): 2344.

Kendall, Roger, and Pantelis Vassilakis. 2006. Perceptual acoustics of consonance and disssonance in multitimbral triads. *Journal of the Acoustical Society of America* 120 (5/2): 3676.

Kersten, Daniel, Pascal Mamassian, and Alan Yuille. 2004. Object perception as Bayesian inference. *Annual Review of Psychology* 55 (1): 271–304.

Klein, Mike E., and Robert J. Zatorre. 2011. A role for the right superior temporal sulcus in categorical perception of musical chords. *Neuropsychologia* 49 (5): 878–887.

Kraus, Nina, Erika Skoe, Alexandra Parbery-Clark, and Richard Ashley. 2009. Experience-induced malleability in neural encoding of pitch, timbre, and timing implications for language and music. *Neurosciences And Music III: Disorders And Plasticity* 1169:543–557.

Kubovy, Michael, and Michael Schutz. 2010. Audio-visual objects. *Review of Philosophy and Psychology* 1 (1): 41–61.

Kubovy, Michael, and David Van Valkenburg. 2001. Auditory and visual objects. *Cognition* 80 (1-2): 97–126.

Ladefoged, Peter, and D. E. Broadbent. 1957. Information conveyed by vowels. *Journal of the Acoustical Society of America* 29 (1): 98–104.

Leman, Marc. 2008. *Embodied music cognition and mediation technology.* Cambridge, MA: MIT Press.

Leopold, David A., and Nikos K. Logothetis. 1999. Multistable phenomena: Changing views in perception. *Trends in Cognitive Sciences* 3 (7): 254–264.

Licklider, J. C. R. 1951. A duplex theory of pitch perception. *Journal of the Acoustical Society of America* 23 (1): 147.

Locke, S., and L. Kellar. 1973. Categorical perception in a non-linguistic mode. *Cortex* 9 (4): 355–369.

Lowinsky, Edward E. 1961. *Tonality and atonality in sixteenth-century music.* Berkeley: University of California Press.

Madsen, Søren Tjagvad, and Gerhard Widmer. 2006. Music complexity measures predicting the listening experience. In *Proceedings of the 9th International Conference on Music Perception and Cognition.* Bologna.

Mauch, Matthias. 2010. Automatic chord transcription from audio using computational models of musical context. Ph.D. thesis, University of London.

McAdams, Stephen. 1984. The auditory image: A metaphor for musical and psychological research on auditory organization. In *Cognitive processes in the perception of art,* ed. W. Ray Crozier and Antony J. Chapman. Burlington: Elsevier Science.

McAdams, Stephen, and Bruno L. Giordano. 2009. The perception of musical timbre. In *The Oxford handbook of music psychology,* ed. Susan Hallam, Ian Cross, and Michael Thaut, 72–80. Oxford: Oxford University Press.

McDermott, Josh H., and Andrew J. Oxenham. 2008. Music perception, pitch, and the auditory system. *Current Opinion in Neurobiology* 18 (4): 452–463.

McKeown, Denis, and David Wellsted. 2009. Auditory memory for timbre. *Journal of Experimental Psychology: Human Perception and Performance* 35 (3): 855–875.

McLachlan, Neil. 2009. A computational model of human pitch strength and height judgments. *Hearing Research* 249 (1-2): 23–35.

McMurray, Bob, Joel L. Dennhardt, and Andrew Struck-Marcell. 2008. Context effects on musical chord categorization: Different forms of top-down feedback in speech and music? *Cognitive Science* 32 (5): 893–920.

Meddis, Ray, and Michael J. Hewitt. 1991. Virtual pitch and phase sensitivity of a computer-model of the auditory periphery: 1. Pitch identification. *Journal of the Acoustical Society of America* 89 (6): 2866–2882.

Micheyl, Christophe, Cynthia Hunter, and Andrew Oxenham. 2010. Auditory stream segregation and the perception of across-frequency synchrony. *Journal of Experimental Psychology: Human Perception and Performance* 36 (4): 1029–1039.

Micheyl, Christophe, and Andrew J. Oxenham. 2007. Across-frequency pitch discrimination interference between complex tones containing resolved harmonics. *Journal of the Acoustical Society of America* 121 (3): 1621–1631.

Miller, George A. 1960. *Plans and the structure of behavior.* New York: Holt.

Neisser, Ulric. 1967. *Cognitive psychology.* New York: Appleton-Century-Crofts.

———. 1976. *Cognition and reality: Principles and implications of cognitive psychology.* San Francisco: W. H. Freeman.

Nettl, Bruno. 2000. An ethnomusicologist contemplates universals in musical sound and musical culture. In *The origins of music,* ed. Nils Lennart Wallin, Björn Merker, and Steven Brown, 463–472. Cambridge, MA: MIT Press.

Palmer, Caroline, and Susan Holleran. 1994. Harmonic, melodic, and frequency height influences in the perception of multivoiced music. *Perception & Psychophysics* 56 (3): 301–312.

Palmer, David James. 1998. Polar processes in stravinsky's neoclassical music. Ph.D. thesis, The University of Rochester, Eastman School of Music.

Parncutt, Richard. 2011. The tonic as triad: Key profiles as pitch salience profiles of tonic triads. *Music Perception* 28 (4): 333–366.

Pastore, Robert E., Mark A. Schmuckler, Lawrence Rosenblum, and Rosemary Szczesiul. 1983. Duplex perception with musical stimuli. *Perception & Psychophysics* 33 (5): 469–474.

Patel, Aniruddh D. 2008. *Music, language, and the brain.* Oxford: Oxford University Press.

Peretz, Isabelle. 2006. The nature of music from a biological perspective. *Cognition* 100 (1): 1–32.

Peterson, Gordon E., and Harold L. Barney. 1952. Control methods used in a study of the vowels. *Journal of the Acoustical Society of America* 24 (2): 175.

Piston, Walter. 1947. Strawinsky's re-discoveries. *Dance Index* 6: 256–257.

Plack, Christopher J., Richard R. Fay, Andrew J. Oxenham, and Arthur N. Popper, eds. 2005. *Pitch: Neural coding and perception.* New York: Springer Science+Business Media.

Plomp, Reinier. 1991. J. C. R. Licklider and the case of the missing fundamental. *Journal of the Acoustical Society of America* 89 (4B): 1887.

Potter, Ralph Kimball, George Adams Kopp, and Harriet C. Green Kopp. 1947. *Visible speech.* New York: D. Van Nostrand.

Pressnitzer, Daniel, and Stephen McAdams. 2000. Acoustics, psychoacoustics and spectral music. *Contemporary music review* 19 (2): 33–59.

Rameau, Jean Philippe. 1722. *Traité de l'harmonie reduite à ses principes naturels.* Paris: J.-B.-C. Ballard.

Rand, Timothy C. 1974. Dichotic release from masking for speech. *Journal of the Acoustical Society of America* 55 (3): 678–680.

Rasch, Rudolf A. 1979. Synchronization in performed ensemble music. *Acustica* 43 (2): 121–131.

Remez, Robert E., Philip E. Rubin, Stefanie M. Berns, Jennifer S. Pardo, and Jessica M. Lang. 1994. On the perceptual organization of speech. *Psychological Review* 101 (1): 129–156.

Repp, Bruno H., and Jacqueline M. Thompson. 2009. Context sensitivity and invariance in perception of octave-ambiguous tones. *Psychological Research* 74 (5): 437–456.

Révész, Géza. 1953. *Introduction to the psychology of music.* London: Longmans, Green.

Rosch, Eleanor, Carolyn B. Mervis, Wayne D. Gray, David M. Johnson, and Penny Boyes-Braem. 1976. Basic objects in natural categories. *Cognitive Psychology* 8 (3): 382–439.

Rosenblum, Lawrence D. 2004. Perceiving articuatory events: Lessons for an ecological psychoacoustics. In *Ecological psychoacoustics,* ed. John G. Neuhoff, 219–248. San Diego: Elsevier Academic Press.

Rumelhart, David E. 1980. Schemata: The building blocks of cognition. In *Theoretical issues in reading comprehension: Perspectives from cognitive psychology, linguistics, artificial intelligence, and education,* ed. Rand J. Spiro, Bertram C. Bruce, and William F. Brewer, 33–58. Hillsdale, NJ: L. Erlbaum.

Ryynänen, Matti P., and Anssi P. Klapuri. 2008. Automatic transcription of melody, bass line, and chords in polyphonic music. *Computer Music Journal* 32 (3): 72–86.

Saffran, Jenny R., Elizabeth K. Johnson, Richard N. Aslin, and Elissa L. Newport. 1999. Statistical learning of tone sequences by human infants and adults. *Cognition* 70 (1): 27–52.

Schank, Roger C., and Robert P. Abelson. 1977. *Scripts, plans, goals and understanding: An inquiry into human knowledge structures.* Hillsdale, NJ: L. Erlbaum.

Scheirer, Eric D. 2000. Music-listening systems. Ph.D. thesis, Massachusetts Institute of Technology.

Schneider, Albrecht. 2001. Complex inharmonic sounds, perceptual ambiguity, and musical imagery. In *Musical imagery,* ed. Rolf Inge Godøy and Harald Jørgensen, 95–117. Lisse, The Netherlands: Swets & Zeitlinger Publishers.

Schneider, Albrecht. 2010. Music theory: speculation, reasoning, experience. a perspective from systematic musicology. *Zeitschrift der Gesellschaft für Musiktheorie* Sonderausgabe: Musiktheorie | Musikwissenschaft. Geschichte - Methoden - Perspektiven. `http://www.gmth.de/zeitschrift/artikel/574.aspx` (accessed Feb. 4, 2011).

Schneider, Albrecht, and Marc Leman. 2002. Sonological and psychoacoustic characteristics of carillon bells. In *Proceedings of the 16th meeting of the FWO Research Society on Foundations of Music Research.* Ghent.

Schneider, Peter, and Martina Wengenroth. 2009. The neural basis of individual holistic and spectral sound perception. *Contemporary Music Review* 28 (3): 315–328.

Scholl, Brian J. 2001. Objects and attention: The state of the art. *Cognition* 80 (1-2): 1–46.

Schouten, J. F. 1970. The residue revisited. In *Frequency analysis and periodicity detection in hearing,* ed. Reinier Plomp and G. F. Smoorenburg, 41–54. Leiden, The Netherlands: Sijthoff.

Schulte, Michael, Arne Knief, Annemarie Seither-Preisler, and Christo Pantev. 2002. Different modes of pitch perception and learning-induced neuronal plasticity of the human auditory cortex. *Neural Plasticity* 9 (3): 161–175.

Scruton, Roger. 1997. *The aesthetics of music.* Oxford New York: Clarendon Press Oxford University Press.

Seebeck, August. 1841. Beobachtungen über einige Bedingungen der Entstehung von Tönen. *Ann. Phys. Chem. Annalen der Physik und Chemie* 129 (7): 417–436.

Seither-Preisler, Annemarie, Linda Johnson, Katrin Krumbholz, Andrea Nobbe, Roy Patterson, Stefan Seither, and Bernd Lütkenhöner. 2007. Tone sequences with conflicting fundamental pitch and timbre changes are heard differently by musicians and nonmusicians. *Journal of Experimental Psychology: Human Perception and Performance* 33 (3): 743–751.

Shamma, Shihab A, and Christophe Micheyl. 2010. Behind the scenes of auditory perception. *Current Opinion in Neurobiology* 20 (3): 361–366.

Shamma, Shihab, and David Klein. 2000. The case of the missing pitch templates: How harmonic templates emerge in the early auditory system. *Journal of the Acoustical Society of America* 107 (5): 2631–2644.

Shepard, Roger N. 1964. Circularity in judgments of relative pitch. *Journal of the Acoustical Society of America* 36 (12): 2346–2353.

Shinn-Cunningham, Barbara G. 2008. Object-based auditory and visual attention. *Trends in Cognitive Sciences* 12 (5): 182–186.

Slim, H. Colin, ed. 1975. *Keyboard music at Castell' Arquato.* Neuhausen-Stuttgart, West Germany: American Institute of Musicology.

Snedecor, George Waddel, and William Gemmell Cochran. 1989. *Statistical methods.* 8th edition. Ames, IA: Iowa State University Press.

Strait, Dana L., Nina Kraus, Alexandra Parbery-Clark, and Richard Ashley. 2010. Musical experience shapes top-down auditory mechanisms: Evidence from masking and auditory attention performance. *Hearing Research* 261 (1-2): 22–29.

Straus, Joseph N. 1982. Stravinsky's tonal axis. *Journal of Music Theory* 26 (2): 261–290.

Stravinsky, Igor. 1942/1947. *Poetics of music in the form of six lessons.* Trans. Arthur Knodel and Ingolf Dahl. Cambridge, MA: Harvard University Press.

Stumpf, Carl. 1890. *Tonpsychologie.* Vol. 2. Hilversum: Frits A. M. Knuf.

———. 1911. Konsonanz und Konkordanz. *Beiträge zur Akustik und Musikwissenschaft* 6: 116–150.

Tanaka, James W., and Marjorie Taylor. 1991. Object categories and expertise: Is the basic level in the eye of the beholder? *Cognitive Psychology* 23 (3): 457–482.

Terhardt, Ernst. 1974. Pitch, consonance, and harmony. *Journal of the Acoustical Society of America* 55 (5): 1061–1069.

———. 1977. The two-component theory of musical consonance. In *Psychophysics and physiology of hearing: An international symposium,* ed. Edward Frank Evans and J. P. Wilson, 381–390. London: Academic Press.

———. 1978. Psychoacoustic evaluation of musical sounds. *Perception & Psychophysics* 23 (6): 483–492.

———. 1979. Calculating virtual pitch. *Hearing Research* 1 (2): 155–182.

———. 1984. The concept of musical consonance: A link between music and psychoacoustics. *Music Perception* 1 (3): 276–295.

———. 2000. Affinity of tones. `http://www.mmk.ei.tum.de/persons/ter/top/affinity.html` (accessed Oct. 5, 2010).

Terhardt, Ernst. 2004. ptp2svp: Conversion of part-tone patterns into spectral-virtual-pitch patterns. `http://www.uni-graz.at/richard.parncutt/ptp2svpCode.html` (accessed Oct. 5, 2010).

Terhardt, Ernst, and Manfred Seewann. 1984. Auditive und objektive Bestimmung der Schlagtonhöhe von historischen Kirchenglocken. *Acustica* 54: 129–144.

Terhardt, Ernst, Gerhard Stoll, and Manfred Seewann. 1982. Algorithm for extraction of pitch and pitch salience from complex tonal signals. *Journal of the Acoustical Society of America* 71 (3): 679–688.

Terhardt, Ernst, Gerhard Stoll, R. Schermbach, and Richard Parncutt. 1986. Tonhöhenmehrdeutigkeit, Tonverwandtschaft und Identifikation von Sukzessivintervallen. *Acustica* 61: 57–66.

Thomson, William. 1993. The harmonic root: A fragile marriage of concept and percept. *Music Perception* 10 (4): 385–415.

Tomasello, Michael, Ann Cale Kruger, and Hilary Horn Ratner. 1993. Cultural Learning. *Behavioral and Brain Sciences* 16 (3): 495–511.

Trainor, Laurel J., and Kathleen A. Corrigall. 2010. Music acquisition and effects of musical experience. In *Music perception,* ed. Mari Riess Jones, 89–127. New York: Springer Science Business Media.

Treisman, Anne M. 1993. The perception of features and objects. In *Attention: selection, awareness, and control. A tribute to Donald Broadbent,* 5–35. Oxford: Clarendon.

Treisman, Anne M., and Garry Gelade. 1980. Feature-integration theory of attention. *Cognitive Psychology* 12 (1): 97–136.

Turgeon, Martine, Albert S. Bregman, and Brian Roberts. 2005. Rhythmic masking release: Effects of asynchrony, temporal overlap, harmonic relations, and source separation on cross-spectral grouping. *Journal of Experimental Psychology: Human Perception and Performance* 31 (5): 939–953.

Tversky, Barbara, and Kathleen Hemenway. 1984. Objects, parts, and categories. *Journal of Experimental Psychology: General* 113 (2): 169–193.

Tyler, James, and Paul Sparks. 2002. *The guitar and its music: From the Renaissance to the Classical era.* New York: Oxford University Press.

van Noorden, Leon. 1975. Temporal coherence in the perception of tone sequences. Ph.D. thesis, Technical University Eindhoven.

Van Valkenburg, David, and Michael Kubovy. 2003. In defense of the theory of indispensable attributes. *Cognition* 87 (3): 225–233.

Vogel, Martin. 1993. *On the relations of tone.* Bonn: Verlag für Systematische Musikwissenschaft.

Warren, J. D., Stefan Uppenkamp, Roy D. Patterson, and Timothy D. Griffiths. 2003. Separating pitch chroma and pitch height in the human brain. *Proceedings of the National Academy of Sciences of the United States of America* 100 (17): 10038–10042.

Warrier, Catherine M., and Robert J. Zatorre. 2002. Influence of tonal context and timbral variation on perception of pitch. *Perception & Psychophysics* 64 (2): 198–207.

Werker, Janet F., and Richard C. Tees. 1984. Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development* 7 (1): 49–63.

Wertheimer, Max. 1923/1938. Gestalt theory. In *A source book of Gestalt psychology,* ed. and trans. Willis D. Ellis, 1–11. New York: Harcourt, Brace.

———. 1923/1938. Laws of organization in perceptual forms. In *A source book of Gestalt psychology,* ed. and trans. Willis D. Ellis, 71–88. New York: Harcourt, Brace.

Wickens, Thomas D. 2002. *Elementary signal detection theory.* Oxford University Press.

Winkler, István, Susan L. Denham, and Israel Nelken. 2009. Modeling the auditory scene: predictive regularity representations and perceptual objects. *Trends in Cognitive Sciences* 13 (12): 532–540.

Wright, James K., and Albert S. Bregman. 1987. Auditory stream segregation and the control of dissonance in polyphonic music. *Contemporary Music Review* 2: 63–92.

Wright, James. 2008. Commentary on "Asynchronous preparation of tonally fused intervals in polyphonic music" by David Huron. *Empirical Musicology Review* 3 (2): 69–72.

Xiang, Juanjuan, Jonathan Simon, and Mounya Elhilali. 2010. Competing streams at the cocktail party: Exploring the mechanisms of attention and temporal integration. *Journal Of Neuroscience* 30 (36): 12084–12093.

Zbikowski, Lawrence M. 2002. *Conceptualizing music: Cognitive structure, theory, and analysis.* New York: Oxford University Press.