5.1 Introduction

The comparison of linguistic and musical syntax is a topic that has generated both warm enthusiasm and cool skepticism. The former reaction is illustrated by a set of lectures given by Leonard Bernstein at Harvard in 1973, later published as *The Unanswered Question* (1976). Bernstein, who had long been interested in the analysis of musical structure and meaning (Bernstein, 1959), found inspiration in the generative linguistic theory of Noam Chomsky (1972) and set out to analyze the grammar of Western tonal music in a linguistic framework. As part of his presentation, Bernstein made comparisons between linguistic and musical syntax. Although his musical knowledge and personal charisma make his lectures well worth watching, the details of his exposition were not persuasive to scholars in either linguistics or music. Keeler (1978) has enumerated several of the problems with Bernstein's approach, which include rather strained analogies between linguistic parts of speech such as nouns and verbs and particular musical elements such as motives and rhythms.

Nevertheless, Bernstein's efforts had an important effect on language-music studies. As a result of his lectures, a seminar on music and linguistics was organized at MIT in the fall of 1974, and two of the participants (the musicologist Fred Lerdahl and the linguist Ray Jackendoff) ultimately produced one of the most influential books in music cognition, *A Generative Theory of Tonal Music* (1983). The use of the term "generative" in their title refers to the use of formal procedures to generate a structural description of a given musical piece. This description focuses on four types of structural relations a listener perceives when hearing music. Two of these relations concern rhythm: grouping structure and metrical structure (cf. Chapter 3). The other two relations are more abstract, and concern hierarchies in the relative structural importance of tones ("time-span reduction") and in the patterning of tension and relaxation over time ("prolongation reduction"). Although Lerdahl and Jackendoff adapted the tools of generative grammar to analyze music (cf. Sundberg & Lindblom, 1976), they did not focus on comparisons of linguistic and musical syntax.

Indeed, they were skeptical of such comparisons, noting that "pointing out superficial analogies between music and language, with or without the help of generative grammar, is an old and largely futile game" (p. 5). In support of their skepticism, they point out specific differences between the two syntactic systems, including the lack of musical equivalents for linguistic parts of speech such as nouns and verbs, and differences in the way linguistic and musical "syntactic trees" are constructed (cf. section 5.3.1 below).

Despite Lerdahl and Jackendoff's skepticism, comparisons between linguistic and musical syntax have continued to fascinate scholars. Theoretical treatments of the issue include work by musicologists and linguists (e.g., Swain, 1997; Horton, 2001; Tojo et al., 2006; Pesetsky, 2007; Rohrmeier, 2007). It is fair to say, however, that for each theorist who approaches the topic with enthusiasm, there is another who sounds a note of warning (e.g., Powers, 1980; Feld, 1974). A particularly fascinating example of this dialectic concerns two articles on Javanese gamelan music by leading ethnomusicological scholars (Becker & Becker, 1979, 1983), the first enthusiastically analyzing the grammar of this music in a linguistic framework and the second (by the same authors a few years later) rejecting this original approach as an exercise in empty formalisms yielding little new insight.

Theoretical comparisons between linguistic and musical syntax will no doubt continue for many years to come, with voices on both sides of the issue. In the past few years, however, something new has happened: Empirical studies of this topic have started to emerge in cognitive neuroscience. The appeal of this topic for modern brain science is easy to understand. Linguistic syntax is emblematic of the special abilities of the human mind and has been claimed to engage "domain-specific" cognitive mechanisms (i.e., mechanisms unique to language; see Fodor, 1983). The presence of a second syntactic system in the human mind naturally leads to the question of the relation between them. Are they mentally isolated, modular systems, or might there be cognitive and neural overlap?

This chapter is divided into three parts. The first provides background on musical syntax. The second part discusses formal differences and similarities between musical and linguistic syntax. The final part discusses what neuroscience has revealed about musical-linguistic syntactic relations in the brain. As we shall see, there is evidence for significant neural overlap in syntactic processing in the two domains. Furthermore, exploring the nature of this overlap provides a novel way to explore the cognitive and neural foundations of human syntactic abilities.

Before embarking, it is worth addressing the question, what is "musical syntax?" Because this term may mean different things to different scholars. In this chapter, syntax in music (just as in language) refers to the principles governing the combination of discrete structural elements into sequences. The vast majority of the world's music is syntactic, meaning that one can identify both conceptually discrete elements (such as tones with distinct pitches or drum
sounds with distinct timbres) and norms for the combination of these elements into sequences. These norms are not "rules" that musicians must obey. On the contrary, composers and performers can and do purposely contravene these norms for artistic purposes. However, such departures are meaningful precisely because there are norms against which they operate. The cognitive significance of the norms is that they become internalized by listeners, who develop expectations that influence how they hear music. Thus the study of syntax deals not only with structural principles but also with the resulting implicit knowledge a listener uses to organize musical sounds into coherent patterns.

As with language, the syntax of music varies across cultures and historical eras. Unlike language, however, in which a number of important syntactic features are shared by all human languages (Van Valin, 2001), syntactic universals in music appear to be limited to a few very general features such as the organization of pitch in terms of musical scales with (typically) 5 to 7 tones per octave (cf. Chapter 2). Such universals hardly provide a basis for a detailed comparison of linguistic and musical syntax. This lack of syntactic unity in human music should not be surprising. Unlike language, music is not constrained to transmit a certain kind of information, so that the range of sonic structures considered "music" by at least some people reflects the vast and ever-growing diversity of human aesthetic creativity and interest.

Meaningful comparison of linguistic and musical syntax thus requires focus on the music of a particular period and style. I have chosen to focus on Western European tonal music (or "tonal music" for short), a music that flourished between about 1650 and 1900 and whose syntactic conventions have been influential since that time. (In this chapter, the term "tonality" is sometimes used as a shorthand term for these conventions.) For example, most of the music heard in Europe and the Americas today is tonal music. Another reason to focus on this tradition is that of all known musical systems, it is the most extensively studied from both a theoretical and an empirical perspective (e.g., Krumhansl, 1990; Lerdahl, 2001).

5.2 The Structural Richness of Musical Syntax

Does musical syntax really merit comparison with linguistic syntax? The simple fact that a nonlinguistic system is syntactic does not guarantee an interesting comparison with language. For example, the songs of the swamp sparrow *Melospiza georgiana* are made up of a few acoustically discrete elements ("notes"), and different geographic populations order these notes in different ways to form larger chunks ("syllables") that are repeated in time to create a song (Figure 5.1).

Elegant experiments have shown that these syntactic differences are learned and are meaningful to the birds. Indeed, they serve as the basis of geographic song "dialects" that the birds use in identifying potential competitors or mates (Balaban, 1988; cf. Thompson & Bakery, 1993). Yet such a system can hardly sustain a meaningful comparison with linguistic syntax. Linguistic syntax is remarkable for its structural richness, attaining a level of complexity that sets it apart from any known nonhuman communication system.

One aspect of this richness is multilayered organization. There are principles for the formation of words from meaningful subunits, or "morphemes" (such as the use of the suffix "-ed" in English to form the regular past tense), for the formation of phrases from words (such as noun phrases and prepositional phrases), and for the formation of sentences from phrases. Furthermore, sentence formation includes principles of recursive structure (such as embedding

---

1 I am dealing here with "substantive universals" of musical syntax: structural patterns that appear in most if not all widespread musical cultures. A different approach to the notion of musical universals is to consider the cognitive universals underlying music processing (cf. Lerdahl & Jackendoff, 1983; Jackendoff, 2002:75).

---

Figure 5.1 Swamp sparrow song organization. (a) Examples of the six categories of minimal acoustic elements (notes) that make up swamp sparrow songs. Scale at the left represents frequency (1–8kHz); time scale bar on the bottom is 100 ms. (b) From two to six (most commonly three or four) notes are put together to form syllables. Two syllables from two different New York songs are shown. Birds in a given geographic location have preferences for placing certain notes in certain positions in a syllable; this constitutes the syntax of a song. (c) Swamp sparrow syllables are repeated to form a ~2-sec song. The two songs depicted here consist of repetitions of the two syllables detailed in (b). From Balaban, 1988.
one noun phrase within another) that appear to set human language apart from nonhuman animal communication systems (Hauser et al., 2002; though see Gennier et al., 2006).

Another aspect of the structural richness of linguistic syntax is the strong relationship between syntax and meaning, so that changes in the order of words (and/or the identity of grammatical morphemes) can greatly alter the meaning of an utterance. For example, "The man with the thin cane saw the girl" means something quite different from "The thin girl with the cane saw the man." Peter Marler (2000) has pointed out that this aspect of human syntax sets it apart from other vertebrate syntactic systems such as bird song and whale song, in which the meaning of the sequence is not intricately related to the order in which the elements occur. Instead, current evidence suggests that these nonhuman vocal displays always mean the same thing: territorial warning and sexual advertisement. In these simple syntactic systems, the order of the elements simply identifies the caller as a member of a particular species or group.

A third and very important aspect of the richness of linguistic syntax is the fact that words can take on abstract grammatical functions (such as subject, direct object, and indirect object) that are determined by their context and structural relations rather than by inherent properties of the words themselves (Jackendoff, 2002). For example, there is nothing about the word "cat" that makes it a subject, object, or indirect object, yet in a sentence context it can take on one of these functions, and as a consequence trigger syntactic phenomena at other locations such as subject-verb agreement for number.

The next four sections discuss the syntax of tonal music, illustrating that musical structure has many of the key features that make linguistic syntax so rich.

5.2.1 Multiple Levels of Organization

Like language, tonal music has syntactic principles at multiple levels. The following sections focus on three levels of pitch organization. As background, recall from Chapter 2 that the basic pitch material of tonal music is drawn in an orderly way from the continuum of physical frequencies. Specifically, each octave (doubling in frequency) contains 12 tones such that the frequency ratio between each tone and the tone below it is constant. This basic ratio, called the semitone, is about 6%, equal to the pitch distance between an adjacent black and white key on a piano. Also recall from Chapter 2 that tonal music exhibits octave equivalence, whereby pitches whose fundamental frequencies are related by a 2/1 ratio are perceived as highly similar in pitch and are thus given the same letter name or pitch class irrespective of the octave in which they occur. These 12 pitch classes are given letter names: A, A♯, B, C, C♯, D, D♯, E, F, G, G♯, A, where ♯ = "sharp" (or equivalently, A, A♭, B, C, D♭, D♯, E♭, E, F, G♭, G♯, A♭ in which ♯ = "flat"). The octave in which a note occurs is indicated by a number after its pitch class name, for example, 220 Hz corresponds to A3, whereas 440 Hz is A4.

Scale Structure

The most basic level of syntactic organization of pitch concerns musical scales (cf. Chapter 2, section 2.2.2 for background on musical scales). In tonal music the pitches played at any given moment are not uniformly distributed across the 12 possible pitch classes per octave but are instead constrained by a musical scale, a subset of 7 tones (or "scale degrees") per octave with an asymmetric pattern of pitch spacing ("intervals") between them. One such scale (the C major scale) is shown in Figure 5.2.

As noted in Chapter 4 (section 4.2.6, which should be read as background for this section), in a musical context, the different scale tones take on different roles in the fabric of the music, with one tone being the structurally most central and stable (the "tonic"). The use of one pitch as a tonic is not restricted to Western European tonal music, but appears repeatedly in diverse musical traditions. This suggests that the organization of pitch around a tonic may be congenial to the human mind, perhaps reflecting the utility of psychological reference points in organizing mental categories (Rusch, 1975; Krumhansl, 1979; cf. Justus & Hutsebaut, 2005).

An interesting aspect of scale structure in tonal music is that listeners' intuitions of the stability of scale degrees is not simply binary, with the tonic being stable and all other tones being equally less stable. Instead, empirical evidence suggests that there is a hierarchy of stability. An important aspect of this hierarchy is the contrast in stability between the tonic and its neighboring scale tones (scale degrees 2 and 7), which creates a psychological pull toward the tonic. This is reflected in the music-theoretic names for the 2nd and 7th scale degrees: the second is called the "supertonic" (i.e., the tone just above the tonic), and the seventh is known as the "leading tone" (i.e., the tone that leads to the tonic). In an early set of studies, Robert Frank's (1988) provided evidence for the "pull to the tonic" by demonstrating that listeners were less sensitive to upward mistunings of the leading tone when it was in an ascending melodic context, in other words, when the mistuning brought it closer to the tonic.

Another approach to the mental representation of scale structure concerns the perceived relatedness (or mental distance) between different tones in a scale. Krumhansl (1979) explored this issue using a paradigm in which listeners first

---

2 By including the notion of a tonal hierarchy in the concept of "scale," I am using the term "scale" to include what Dowling (1978) has called "mode" in his discussion of scale structure in music.
heard a tonal context (e.g., an ascending or descending C major scale) and then heard two comparison tones from the scale. The task was to judge how closely related the first tone was to the second tone in the tonal system suggested by the context. The results of this task were analyzed using multidimensional scaling. This technique translates judgments of relatedness into a spatial display so that the closer the perceived relatedness, the closer the elements are in the resulting graph. The three-dimensional solution for the similarity ratings is shown in Figure 5.3.

As can be seen, scale degrees 1, 3, and 5 of the C major scale (C, E, and G) are perceived as closely related, whereas the remaining scale tones are less closely related, and nonscale tones are distant and related. One very notable feature of this figure is the large distance that separates tones that are adjacent in frequency, such as C and G. This contrast between the psychological and physical proximity of pitches is likely to be part of what animates tonal music.

Listeners appear to be quite sensitive to scale structure in tonal music, as evidenced by the familiar phenomenon of the “sour note,” as discussed in Chapter 4, section 4.2.6. At what age does this sensitivity emerge? Trainor and Trehub (1992) examined 8-month-old infants’ and nonmusician adults’ ability to detect two types of changes in a repeating 10-note melody. (The melody was transposed on each repetition, so the task involved discerning a change in pitch relationships rather than simply detecting an absolute pitch change.) In both cases, one note in the middle of the melody was changed: In one case, it was raised by four semitones, but remained within the scale of the melody; in another case, it was raised just one semitone, but now departed from the scale of the melody. Thus the two changes cleverly pitted physical distance against scale membership. Infants detected both kinds of changes equally well.

Adults performed better than infants overall, but crucially, they detected the change that violated scale structure significantly better than the within-scale change, even though the former change was a smaller physical change than the latter. This reflects the fact that for the adults the nonscale tone “popped out” as a sour note. These results show that the infants had not yet developed an implicit knowledge of scale structure, as one might expect. It is interesting to note that infants have already started acquiring learned sound categories for language at 10 months (e.g., the vowels of their native language; Kuhl et al., 1992), which may reflect the greater amount of linguistic versus musical input that infants have experienced by that age.

What is the earliest age that sensitivity to scale membership can be demonstrated? In a follow-up study using a very similar paradigm, Trainor and Trehub (1994) showed that 5-year-old children with no formal training in music detected out-of-scale melodic alterations better than within-scale alterations, even though the former changes were physically smaller than the latter.3 This naturally raises the question of the ontogeny of scale-membership sensitivity.

3 It should be noted that in this study, melodies were not presented in transposition, because in the same-different task the children tended to respond “different” to transpositional changes.
between 10 months and 5 years. Because behavioral tests with young children can be difficult, it may be preferable to use event-related brain potentials (ERPs) in such studies. ERPs do not require a behavioral response from the listener, and distinct ERP responses to out-of-scale notes have been observed in adults (e.g., Besson & Faita, 1995).

Chord Structure
A very important aspect of tonal music's syntax is the simultaneous combination of scale tones into chords, creating harmony. Chords are formed in principled ways: basic “triads” are built from scale degrees separated by musical thirds, in other words, by a distance of two scale steps. Because of the asymmetric interval structure of Western scales, a distance of two scale steps can correspond to a distance of either three or four semitones, in other words, to a minor or major third. For example, in the C major scale (cf. Figure 5.2), the chord C-E-G consists of a major third between C and E and a minor third between E and G, whereas D-F-A consists of a minor third between D and F and a major third between F and A. These two chords represent “major” and “minor” triads, respectively. As shown in Figure 5.4, building triads from a major scale results in three major triads (built on scale degrees, 1, 4, and 5), three minor triads (built on scale degrees 2, 3, and 6), and one “diminished” triad in which both intervals are minor thirds (built on scale degree 7).

In chordal syntax, one tone of each chord acts as its “root” or structurally most significant pitch. This is the lowest note in each triad in Figure 5.4, and is the note that gives the chord its name as well as its Roman numeral harmonic label. For example, in Figure 5.4, the chord with a root of E (the 3rd scale degree) is an E-minor chord (E-G-B), with a harmonic label of “iii.” (The use of a lower case roman numeral indicates that this chord is a minor chord.) Similarly, in Figure 5.4, the chord with a root of G (the 5th scale degree) is a G major chord (G-B-D), with a harmonic label of V. (The use of an upper case roman numeral indicates that this chord is a major chord.) Even when the notes of a triad occur in a different vertical ordering, the root and harmonic label remain the same, and the chord is treated as having the same basic harmonic status. Thus C-E-G and G-E-C both have C as the root and harmonically labeled as “I”; the latter is simply considered an “inversion” of the C-E-G chord.

Chord syntax also includes principles for modifying triads with additional tones. For example, one very common modification is to add a fourth tone to a triad to convert it to a “seventh” chord, so called because the added tone is seven scale steps above the root of the chord. For example, in a C major scale, the chord G-B-D-F would be a seventh chord built on the root G, or G7 (cf. Figure 5.4), and its harmonic label would be V7. Seventh chords play an important role in chord progressions, by implying forward motion toward a point of rest that has not yet been reached.

The above discussion of chord syntax concerns the “vertical” organization of tones in music. Another important aspect of chord syntax concerns the “horizontal” patterning of chords in time. In tonal music, there are norms for how chords follow one another (Piston, 1987; Huron, 2006), and these norms play a role in governing the sense of progress and closure in musical phrases. A prime example of this is the “cadence,” a harmonic resting point in music. An “authentic cadence” involves movement from a V chord (or a V7 chord) to a I chord and leads to a sense of repose. Moving beyond this simple two-chord progression, some longer chord progressions can be identified as prototypical in tonal music, such as I-V-I, I-IV-V-I, II-V-I, and so on. One of the governing patterns behind these progressions is the “cycle of fifths” for chords, a sequence in which the roots of successive chords are related by descending fifths. In its entirety, the progression is I-IV-VII-III-II-V-I. Smith and Melara (1990) have shown that even musical novices are sensitive to syntactic prototypicality in chord progressions, showing that implicit knowledge of these progressions is widespread among listeners.

Chord sequences are also important in melody perception, in which the chords are implied by important melody tones rather than explicitly played as simultaneities of tones (cf. Chapter 4, section 4.2.8). Listeners are sensitive to this implied harmony. Cuddy et al. (1981) have shown that melodic sequences that imply prototypical chord sequences are better remembered than other sequences. Furthermore, Trainor and Trehub (1994) have shown that musically unselected adults are more sensitive to melodic changes that violate the implied harmony than to physically larger changes that remain within the implied harmony (cf. Holter et al., 1998).
Like the tones of the scale, different chords built on the 7 different scale degrees are not equal players in musical contexts. Instead, one chord (the tonic chord, built on the 1st scale degree) is the most central, followed by the dominant chord (built on the 5th scale degree) and the subdominant chord (built on the 4th scale degree). Informal evidence for the structural importance of the tonic, subdominant, and dominant chords (harmonically labeled as I, IV, and V chords) comes from the fact that many popular and folk songs can be played using just these three chords as the underlying harmony. More formal evidence comes from a study by Krumhansl et al. (1982) in which a musical context (an ascending scale) was followed by two target chords. Listeners were asked to judge how well the second chord followed the first in the context of the preceding scale. The judgments were then subject to multidimensional scaling in order to represent perceived relatedness as spatial proximity. Figure 5.5 shows the multidimensional scaling solution, and reveals that chords I, IV, and V form a central cluster around which the other chords are arrayed.

![Figure 5.5](image)

**Figure 5.5** Psychological relatedness of different chords in a musical context. Chords are indicated by their harmonic labels, with uppercase Roman numerals used in a generic fashion (i.e., major, minor, and diminished chords are not distinguished). From Krumhansl et al., 1982.

---

**Key Structure**

A scale and its tonal hierarchy, plus its system of chords and chord relations, defines a “key” or tonal region in Western European music. Because there are 12 pitch classes in tonal music, each of which can serve as the tonic of a scale, and because there are two commonly used scale structures (major and minor), there are 24 keys in tonal music. Keys are named for their principal note and their scale structure, for example, C major, B minor. Thus keys and scales are named in a similar way, which may be one reason that people sometimes confuse the two.

A great deal of tonal music moves between keys during the course of a composition. These “modulations” of key allow a composer to explore different tonal regions and add diversity to the tonal journey outlined by a piece. Part of the syntax of tonal music is the pattern of key movement in music, which is far from a random walk between the 24 possible keys. Instead, modulations tend to occur between related keys, in which relatedness is defined in particular ways. Major keys are considered closely related if they share many of their basic scale tones. For example, the notes of the C major scale (C, D, E, F, G, A, B, C) and the G major scale (G, A, B, C, D, E, F, G) differ only in terms of one pitch class. (Recall that a major scale is obtained by starting on one note and choosing subsequent tones according to the major scale interval pattern [2 2 1 2 2 1 2].) Generalizing this relationship, any two keys whose 1st scale degrees are separated by a musical perfect fifth are closely related, because their scales share all but one pitch class. This pattern of relations can be represented as a “circle of fifths” for major keys (Figure 5.6).

Music theory also suggests that each major key is also closely related to two different minor keys. One is the “relative minor,” which shares the same notes of the scale but has a different tonic. For example, A minor (A, B, C, D, E, F, G, A) is the relative minor of C major, because it has all the same pitch.

![Figure 5.6](image)

**Figure 5.6** The circle of fifths for major keys. Each key is represented by a letter standing for its tonic. Keys that are adjacent on the circle share all but one pitch class.
classes. (Recall that the minor scale is obtained by starting on one note and choosing subsequent notes by following the minor-scale interval pattern of 2 1 2 2 1 2 1.) The other minor key related to a given major key is the “parallel minor,” which shares the same tonic but has different scale tones. Thus C-minor (C, D, Eb, F, G, A, Bb, C) is the parallel minor of C major.

One way to represent this pattern of relationship among musical keys is via a geometric diagram in which psychological distance between keys is reflected by spatial distance. One such diagram, proposed by Krumhansl and Kessler (1982) on the basis of perceptual experiments, is shown in Figure 5.7. An important aspect of this two-dimensional diagram is that the left and right edges are equivalent, as are the top and bottom edges. That is, the map is actually an unfolded version of a shape that is circular in both dimensions (a torus), reflecting the circular nature of perceived key relations.

An interesting form of evidence for implicit knowledge of key distances comes from experiments in which listeners hear a melody followed by a transposed version of the same melody and must judge whether the two melodies are the same or different. A number of researchers have found that performance on this task is better if the melody is transposed to nearby versus a distant key (e.g., Cuddy et al., 1981; Trainor & Treby, 1993; cf. Thompson & Cuddy, 1992). There is also neural evidence for implicit knowledge of key structure. When listening to a chord sequence in a particular key, an “alien” chord from a distant key produces a larger P600 (an event-related brain potential elicited by structural incongruity) than an alien chord from a nearby key, even when both chords contain the same number of out-of-key notes (Patel, Gibson, et al., 1998). Janata et al. (2002) have also provided evidence for maps of key distance in the brain, using the technique of functional magnetic resonance imaging (fMRI).

### 5.2.2 The Hierarchical Structure of Sequences

One of the principal features of linguistic syntax is that relationships between words are not simply based on nearest neighbor relations. For example, consider the sentence, “The girl who kissed the boy opened the door.” Although the sentence contains the sequence of words “the boy opened the door,” a speaker of English knows that the boy did not do the opening. This is because words are not interpreted in a simple left-to-right fashion, but via their combination into phrases and the combination of phrases into sentences. Figure 5.8 shows a syntactic tree diagram for this sentence specifying the hierarchical organization of words in relation to each other.

As with language, structural relations in tonal music are not merely based on adjacency. Instead, events are organized in a hierarchical fashion. The structure

![Figure 5.7 A map of psychological distances between musical keys. Major keys are indicated by uppercase letters and minor keys by lowercase letters. Dashed lines extending from the key of C major indicate related keys: two adjacent major keys along the circle of fifths (G and F; cf. Figure 5.6) and two related minor keys (see text for details). Modified from Krumhansl & Kessler, 1982.](image)

![Figure 5.8 The hierarchical syntactic structure of an English sentence. (S = sentence; NP = noun phrase, VP = verb phrase, S' = sentence modifier [relative clause], N = noun; V = verb; Det = determiner; Rel-Pro = relative pronoun.) Within the clause, the relative pronoun “who” is referred to as a filler and is interpreted as the actor for the verb “kissed.” This relationship is identified by the presence of a coindexed empty element e, in the subject position of the relative clause. Modified from Patel, 2003b.](image)
of these hierarchies is a major focus of modern cognitively oriented music theory, as exemplified by Lerdahl and Jackendoff's generative theory of tonal music (GTDM). Figure 5.9 shows a hierarchical structure for the tones in a musical passage according to GTDM. The details of this syntactic tree will be explained in the next section. For now, two conceptual points need to be made. First, this type of hierarchy is an "event hierarchy," which describes structural relations in a particular sequence of music. This must be clearly distinguished from the "pitch hierarchies" described in previous sections, which concern overall, atemporal aspects of the tonal musical style, for example, the fact that scale degree 1 is the most structurally stable tone in the scale (Bharucha, 1984a). Pitch hierarchies are only one factor that influences the construction of event hierarchies. The second point is that music theory posits two types of event hierarchies for musical sequences, describing different kinds of structural relations between musical events (Lerdahl & Jackendoff, 1983; cf. Jackendoff & Lerdahl, 2006). These will be discussed in turn below.

Musical Event Hierarchies I: Structure and Ornamentation

The concept that some pitches serve to elaborate or ornament others is central to Western European music theory (see, e.g., the theories of Schenker, 1969; Meyer, 1973; and Lerdahl & Jackendoff, 1983; cf. Cook, 1987a). The ability to recognize a familiar tune in a richly ornamented jazz version, or more generally, the ability to hear one passage as an elaborated version of another, implies that not all events in a musical sequence are perceived as equally important. Instead, some events are heard as more important than others. Note that calling some pitches "ornamental" is not meant to imply that they are aesthetically more or less important than other pitches. Rather, the distinction is meant to capture the fact that not all pitches are equal in forming the mental gist of a musical sequence. It is also worth noting that the concept of melodic elaboration is not unique to Western European tonal music (cf. Chapter 4, section 4.2.7).

One particularly clear treatment of event hierarchies of structure and ornamentation is Lerdahl and Jackendoff's theory of "time-span reduction." The tree in Figure 5.9a is a time-span reduction of two phrases of a musical melody, showing the hierarchy of structural importance for the tones in this passage. Shorter branches terminate on less important pitches, whereas longer branches terminate on more important pitches.¹

¹ The oval shape in the right part of the tree is meant to indicate that the short motif that projects upward to it from the left (the first four notes of the second phrase, C-C-E-G) is subordinate to both branches touched by the oval, not just the left branch (cf. Lerdahl & Jackendoff, 1983:138). This sublety is not essential for the current discussion.

Construction of such a tree requires decisions about which pitches are more structural than others; these decisions are influenced by tonal hierarchies, but also take rhythmic and motivic information into account. The use of a tree structure to indicate structure versus ornament (rather than a simple binary scheme whereby each pitch is either structural or ornamental) is based on the hypothesis that music is organized into structural levels, so that a pitch that is structural at one level may be ornamental at a deeper level. Thus taking a cross section of the tree at any particular height leaves one with the dominant events at that level (cf. Figure 5.9b). The trees are thus meant to model an experienced listener's intuitions about levels of relative structural importance of tones.

Although the notion of structure versus ornament is deeply ingrained in theories of tonal syntax, empirical studies of this issue have been relatively rare. One important study was conducted by Large et al. (1995), who examined pianist's improvised variations on children's melodies, such as the one shown in Figure 5.9. The pianists began by playing a melody from notation, and then produced five simple improvisations on the same melody. Large et al. reasoned that the structural importance of a pitch would be reflected in the number of times it was preserved in the same relative location across the variations. Consistent with this idea, the authors found substantial variation in the extent to which different pitches were retained across improvisations, suggesting that the pianists had a notion of the structural skeleton of the melody. The pattern of pitch retention could be accounted for largely on the basis of the pitch hierarchy of different scale degrees in combination with note duration and degree ofmetrical accent on a given note (all of which are incorporated in Lerdahl and Jackendoff's time-span reduction). Thus in these melodies, a typical "elaboration" pitch was one that occurred on a scale degree of low tonal stability, was of short duration, and was not aligned with an accented beat.

Large et al.'s study provides insights on elaboration in music performance, but leaves open the question of the perception of elaboration relations by listeners. Ideally, one would like a measure of the perceived structural importance of each pitch in a sequence, resulting in a tone-by-tone profile that could be analyzed in a quantitative fashion. Such measurements are difficult to make in practice, however, and creative approaches are needed in this area. One relevant study is that of Bharucha (1984b), who used a memory experiment to show that the salience of a tonally unstable note was influenced by its serial position relative to the following note. Specifically, Bharucha demonstrated that an unstable note that is immediately followed by a tonally stable pitch neighbor (e.g., B-C in a C-major context) is less prominent/detectable than an unstable note that is not "anchored" in this way. It is as if the stable tone subordinates the preceding tone as a local ornament and makes it less conspicuous than if
that same tone were inserted randomly into the sequence. This suggests that the
tonal hierarchy is involved in perceived elaboration relations in music. There
is clearly room for more work in this area, however, aimed at generating a
note-by-note metric of the perceived structural importance of events in musical
sequences.

Musical Event Hierarchies II: Tension
and Resolution

Central to the experience of tonal music is a listener's sense of tension and
resolution as a piece unfolds in time (Swain, 1997). Lerdahl and Jackendoff

Figure 5.9 (A) A time-span reduction of the first two phrases of the children's song
"Hush Little Baby." Shorter branches terminate on less important pitches, whereas
longer branches terminate on more important pitches. (b) The lower staves show the
dominant events at successively higher levels of tree structure. Modified from Large
et al., 1995.

Figure 5.10 A prolongation reduction of a phrase from a composition by J. S. Bach
(Christus, der ist mein Leben). In this type of tree, right branching indicates an increase
in tension and left branching a decrease in tension (i.e., a relaxation). The tree shows
how local tensing and relaxing motions are embedded in larger scale ones. Modified
from Lerdahl, 2001:32.
chord locally tenses into the third. The fourth chord (the point of maximum tension in the phrase) is the first event originating from a right branch that attatches high up in the tree, and represents an increase in tension at a larger level. Following this chord, local relaxations into chords 5 and 6 are followed by local tensing movements before a more global relaxation, indicated by the left branch connecting chord 6 and the final chord. Note that the construction of trees such as that of Figure 5.10 relies on time-span reduction but is not determined by it: The two kinds of trees can organize the musical surface in different ways (see Lerdahl & Jackendoff, 1983, Chs. 5-9 for details). Thus prolongation reduction is another kind of event hierarchy that relates events (typically chords) to each other in ways that are more complex than simple nearest-neighbor relations.

Evidence that tension is actually perceived in a hierarchical fashion comes from studies in which listeners rate perceived tension as they listen to musical passages (Krumhansl, 1996). Such empirically measured "tension profiles" can then be compared to predictions based on numerical models of tonal tension based on computing psychological distances between pitches, chords, and keys. One such model is discussed in section 5.4.3 (subsection on tonal pitch space theory). The tonal pitch space (TPS) model can generate predictions based on hierarchical versus purely sequential analyses of tension-relaxation relationships, so that one can determine which type of structure better fits the empirical data. Research in this area has produced apparently contradictory evidence, with some studies favoring a purely sequential structure of perceived tension-relaxation patterns, whereas other studies support a hierarchical structure. Closer examination of these studies suggests similar differences might arise from different paradigms used to collect tension ratings. For example, in Bigand and Parnicutt’s (1999) study, listeners heard increasingly long fragments of chord sequences and made tension ratings at the end of each fragment. This "stop-tension" task has the advantage of temporal precision, but suffers from an unnatural listening situation that may encourage local rather than global listening. Indeed, Bigand and Parnicutt found that the tension profiles of their listeners were well modeled using local harmonic structure (especially cadences), with a negligible contribution of hierarchical structure. In contrast, research using a "continuous-tension" task, in which a listener moves a slider while listening to an ongoing piece, has provided evidence for hierarchical structure in perceived tension patterns (Lerdahl & Krumhansl, 2007; Smith & Cuddy, 2003). Thus there is evidence that musical tension and relaxation is in fact organized in a hierarchical fashion, though more work is needed in this area.1

1 One important direction for future research using TPS is to design musical sequences in which predictions based on purely sequential relations are very different from predictions based on hierarchical relations (cf. Smith & Cuddy, 2003). It would be particularly interesting to determine if some listeners hear more sequentially, whereas others hear

5.2.3 Context Dependent Structural Functions

It is evident from the discussion above that tonal music has a rich syntax, but one important issue has not yet been addressed. To what extent does this syntax reflect abstract cognitive relationships between sounds, versus psychoacoustic relationships? Put another way, are syntactic relationships in tonal music merely a natural outcome of the psychoacoustic facts of sound, such as the overtones series and the smoothness or roughness of certain frequency intervals between tones? (Cf. Chapter 2 for a review of the overtones series and the sensory qualities of different frequency intervals.) If so, this might imply that tonal syntax lacks the abstractness of linguistic syntax, because psychological relations between elements reflect physical properties of sounds rather than purely conventional structural relations (cf. Bigand et al., 2006).

The search for a physical basis for tonal syntax dates back to the theorist Rameau (1722), and has had strong advocates ever since. Its appeal is understandable, in that it links music perception to a more basic level of sound perception. The advocates of this view in the last century include Bernstein (1976), who argued that the overtones series provides the foundation for musical scales and for the existence of a tonal center in music. More recently, Parnicutt (1989) has provided a sophisticated quantitative analysis of harmony that seeks to understand structural relationships between chords in tonal music on the basis of their psychoacoustic properties (see also Huron & Parnicutt, 1993; Leman, 1995; Parnicutt & Bregman, 2000).

Debates between "physicalist" and "cognitivist" views of musical syntax have existed for some time. For example, the culturally widespread importance of the octave and fifth in musical systems is likely due to universal acoustic/auditory mechanisms (cf. Chapter 2). On the other hand, in a review of Bernstein's The Unanswered Question, Jackendoff (1977) points out that the match between the natural harmonic series and the details of musical scales (melodic,
pentatonic scales) is actually not that good, and there are cultures in which musical scales have little relationship to the overtone series, yet whose music still has a tonal center. Back on the physicist side, Leman (2000) has shown that some of Krumhansl's probe-tone ratings can be accounted for on the basis of a particular model of auditory short-term memory. Back on the cognitivist side, research on the perception of chord relations that has directly pitted predictions based on psychoacoustics against those based on conventional harmonic relations has found evidence for the latter (Telman & Bharucha, 1998; Bigand et al., 2003). There is little doubt that this debate will continue. At the current time, the evidence suggests that Western European musical tonality is not a simple byproduct of psychoacoustics, but is not a blatant contradiction of it either. That is, the psychoacoustic properties of musical sounds appear to provide a "necessary but insufficient" basis for tonal syntax (Lerdahl, 2001; cf. Koelsch et al., 2007, for relevant neural data).

One strong piece of evidence for a cognitivist view of tonal syntax is that certain psychological properties of musical elements derive from their context and structural relations rather than from their intrinsic physical features. One such a property is the "harmonic function" of different chords. The harmonic function of a chord refers to the structural role it plays in a particular key. For example, consider the two chords G-B-D and C-E-G. In the key of C major, these chords play the role of a V chord and a I chord, respectively, whereas in the key of G major, they play the role of a I chord and a IV chord. The different feeling conveyed by these different functions is illustrated by Sound Examples 5.2a and 5.2b, in which these two chords form the final chords of sequences in C major and G major, respectively. In the former case, the two chords form an authentic cadence (IV), and bring the phrase to a musical conclusion. In the second case, the same two chords (and thus physically identical sound waves) act as a IV progression and leave the phrase sounding unfinished. Numerous studies have shown that musically untrained listeners are sensitive to this syntactic difference. For example, they react more quickly and accurately in judging whether the final chord is mismatched when it functions as a I chord than when it functions as a IV chord (Bigand & Pineau, 1997; Tillmann et al., 1998), and show different brain responses to the same chord when it functions in these two distinct ways (Regnault et al., 2001; cf. Poulin-Charron and et al., 2006). Bigand et al. (2001) have used such "harmonic priming" experiments to show that the difference between chord functions even influences linguistic processes. They constructed sequences in which chord progressions were sung in four-part harmony, with each chord corresponding to a different nonsense syllable. The listeners' task was simply to indicate if the final syllable contained the phone /I/ or /U/. The correct responses of both musicians and nonmusicians were faster when the final chord functioned as a I chord rather than as a IV chord.

Evidence for an even broader level of abstraction regarding harmonic function comes from research examining chord perception in isolation versus in musical contexts. This work suggests that physically different chords are difficult to tell apart in a musical context if they have a similar harmonic function (in the sense of providing a similar prediction for a subsequent chord). Specifically, it has been shown that listeners can easily distinguish a IV chord from a I$^6$ chord in isolation. Yet the same listeners find it very difficult to distinguish a I-IV-V chord sequence from a I-$\text{V}^7$-V sequence (Hutchins, 2003). This is likely due to the fact that it is quite common for the IV or I$^6$ chord to function in a similar way in tonal music, in other words, as a preparation for a V chord. (Theorists refer to the function played by the IV or I$^6$ chord in this context as the "subdominant function," named after the IV chord, which is the prototypical chord fulfilling this function; cf. Lerdahl & Jackendoff, 1983:192.) A crucial condition in this study demonstrated that listeners could distinguish a V from a I$^6$ chord both in isolation and in a musical context (IV-V vs. I-$\text{V}^7$-V), though the difference between these chords is akin to that between the IV and I$^6$ chords. Importantly, the V and I$^6$ chords are not functionally similar in tonal music in terms of predicting a following chord. (For similar evidence on the influence of tonal functions in melody perception, see Cuddy & Lyons, 1981.)

5.2.4 Some Final Comments on Musical Syntax

From the preceding discussion, it is clear that musical syntax is structurally complex, suggesting that it can sustain a meaningful comparison with linguistic syntax. Before embarking on such a comparison, it is worth making three final points about musical syntax.

First, syntax allows music to achieve perceptual coherence based on contrast rather than on similarity. A defining feature of music perception is hearing sounds in significant relation to one another rather than as a succession of isolated events (Sloboda, 1985). The most pervasive type of relationship used in music is similarity, such as the similarity created via the repetition or variation of a melodic phrase, or via the similarity of a theme to a stock of musical themes known in a culture (cf. Gjerdingen, 2007). Syntactic organization complements similarity-based coherence by creating coherent patterns based on contrast. For example, a chord progression involves the sequential use of different chords, yet these differences articulate a coherent journey, such as from repose to tension and back to repose. Musical syntax also creates coherence via hierarchical contrasts. For example, by hearing the difference between structural and ornamental tones, a listener can extract the gist of a musical sequence and recognize

The superscript 6 refers to the fact that the standard I$^6$ chord, in other words, D-F-A in C major, is played in an "inverted" form such that the F is the lowest note. Chord inversions are common in tonal music.
its similarity to (or contrast with) other passages in the same musical idiom. As suggested by this latter example, much of the richness of tonal music comes from the way that contrast-based and similarity-based cognitive processes interact in forming perceptually coherent patterns.

Second, although the different levels of syntactic organization of music have been described independently above (e.g., scales, chords, and keys), there is a good deal of interaction between levels. For example, for an experienced listener, even one or two chords can suggest a key, and a melody of single tones can suggest an underlying chord structure. Specifying the precise nature of inter-level interactions has been the focus of various models of tonal perception. For example, in one model suggested by Bharucha (1987), tones, chords, and keys are represented as nodes in different layers of an artificial neural network, and activity at one level can propagate to other levels according to the pattern of inter-level connectivity (cf. Tillmann et al., 2000). This model has been able to account for the results of harmonic priming experiments that show how a chord's processing is influenced by its harmonic relation to preceding chords (Bigand et al., 1999). Another artificial neural network model that addresses the relations between syntactic levels involves a self-organizing map that is used to model how a sense of key develops and shifts over time as a sequence of musical chords is heard (Toivainen & Krumhansl, 2003).

Third, a very important issue about tonal syntax concerns the relationship between its acquisition during development and its application during perception. Statistical research suggests that pitch hierarchies among scale tones and chords reflect the relative frequency of different notes and chords in tonal music (Krumhansl, 1990). Thus there is reason to believe that the acquisition of tonal syntax reflects the statistics of a particular musical environment. However, once a musical syntax is acquired, it can be activated by patterns that do not themselves conform to the global statistics that helped form the syntactic knowledge. For example, Bigand et al. (2003) found that listeners reacted more quickly and accurately to misrings of the final chord of a sequence when it functioned as a 1 chord versus a IV chord (cf. section 5.2.3 above), even when the preceding chords had more instances of IV chords than 1 chords. Thus cognitive factors, namely the structural centrality of the 1 chord in a musical key, prevailed over the frequency of the IV chord in influencing perception and behavior. This provides strong evidence for a syntactic knowledge system that influences how we hear our musical world.

5.3 Formal Differences and Similarities Between Musical and Linguistic Syntax

The preceding sections provide the background for comparing linguistic and musical syntax in formal terms. Such a comparison is meaningful because language and music are rich syntactic systems that are not trivial variants of one another. Formal similarities between the systems are best appreciated in light of a clear understanding of formal differences, to which I turn first.

5.3.1 Formal Differences

Perhaps the most obvious difference between syntax in the two domains is the presence of grammatical categories in language, such as nouns, verbs, and adjectives, that have no counterparts in music. The attempts to find musical analogs for such categories is a trap that Bernstein fell into in The Unanswered Question, and is part of what Lerdahl and Jackendoff correctly labeled as "an old and largely futile game." Another example of uniquely linguistic syntactic entities are the linguistic grammatical functions that words play in sentences, in other words, subject, direct object, and indirect object (Jackendoff, 2002). Searching for direct musical equivalents of such functions is a misguided enterprise.

Beyond these differences in category identity, the hierarchical organization of grammatical categories in sequences also shows important differences in the two domains. Syntactic trees in language, such as that in Figure 5.8, convey the relationship of constituency: A determiner plus a noun is a noun phrase; a noun phrase plus a verb phrase is a sentence, and so forth. Syntactic trees in music, such as that in Figures 5.9 and 5.10, are not constituent trees. In the case of structure-elaboration trees (time-span reduction trees in GTTM), the branches of the tree join in ways that indicate which events are more structurally important. In tension-relaxation trees (prolongation-reduction trees in GTTM), the branching pattern indicates whether an event represents a tension or a relaxing movement in relation to another event.

Another difference between linguistic and musical syntax concerns long-distance dependencies. Such relations, such as between "girl" and "opened" in Figure 5.8, are ubiquitous in language and every normal listener can be assumed to perceive them (Chomsky, 1965). In contrast, the long-distance relations posited by musical syntax, such as the relations embodied in tension-relaxation trees (Figure 5.10), cannot simply be assumed to be perceived and are better viewed as hypotheses subject to empirical test, for example, using the tension-rating experiments described in section 5.2.2 (subsection on tension and resolution). Put another way, a particular sequence of notes or chords does not

7 In GTTM, there is a kind of pseudoconstituency in prolongation-reduction trees, in that nodes in a tree have a distinct syntactic identity as a progression, a weak prolongation, or a strong prolongation (Lerdahl & Jackendoff, 1983:182). However, even in these cases, the chords are not really in a constituent relationship in a linguistic sense, in that one cannot define a priori what the grammatical categories of a constituent must be.
constrain perceived dependencies to the same degree as a particular sequence of words, suggesting that words have more intricate syntactic features built into them than do notes or chords. (For example, the mental representation of a verb is thought to include its syntactic category and thematic role information, in addition to its semantic meaning; cf. Levelt, 1999.)

One final formal difference that can be mentioned concerns the role of syntactic ambiguity in the two domains. In language, syntactic ambiguity is generally eschewed by the cognitive system, which seeks to arrive at a single structural analysis for a sentence. For example, in the sentence “The fireman left the building with the large sign,” most individuals will parse the sentence so that “the large sign” is a modifier of “the building” rather than of “the fireman,” even though the sentence is structurally ambiguous. In contrast, there is much greater tolerance for syntactic ambiguity in music. Krumhansl (1992) notes that “a chord may be heard simultaneously in its multiple roles in different keys, with the effect that modulations between closely related keys are easily assimilated.” The role of a chord need never be disambiguated” (p. 199). Examples of such “pivot chords” abound in tonal music, and show that music not only tolerates syntactic ambiguity, it exploits it for aesthetic ends.

### 5.3.2 Formal Similarities: Hierarchical Structure

Having reviewed several important formal differences between musical and linguistic syntax, we can now turn to similarities. As reviewed in section 5.2.1, one similarity is the existence of multiple levels of organization. In language, there are syntactic principles that guide how basic lexical units (morphemes) are combined to form words, how words are combined to form phrases, and how phrases are combined to form sentences. In music, there are syntactic principles that govern how tones combine to form chords, how chords combine to form chord progressions, and how the resulting keys or tonal areas are regulated in terms of structural movement from one to another. In both domains, this multilayered organization allows the mind to accomplish a remarkable feat: A linear sequence of elements is perceived in terms of hierarchical relations that convey organized patterns of meaning. In language, one meaning supported by syntax is “who did what to whom,” in other words, the conceptual structure of reference and predication in sentences. In music, one meaning supported by syntax is the pattern of tension and resolution experienced as music unfolds in time.

As noted earlier, the hierarchical organization of linguistic and musical syntactic structures follows different principles. Specifically, linguistic syntactic trees embody constituent structure, whereas musical syntactic trees do not (cf. section 5.3.1). Nevertheless, some interesting parallels exist between linguistic and musical syntactic trees at an abstract level, particularly between linguistic trees and prolongation-reduction trees in GTTM. First, just as each node of a linguistic tree terminates on a linguistic grammatical category (e.g., noun, verb, preposition), each node of a prolongation-reduction tree terminates on musical grammatical category: a chord assigned to a particular harmonic function in a given key (e.g., I, IV). That is, in both cases the tree structures relate grammatical categories in a hierarchical fashion, and in both cases the same grammatical categories can be filled by different members of the same category. That is, one can have the same sentence structure with different words, and the same harmonic structure with different chords (such as chords in different inversions, or if the key is changed, an entirely different set of chords). Note that in making this comparison, there is no claim for a direct correspondence between categories in the two domains (e.g., between tonic chords and nouns).

One final similarity regarding hierarchical syntactic relations in language and music bears mention, namely the shared capacity for recursive syntactic structure. In language, phrases can be embedded within phrases of the same type, for example, in Figure 5.8 the noun phrase “the boy” is embedded within the larger noun phrase “the girl who kissed the boy.” In music, small-scale patterns of tension and relaxation can be embedded in larger tension-relaxation patterns of identical geometry but of a longer timescale (see Lerdahl & Jackendoff, 1983:207, for an example). Recursive syntactic structure has been proposed as a feature that distinguishes human language from nonhuman communication systems (Hauser et al., 2002; though see Gentner et al., 2006). If this is the case, then human music, no less than human language, is fundamentally different from animal acoustic displays such as bird song and whale song.

### 5.3.3 Formal Similarities: Logical Structure

Although the previous section dealt with hierarchical structure, this section is concerned with nonhierarchical aspects of “logical structure” of syntax in the two domains. For example, both domains recognize a distinction between “structural” and “elaborative” elements in sequences. In language, elaborative elements take the form of modifiers such as adjectives and adverbs. In tonal music, as discussed in a previous section, elaborative elements are identified on the basis of relative importance in the tonal hierarchy, together with rhythmic and motivic information. Although the means for distinguishing structure from elaboration are quite different in language and music, in both domains this conceptual distinction plays a role in organizing communicative sequences.

Another parallel in logical structure concerns grammatical functions in language and music. In the syntax of language, such functions include subject, object, and indirect object. These are logical functions that words take on with respect to other words in a sentence context, rather than being inherent properties of isolated words. The evidence that such a level of organization exists is that there are a number of grammatical principles that refer to these
functions, such as verb agreement, which requires agreement between a subject or an object and its verb (Jackendoff, 2002:149). Tonal music also has a system of grammatical functions, discussed as "harmonic functions" in section 5.2.3 above. Such functions pertain to the structural role a chord plays in a particular key. As noted in that section, the harmonic function of a chord derives from its context and its relation to other chords rather than to intrinsic properties of the chord itself. Typically, three such functions are recognized: tonic, subdominant, and dominant, prototypically instantiated by the I, IV, and V chords of a key, respectively. The same chord (e.g., C-E-G) can be a tonic chord in one key but a dominant or subdominant chord in other keys, and empirical research shows that listeners are quite sensitive to this functional difference, as discussed in section 5.2.3. (Note that the tonic function is not limited to major chords: in minor keys, the tonic function is played by a minor chord, e.g., A-C-E in A minor; Krumhansl et al., 1982.) Conversely, two distinct chords in the same key—for example, a IV chord and a II° chord—can have the same harmonic function by virtue of the way in which they are used (cf. section 5.2.3). The salient point is that a chord's harmonic function is a psychological property derived from its relation to other chords. Thus music, like language, has a system of context-dependent grammatical functions that are part of the logical structure of communicative sequences. There is, of course, no claim for any mapping between functions in the two domains, for example, between subjects in language and tonics in music.

As an aside, readers may be curious why the number of harmonic functions in music has generally been recognized as three, when there are seven distinct chord categories in any key. This theory is based on the idea that chords whose roots are separated by a musical fifth or a musical second (e.g., I and V, or I and ii) are functionally different, whereas chords whose roots are separated by a musical third (e.g., I and vi, or ii and IV) are functionally similar (Dahlhaus, 1990:58). The philosophical basis of this theory lies in the work of the 19th-century music theorist Hugo Riemann, who related the different chord functions to three different logical aspects of thought: thesis (tonic), antithesis (subdominant), and synthesis (dominant; Dahlhaus, 1990:51–52). As noted by Dahlhaus,

---

5.3.4 Formal Differences and Similarities: Summary

The above sections have confirmed that there are important differences between linguistic and musical syntax, but have also shown that these differences need not prevent the recognition and exploration of formal similarities between syntax in the two domains. The key to successful comparison is to avoid the pitfall of looking for musical analogies of linguistic syntactic entities and relations, such as nouns, verbs, and the constituent structure of linguistic syntactic trees. Once this pitfall is avoided, one can recognize interesting similarities at a more abstract level, in what one might call the "syntactic architecture" of linguistic and musical sequences. These include the existence of multiple levels of combinatorial organization, hierarchical (and recursive) structuring between elements in sequences, grammatical categories that can be filled by different physical entities, relationships of structure versus elaboration, and context-dependent grammatical functions involving interdependent relations between elements. These similarities are interesting because they suggest basic principles of syntactic organization employed by the human mind.

5.4 Neural Resources for Syntactic Integration as a Key Link

The past decade has seen the birth of a new approach to music-language syntactic relations, based on empirical research in cognitive neuroscience. A primary motivation for this work has been a debate over the extent to which linguistic syntactic operations are "modular," conducted by a cognitive subsystem dedicated to linguistic function and largely independent of other forms of brain processing (Fodor, 1983; Elman et al., 1996). Music provides an ideal case for testing this claim. As we have seen, musical and linguistic syntax are rich systems that are not trivial variants of one another. Are musical and linguistic syntax neurally independent, or is there significant overlap? Any significant overlap would serve as compelling evidence in the debate over modularity. Furthermore, exploring this overlap could provide novel insights into fundamental syntactic operations in the human brain.

Although the prospect of overlap is stimulating, evidence from neuroscience long seemed to disfavor it. Specifically, neuropsychology provided
well-documented cases of dissociations between musical and linguistic syntactic abilities. For example, individuals with normal speech and language abilities may show impaired perception of musical tonality following brain damage or due to a lifelong condition of musical tone-deafness (amusia without aphasia; Peretz, 1993b; Peretz et al., 1994; Ayotte et al., 2000; Ayotte et al., 2002). Conversely, there are persons with severe language impairments following brain damage but with spared musical syntactic abilities (aphasia without amusia; e.g., Luria et al., 1965). This double dissociation between amusia and aphasia has led to strong claims about the independence of music and language in the brain. For example, Marin and Perry (1999) state that “these cases of total dissociation are of particular interest because they decisively contradict the hypothesis that language and music share common neural substrates” (p. 665). Similarly, Peretz and colleagues have used such dissociations to argue for a highly modular view of musical tonality processing (Peretz & Coltheart, 2003; Peretz, 2006).

As we shall see, there are reasons to doubt that this view is correct. One reason concerns neuroimaging evidence from healthy individuals processing syntactic relations in music and language. This evidence suggests far more neural overlap than one would expect based on dissociations in brain-damaged cases. Another reason concerns the nature of the evidence for aphasia without amusia, as discussed in the next section. These two factors have led to a reevaluation of syntactic relations between music and language in the brain. One outcome of this reevaluation is the hypothesis that the two domains have distinct and domain-specific syntactic representations (e.g., chords vs. words), but that they share neural resources for activating and integrating these representations during syntactic processing (Pattel, 2003b). This “shared syntactic integration resource hypothesis” (SSIRH) is explained in more detail in section 5.4.3. For now, suffice it to say that the SSIRH can account for the apparent contradiction between neuropsychology and neuroimaging, and that it suggests a deep connection between musical and linguistic syntax in the brain.

The remainder of section 5.4 is divided into four parts. The first part reviews some of the classical evidence for neural dissociations between musical and linguistic syntax. The second part discusses neuroimaging research that contradicts this traditional picture. The third part discusses the use of cognitive theory to resolve the apparent paradox between neuropsychology and neuroimaging, and introduces the SSIRH. The fourth part discusses predictions of the SSIRH and how these predictions are faring in empirical research.

5.4.1 Neuropsychology and Dissociation

There is good evidence that musical syntactic deficits can exist in the absence of linguistic difficulties. An exemplary case for this is provided by the patient G.L., investigated by Peretz and colleagues (Peretz, 1993; Peretz et al., 1994). G.L. had bilateral temporal lobe damage, with infarctions on both sides of the brain due to strokes. This is a rare neurological occurrence, but is not infrequent among cases of acquired amusia. In G.L.’s case, primary auditory cortex was spared, but there was damage to rostral superior temporal gyrus, which encompasses several auditory association areas (Peretz et al., 1994; cf. Tramo et al., 1990). G.L. was a well-educated individual who had been an avid music listener, though he had no formal musical training. Ten years after his brain damage, G.L. was referred for neuropsychological testing because of persistent problems with music perception. Peretz and colleagues administered a large battery of tests to study the nature of G.L.’s musical deficits, ranging from simple pitch discrimination to melody discrimination and tests for sensitivity to tonality. G.L. could discriminate changes between single pitches and was sensitive to differences in melodic contour in short melodies. He also showed some residual sensitivity to patterns of pitch intervals (e.g., in tests involving discrimination of melodies with the same contour but different intervals). What was most striking about his case, however, was his complete absence of sensitivity to tonality. For example, G.L. was given a probe-tone task in which a few notes (which establish a musical key) were followed by a target tone. The task was to rate the how well the target fit with the preceding context (cf. Cuddy & Badertscher, 1987; and Chapter 4, section 4.2.6, for background on probe-tone studies). Normal controls showed the standard effect of tonality: Tones from the key were rated higher than out-of-key tones. G.L., in contrast, showed no such effect, and tended to base his judgments on the pitch distance between the penultimate and final tone. He also failed to show an advantage for tonal versus atonal melodies in short-term memory tasks, in contrast to controls. Additional experiments showed that his problems could not be accounted for by a general auditory memory deficit. Most importantly for the current purposes, G.L. scored in the normal range on standardized aphasia tests, showing that he had no linguistic syntactic deficit.

G.L. is one of a handful of well-documented cases of acquired amusia. Cases of musical tone-deafness or congenital amusia are much more common (cf. Chapter 4, section 4.5.2, second subsection, for background on musical tone deafness). For example, Ayotte et al. (2002) presented a group study of 11 such individuals. These were well-educated persons with no other neurological or psychiatric problems, whose tone deafness could not be attributed to lack of exposure to music (indeed, all had music lessons in childhood). When tested, all showed a variety of musical deficits. Crucially, all failed a melodic sour-note detection task, which is a simple test of musical syntactic abilities. Normal individuals enculturated in Western music find this an easy task, even if they have had no formal musical training.

Acquired and congenital amusia thus provide compelling evidence that musical syntax can be disrupted without associated linguistic syntactic deficits. What about the reverse condition? The relevant evidence comes from cases of aphasia without amusia. The most often-cited case is that of the Russian
componist Vissarion Shebalin (1902–1963). Shebalin attracted the interest of the famous neuropsychologist A. R. Luria (Luria et al., 1965), who commented that “the relationship of the two kinds of acoustic processes, namely verbal and musical, constitutes one of the most interesting problems of cortical neurology” (p. 288). Shebalin suffered two strokes in his left hemisphere, affecting the temporal and parietal regions. After the second stroke, he had severe difficulties in comprehending and producing language. Shebalin died 4 years after this second stroke, but in those few years he composed at least nine new pieces, including a symphony hailed by the Soviet composer Shostakovich as a “brilliant creative work” (p. 292).

Shebalin’s case is not unique. Tzotzis et al. (2000, Table 4) list six published cases ofaphasia without amnesia, and report a seventh case in their own paper. However, Tzotzis et al. also point out a crucial fact about these cases: They are all professional musicians. Indeed, most cases represent composers or conductors, individuals with an extraordinarily high degree of musical training and achievement. (Note that this stands in sharp contrast to case studies of amnesia without aphasia, which typically involve nonmusicians.) The question, then, is whether findings based on highly trained musicians can be generalized to ordinary individuals. There are reasons to suspect that the answer to this question is “no.” Research on neural plasticity has revealed that the brains of professional musicians differ from those of nonmusicians in a variety of ways, including increased gray matter density in specific regions of frontal cortex and increased corpus callosum size (Schlaug et al., 1995; Gaser & Schlaug, 2003). This suggests that generalizations about language-music relations in aphasia cannot be drawn on the basis of case studies of professional musicians. To conclusively show a double dissociation between amnesia and aphasia, one needs evidence of aphasia without amnesia in nonmusicians. Peretz and colleagues (2004) have argued that such cases exist, but closer examination of their cited cases reveals that these individuals suffered from a phenomenon known as “pure word deafness.” Although pure word deafness is sometimes referred to as a form of aphasia, it is in fact an auditory agnosia. An individual with pure word deafness can no longer understand spoken material but can understand and/or produce language in other modalities (i.e., writing). This is qualitatively different from true aphasia, which is a deficit of core language functions that cuts across modalities (Caplan, 1992).

The relevant point for the current discussion is that there has not been a convincing demonstration of a double dissociation between musical and linguistic syntactic abilities in ordinary individuals with brain damage. Indeed, as discussed later, new evidence from aphasia (in nonmusicians) points to an association between linguistic and musical syntactic disorders. Furthermore, the phenomenon of musical tone deafness, in which otherwise normal individuals exhibit musical syntactic deficits, is largely irrelevant to the question of music-language syntactic relations, as we shall see. Before discussing these issues, however, it is important to review some of the neuroimaging evidence that challenged the idea of separate processing for linguistic and musical syntax in the human brain.

5.4.2 Neuroimaging and Overlap

One of the first studies to compare brain responses to linguistic and musical syntactic processing in the same set of individuals was that of Patel, Gibson, et al. (1998). This study was inspired by two earlier lines of work. First, research on brain responses to harmonic incongruities in music, such as an out-of-key note at the end of a melodic sequence (Besson & Faita, 1995), had revealed that these incongruities elicited a positive-going event-related brain potential (ERP). This positive ERP stood in sharp contrast to the commonly observed negative-going ERP associated with semantic anomalies in language (the N400, which peaks about 400 ms after the onset of a semantically incongruous word, such as the word “dog” in “I take my coffee with cream and dog”; Kutas & Hillyard, 1984). The second line of research concerned brain responses to syntactic (rather than semantic) incongruities in language (Osterhout & Holcomb, 1992, 1993; Hagoort et al., 1993). This research has revealed that when a word disrupted the syntactic form of a sentence, a positive-going ERP was generated. This ERP was referred to as the P600, because it peaked about 600 ms after the onset of the incongruous word (e.g., after the onset of “was” in “The broker hoped to sell the stock was sent to jail”).

The question asked by Patel, Gibson, et al. (1998) was whether harmonically incongruous chords within chord sequences would generate P600s akin to those elicited by syntactically incongruous words in sentences. If so, then this would suggest that some aspect of syntactic processing was shared between the two domains.

Before discussing this study in more detail, it is worth making an important point about the N400 and P600. These ERPs have been most often studied in the context of incongruities (semantic or syntactic), but it is a mistake to think that these are simply “error signals” emitted by the brain due to surprise, attention switching, and so forth. Crucially, both of these ERPs can be elicited in sentences without any semantic or syntactic errors. For example, if ERPs are measured to each word of the sentences “The girl put the sweet in her mouth after the lesson.”

* The fact that the P600 peaks after the N400 does not necessarily mean that semantic operations precede syntactic ones. ERPs are an indirect measure of brain activity that are especially sensitive to locally synchronous activity patterns. Although one can infer from ERPs that processing occurs no later than a certain time, one cannot infer from ERPs how early processing begins, because it is always possible that relevant processing commences before one sees the ERP, but is not synchronous enough to be detected.
and "The girl put the sweet in her pocket after the lesson," a comparison of the ERP at "pocket" versus "mouth" reveals an N400 to the former word (Hagoort et al., 1999). This reflects the fact that "pocket" is a less semantically predictable final word than "mouth" given the context up to the final word. The N400 is thus a sensitive measure of semantic integration in language.

Similarly, the P600 can be elicited without any syntactic error (cf. Kaan et al., 2000; Gouvea et al., submitted). For example, if the word "to" is compared in the sentences "The broker hoped to sell the stock" and "The broker persuaded to sell the stock was sent to jail," a P600 is observed to "to" in the latter sentence, even though it is a perfectly grammatical sentence (Osterhout & Holcomb, 1992). This is because in the former sentence, the verb "hoped" unambiguously requires a sentential complement, so that "to" is structurally allowed. In the latter sentence, however, when the word "persuaded" is first encountered, a simple active-verb interpretation is possible and tends to be preferred (e.g., "The broker persuaded his client to sell his shares"). This interpretation does not permit the attachment of a constituent beginning with "to." As a consequence, there is some syntactic integration difficulty at "to," although it soon becomes obvious that the verb "persuaded" is actually functioning as a reduced relative clause (i.e., "The broker who was persuaded . . ."). Sentence understanding proceeds apace. Thus although frank anomalies are often used to eliciting the N400 and P600, it is important to note that this is simply a matter of expediency rather than of necessity.

Returning to our study (Parel, Gibson, et al., 1998), we constructed sentences in which a target phrase was either easy, difficult, or impossible to integrate with the preceding syntactic context, such as the following:

(5.1a) Some of the senators had promoted an old idea of justice.
(5.1b) Some of the senators endorsed promoted an old idea of justice.
(5.1c) Some of the senators endorsed the promoted an old idea of justice.

The syntactic structure of sentence 5.1b is considerably more complex than that of 5.1a, as shown in Figure 5.11 (Note that in sentence 5.1b, the verb "endorsed" has the same type of ambiguity as the verb "persuaded" in the discussion above.)

Thus in sentence 5.1b, the target phrase should be more difficult to integrate with the preceding structure than in 5.1a. In contrast to sentences 5.1a and 5.1b, which are grammatical sentences, 5.1c is ungrammatical, making the target impossible to integrate with the previous context.

We also constructed sequences of 7-12 chords in which a target chord within the middle part of the phrase was designed to vary in its ease of integration with the prior context. We based our musical design principles on previous research showing that listeners were sensitive to key distance in music. Thus the target chord was either the tonic chord of the key of the sequence, or the tonic chord of a nearby or distant key. "Nearby" and "distant" were defined using the circle of fifths for major keys: A nearby key was three counterclockwise steps away on the circle of fifths, and a distant key was five counterclockwise steps away (Figure 5.12).

For example, if the chord sequence was in the key of C major, then the target chords were: C major (C-E-G), E-flat major (E-G-Bb), or D-flat major (D-F-A). This design had the advantage that the two out-of-key chords had the same number of out-of-key notes relative to C major, so that differences in harmonic incongruity could not be attributed to differences in the number of out-of-key notes within the chords.

Two other aspects of the chord sequences bear mention. First, the target chord always occurred after a dominant (V or V7) and was thus always in

---

**Figure 5.11** Syntactic structures for the simple and complex sentences in the study of Parel, Gibson et al. (1998). Most symbols are explained in the caption of Figure 5.8. N* = noun phrase projection, O = operator, t = trace. The sentence in (B) is substantially more complex than the sentence in (A) because the verb "endorsed" functions as a reduced relative clause (i.e., "some of the senators who were endorsed . . .").
Figure 5.12 Example chord sequences from Patel, Gibson et al.'s (1998) study. The position of the target chord is indicated by the arrow. (A) The target chord is the tonic of the key of the phrase (a C major chord in this case, as the phrase is in the key of C). (B) The target chord is the tonic of a nearby key (G-flat major). (C) The target chord is the tonic of a distant key (D-flat major). Nearby and distant keys are defined as three versus five counterclockwise steps away from key of the phrase on the circle of fifths for keys (cf. Figure 5.6). Note that each stimulus has the same chord progression but differs slightly in terms of the inversion of chords before and after the target (though the two chords before and one chord after the target are held constant).

a specific harmonic context. Second, the chord sequences were composed in a popular, rather than a classical style: They sound like musical jingles rather than like traditional four-part harmony (Sound Examples 5.3a-c). Thus for both language and music, we used syntactic principles to construct sequences with varying degrees of syntactic incongruity. A single group of 15 musically trained listeners heard numerous linguistic and musical sequences like those shown above, and for each sequence, judged whether it sounded normal or structurally odd (we chose to work with musically trained listeners because we wanted to ensure sensitivity to musical syntax). ERPs to the target phrases in language were compared to ERPs to the target chords in music. The primary result of interest was that incongruities in both domains elicited P600s, and that these ERPs were statistically indistinguishable in amplitude and scalp distribution at both the moderate and strong levels of incongruity (Figure 5.13 shows ERPs to linguistic and musical targets at the strong level of incongruity.) This demonstrated that the P600 was not a signature of a language-specific syntactic process. Patel et al. suggested that the P600 may reflect domain-general structural integration processes in both domains (I return to this point in section 5.4.3, subsection “Reconciling the Paradox”).

What can be said about the underlying neural source of the P600s in this study? Both the linguistic and musical P600s were maximal over temporal/posterior regions of the brain, but it is difficult to make any conclusions about the precise location of their underlying sources on this basis. This is because

Figure 5.13 Traces show ERPs to linguistic (solid line) and musical (dashed line) syntactic incongruities from three electrodes along the middle of the head (Fz = front; Cz = vertex; Pz = back). (The schematic on the left side of the figure shows the electrode positions as if looking down on the head from above.) The ERP responses are highly similar in the vicinity of 600 ms. The continued positivity of the linguistic P600 beyond 600 ms is due to the continuing ungrammaticality of the sentence beyond this point. See Patel, Gibson et al., 1998, for details.

the ERP technique has excellent time resolution but poor spatial resolution. One thing that can be said with confidence, however, is that the generators of the P600 are unlikely to be strongly lateralized, as the ERP was symmetric across the left and right sides of the brain. (We also observed an unexpected and highly lateralized right anterior-temporal negativity elicited by out-of-key chords, which peaked at about 350 ms posttarget-onset. See Patel, Gibson et al., 1998, for discussion, and Koelsch and Mulder, 2002, for a similar finding using more naturalistic musical materials and nonmusician listeners.) Subsequent work on musical syntactic processing has supported the case for syntactic overlap between language and music by showing that musical syntactic processing activates “language” areas of the brain. Maess et al. (2001)
provided evidence from MEG that an early right anterior negativity (ERAN) associated with harmonic processing in music originates in a left frontal brain area known as Broca's area and its right hemisphere homologue (cf. Koelsch & Siebel, 2005). An fMRI study of harmonic processing (Tillmann et al., 2003; cf. Tillmann, 2005) also reported activation of these areas, and a second such study (Koelsch et al., 2002) implicated both Broca's and Wernicke's areas in musical harmonic processing (cf. Leventi & Menon, 2003; Brown et al., 2006).

These findings from neuroimaging, which point to overlap between linguistic and musical syntax in the brain, stand in sharp contrast to the evidence for music-language dissociations provided by neuropsychology (cf. section 5.4.1). Any attempt to understand the neural relationship of musical and linguistic syntax must come to terms with this paradox. One such attempt, based on cognitive theory in psycholinguistics and music cognition, was offered by Patel (2003b). An updated version of this approach is offered in the next section.

5.4.3 Using Cognitive Theory to Resolve the Paradox

Cognitive theories of language and music suggest that the mental representations of linguistic and musical syntax are quite different. For example, as discussed in section 5.3.1, one very important difference between syntax in language and music is that language has universal grammatical categories (such as nouns and verbs) and grammatical functions (subject, direct object, and indirect object) that are unique to language. Furthermore, the perception of long-distance syntactic dependencies is much more constrained by linguistic structure than by musical structure. In music, proposed hierarchical patterns (such as that in Figure 5.10) are best viewed as hypotheses subject to empirical test.

These observations suggest that the overlap in linguistic and musical syntax is not at the level of representation. Thus one way to break the paradox outlined above is to propose a conceptual distinction between syntactic representation and syntactic processing. This can be understood as the distinction between long-term structural knowledge in a domain (i.e., in associative networks that store knowledge of words and chords) and operations conducted on that knowledge for the purpose of building coherent percepts. A key idea of this approach is that some of the processes involved in syntactic comprehension rely on brain areas separate from those areas in which syntactic representations reside. Such "dual system" approaches have been proposed by several researchers concerned with the neurolinguistics of syntax. For example, Caplan and Waters (1999) have suggested that frontal areas of the brain support a special working memory system for linguistic syntactic operations, and Ullman (2001) has suggested that frontal areas contain a symbol-manipulation system for linguistic syntax. The approach taken here is a dual system approach, but does not propose that that linguistic and musical syntax share a special memory system or symbol manipulation system. Instead, a hypothesis for what is shared by linguistic and musical syntactic processing is derived from comparison of cognitive theories of syntactic processing in the two domains.

Before introducing these theories, two related points should be made. First, there are theoretical approaches to linguistic syntax that reject a separation between representation and processing, and artificial neural network ("connectionist") models in which syntactic representation and processing occur in the same network (MacDonald & Christiansen, 2002), and thus by implication in the same brain areas. Second, the theories considered below are by no means the only theories of syntactic processing in language and music. They were chosen because of their strong empirical basis and because they show a remarkable point of convergence.

Syntactic Processing in Language I: Dependency Locality Theory

Gibson's dependency locality theory (DLT; Gibson, 1998, 2000) was developed to account for differences in the perceived complexity of grammatical sentences and for preferences in the interpretation of syntactically ambiguous sentences. DLT posits that linguistic sentence comprehension involves two distinct components, each of which consume neural resources. One component is structural storage, which involves keeping track of predicted syntactic categories as a sentence is perceived in time (e.g., when a noun is encountered, a verb is predicted in order to form a complete clause). The other component is structural integration, in other words, connecting each incoming word to a prior word on which it depends in the sentence structure. A basic premise of this theory is that the cost of integration is influenced by locality: Cost increases with the distance between the new element and the site of integration. Distance is measured as the number of new "discourse referents" (nouns and verbs) since the site of integration. Thus DLT uses a linear measure of distance rather than a hierarchical one (e.g., based on counting nodes in a syntactic tree), and thus does not depend on the details of any particular phrase structure theory.

To illustrate DLT's approach, consider the relationship between the words reporter and sent in the sentences:

(5.2a) The reporter who sent the photographer to the editor hoped for a story.
(5.2b) The reporter who the photographer sent to the editor hoped for a story.

In sentence 5.2a, when "sent" is reached, integration with its dependent "reporter" is relatively easy because the words are nearly adjacent in the sentence. In sentence 5.2b, however, the integration between "sent" and "reporter" (now the object of the verb) is more difficult, because it must cross an intervening
noun phrase, "the photographer." A strength of this theory is its ability to provide numerical predictions of the processing (storage plus integration) cost at each word in a sentence. Figure 5.14 illustrates the numerical accounting system for integration cost using the example sentences given above.

The numerical predictions of DLT can be empirically tested in reading time experiments in which the amount of time spent viewing each word of a sentence on a computer screen is quantified. The assumption of such experiments is that longer reading time is a reflection of greater processing cost. DLT has been supported by empirical research on sentence processing in English and other languages (e.g., Warren & Gibson, 2002; Grodner & Gibson, 2005). The relevant aspect of the theory for the current purpose is the idea that mentally connecting distant elements requires more resources.

Syntactic Processing in Language II: Expectancy Theory

DLT provides one account of syntactic integration difficulty in language, namely, due to distance between an incoming word and its prior dependent word. A different theoretical perspective suggests that syntactic integration difficulty is associated with how well a word fits a perceivers syntactic expectations at that point. The underlying assumption of this view is that at each point during sentence comprehension, a perceivers specific expectations for upcoming syntactic categories of words (Narayanan & Jurafsky, 1998, 2002; Hale, 2001; Levy, in press; cf. Lau et al., 2006, for neural data). These expectations reflect structural analyses of the sentence currently being considered by the parsing mechanism. When a word is encountered that does not match the most favored analysis, resources must be reallocated in order to change the preferred structural interpretation. Such an explanation can account for a number of different sentence processing effects, including difficulty caused by "garden-path" sentences in which a comprehender encounters a syntactically unexpected word, such as "to" in "The broker persuaded to sell the stock was sent to jail" (cf. section 5.4.2).

The expectancy theory of syntactic processing has old roots (Marslen-Wilson, 1973; cf. Jurafsky, 2003 for a historical perspective) but has only recently begun to be systematically investigated using psycholinguistic methods. Notably, the approach can successfully account for sentence processing effects not predicted by DLT. For example, Jaeger et al. (2005, described in Levy, in press) had participants read sentences that varied in the size of an embedded relative clause (marked by brackets below):

(5.3a) The player [that the coach met at 8 o'clock] bought the house ...

(5.3b) The player [that the coach met by the river at 8 o'clock] bought the house ...
According to DLT, the verb "bought" should be harder to integrate with its prior dependent ("player") as the size of the intervening relative clause increases. In fact, precisely the opposite pattern of results was found: Reading times on "bought" became shorter as the size of the relative clause increased. Expectancy theory predicts this result, because as each additional modifying phrase occurs within the relative clause, a main verb ("bought," in this case) becomes more expected. This in turn likely reflects experience with the statistics of language, in which short relative clauses are more common than long ones (as confirmed by Jaeger et al.'s measurements of corpus statistics). For other studies supporting expectancy theory, see Koniecny (2000), and Vasishth and Lewis (2006).

Notably, however, there are cases in which DLT makes more accurate predictions than expectancy theory. For example, in the same sentences in the previous subsection, when "sent" is encountered in sentence 5.2b, its grammatical category (noun) is highly expected, yet the word is still difficult to integrate due to its distance from "reporter." Hence at the current time, it appears that DLT and expectancy theory are successful in different circumstances, meaning that work is needed to reconcile these two theories. (For one recent framework that seeks to unify distance-based effects and expectancy-based effects, see Lewis et al., 2006.) For the current purposes, however, the relevant point is that both DLT and expectancy theory posit that difficult syntactic integrations consume processing resources used in building structural representations of sentences.

Syntactic Processing in Music: Tonal Pitch Space Theory

Lerdahl's (2001) tonal pitch space (TPS) theory concerns the perception of pitch in a musical context. It builds on the empirical findings about the perceived relations between scale tones, chords, and keys outlined in section 5.2.1, and illustrated in Figures 5.3, 5.5, and 5.7. The main formalism used to represent these relations is a "basic space" organized as a hierarchy of pitch alphabets (based on Deutsch & Feroe, 1981). Figure 5.15 shows a basic-space representation of a C major chord in the context of the C major key.

As noted by Lerdahl, "Each level of the space elaborates into less stable pitch classes at the next lower level; conversely, the more stable pitch classes at one level continue on to the next upper level. The structure is asymmetrical and represents the diatonic scale and the triad directly" (p. 48). The basic space provides a mechanism for computing the psychological distance between any two musical chords in a sequence. The algorithm for computing distance involves measuring how much one has to shift a chord's representation in the basic space to transform it into another chord. The details of this algorithm are beyond the scope of this book (see Lerdahl, 2001 for details); what is important is that the basic space provides an algebraic method for computing chord distances in a manner incorporates the tripartite distances of pitch classes, chords, and keys, and yields a single distance value that can be expressed as an integer.

TPS also provides a method for deriving tree structures such as that in Figure 5.10, which serve as a hypothesis for the perceived dependencies between chords. Using the tree structure, one computes the distance of each chord from the chord to which it attaches in the tree, with the added stipulation that a chord "inherits" distances from the chords under which it is embedded. Thus each chord is associated with a numerical distance value from another chord. This distance plays an important role in predicting the perceived ebb and flow of tension in musical sequences, with the basic idea being that tension increases with tonal distance between chords. For example, when a chord is introduced from a new key area, tension increases (cf. Steinbeis et al., 2006). The numerical predictions of TPS can be compared to tension profiles produced by listeners who rate perceived tension over time in musical passages (cf. section 5.2.2, second subsection). Such experiments have provided support for TPS, and suggest that listeners do in fact hear relations between chords in a hierarchical rather than a purely sequential manner (Lerdahl & Krumhansl, 2007).
It is important to note, however, that the essential feature of TPS for the current purposes is not the tree structures it proposes. This is because one cannot simply assume that listeners hear long-distance harmonic relations in music, so that such tree structures are best viewed as hypotheses subject to empirical test, as previously mentioned (cf. section 5.3.1). Instead, the essential feature of TPS is that chord relations are perceived in terms of distances in a structured cognitive space of pitch classes, chords, and keys. Such harmonic distances apply even when chords are heard in a purely sequential way, so that each chord’s harmonic distance is computed from the immediately preceding chord (cf. section 5.2.2, second subsection). It is this notion of distance-based syntactic processing that provides a key link to language processing, as discussed in the next section.

Convergence Between Syntactic Processing in Language and Music

We have seen that words can be difficult to integrate syntactically into sentences when they are distant from their dependents or when they are syntactically unexpected. In both cases, resources are consumed as part of constructing the structural interpretation of a sentence. In DLT, distant integrations are costly because they require reactivating a prior dependent word whose activation has decayed in proportion to the distance between the words. In expectancy theory, unexpected syntactic categories are costly because they require changing the preferred structural interpretation of a sentence, which amounts to boosting the activation of a structure that previously had a low activation level. In other words, both language theories posit that difficult integrations arise from activating low-activation items.

In music, as in language, a listener is continuously involved in building a structural interpretation of a sequence, including a sense of the local key. In music, a harmonically unexpected note or chord creates a processing cost due to its tonal distance (in the sense of TPS) from the current musical context. This cost arises because the incoming note or chord has a low activation level in the associative networks that store information about chord relationships (cf. Bharucha, 1987; Tillmann et al., 2000), yet its representation must be rapidly and selectively activated in order for it to be integrated with the existing context. In other words, harmonic distance translates into processing cost due to the need to activate a low-activation item. (Note that according to this idea, harmonically unexpected notes or chords are precisely those that are harmonically distant from the local key, because listeners tend to expect chords from the local key; cf. Schmuckes, 1989; Huron, 2006.)

Having a sense of key permits tones to be perceived in terms of varying degrees of stability, which contributes strongly to the dynamic perceptual quality of music.

Overlap in the syntactic processing of language and music can thus be conceived of as overlap in the neural areas and operations that provide the resources for difficult syntactic integrations, an idea termed the “shared syntactic integration resource hypothesis” (SSIRH). According to the SSIRH, the brain networks providing the resources for syntactic integration are “resource networks” that serve to rapidly and selectively bring low-activation items in “representation networks” up to the activation threshold needed for integration to take place (Figure 5.16).

The neural location of the hypothesized overlapping resource networks for language and music is an important question that does not yet have a firm answer. One idea consistent with current research on language processing is that they are in frontal brain regions that do not themselves contain syntactic representations but that provide resources for computations in posterior regions where syntactic representations reside (Haarmann & Kolk, 1991; Kaan & Swaab, 2002). Defining the neural locus of overlap will require within-subjects comparative studies of language and music using techniques that localize brain activity, such as fMRI. For example, if independent linguistic and musical tasks are designed with two distinct levels of syntactic integration demands within

---

![Figure 5.16 Schematic diagram of the functional relationship between linguistic and musical syntactic processing. L = language, M = music. The diagram represents the hypothesis that linguistic and musical syntactic representations are stored in distinct brain networks, whereas there is overlap in the networks which provide neural resources for the activation of stored syntactic representations. Arrows indicate functional connections between networks. Note that the circles do not necessarily imply highly focal brain areas. For example, linguistic and musical representation networks could extend across a number of brain regions, or exist as functionally segregated networks within the same brain regions.](image-url)
them, one could search for brain regions that show increased activation as a function of integration cost in both language and music (a technique known as "cognitive conjunction" neuroimaging; Price & Friston, 1997). These regions would be strong candidates for the overlapping resource networks proposed by the SSIRH.

Reconciling the Paradox

One appeal of the SSIRH is that it can reconcile the apparent contradiction between neuroimaging and neuropsychology described earlier in this section. With respect to neuroimaging, the SSIRH is consistent with the findings of Patel, Gibson, et al. (1998) under the assumption that the P600 reflects syntactic integration processes that take place in posterior temporal brain regions (cf. Kaan et al., 2000). It is also consistent with localization studies that find that musical harmonic processing activates frontal language areas (Koelsch et al., 2002; Tillmann et al., 2003) under the view that these anterior loci house the resource networks that serve to activate representations in posterior regions. (It should be noted, however, that the precise localization of overlapping resource networks requires a within-subjects design comparing language and music, cf. above.)

With respect to neuropsychology, the SSIRH proposes that the reported dissociations between musical and linguistic syntactic processing in acquired amusia are due to damage to domain-specific representations of musical syntax (e.g., long-term knowledge of harmonic relations), rather than a problem with syntactic integration processes. Consistent with this idea, most such cases have been associated with damage to superior temporal gyr (Peretz, 1993; Peretz et al., 1994; Patel, Peretz, et al., 1998; Ayotte et al., 2000), which are likely to be important in the long-term representation of harmonic knowledge. The SSIRH also proposes that musico-linguistic syntactic dissociations in musical tone deafness (congenital amusia) are due to a developmental failure to form cognitive representations of musical pitch (Krumhansl, 1990). Consistent with this idea, research by Peretz and Hyde (2003) and Foxton et al. (2004) has revealed that congenital amusics have basic psychophysical deficits in pitch discrimination and in judging the direction of pitch changes. As discussed in section 4.5.2 (subsection "The Melodic Contour Deafness Hypothesis"), these problems likely prevent such individuals from forming normal cognitive representations of musical scale, chord, and key structure. Without such representations, there is no basis on which musical syntactic processes can operate. This is the reason why musical tone deafness is largely irrelevant to the study of music-language syntactic relations, as alluded to in section 5.4.1.

How can the SSIRH account for reports of aphasia without amusia, for example, a stroke that results in severe language impairment but spared musical abilities? As discussed in section 5.4.1, such reports often focus on case studies of individuals with extraordinary musical abilities, and may not be relevant to music processing in the larger population. Furthermore, most reports of aphasia without amusia are seriously out of date. For example, in Marin and Perry's (1999) review of 13 cases, many cases were from the 1800s, and the most recent was from 1987. Needless to say, none contain any systematic tests of musical syntactic processing—e.g., harmonic processing of chords—in individuals with well-defined linguistic syntactic processing deficits. In fact, it is striking that there are no modern studies of harmonic processing in aphasia, despite suggestive older research (Francés et al., 1973). This is an area that merits careful study, and is a crucial testing ground for SSIRH, as discussed in the next section.

5.4.4 Predictions of a Shared-Resource Hypothesis

A principal motivation for developing the SSIRH was to generate predictions to guide future research into the relation of linguistic and musical syntactic processing. One salient prediction regards the interaction of musical and linguistic syntactic processing. In particular, because the SSIRH proposes that linguistic and musical syntactic integration rely on common neural resources, and because syntactic processing resources are limited (Gibson, 2000), it predicts that tasks that combine linguistic and musical syntactic integration will show interference between the two. In particular, the SSIRH predicts that integrating distant harmonic elements will interfere with concurrent difficult syntactic integration in language. This idea can be tested in paradigms in which a harmonic and a linguistic sequence are presented together and the influence of harmonic structure on syntactic processing in language is studied. Several relevant studies are discussed in the following subsection.

A second prediction made by the SSIRH regards aphasia. Several language researchers have argued that syntactic comprehension deficits in Broca's aphasia can be due to disruption of processes that activate and integrate linguistic representations in posterior language areas, rather than damage to these representations per se (Kolk & Friederici, 1985; Haarmann & Kolk, 1991; Swaab, Brown, & Hagoort, 1998; Kaan & Swaab, 2002). For these aphasics, the SSIRH predicts that syntactic comprehension deficits in language will be related to harmonic processing deficits in music. Relevant evidence is discussed in the second subsection below.

Interference Between Linguistic and Musical Syntactic Processing

If music and language draw on a common pool of limited resources for syntactic processing, then one should observe interference between concurrent
difficult musical and linguistic syntactic integrations. Testing this prediction requires experiments in which music and language are presented together. Although a number of past studies have paired linguistic and harmonic manipulations, they have largely focused on the relationship between musical harmonic processing and linguistic semantic processing. What is notable about these studies is that they either find a lack of interaction in processing (Besson et al., 1998; Bonnet et al., 2001) or report an interaction that is likely due to nonspecific factors having to do with attention (Poulin-Charronnat et al., 2005). These studies are briefly reviewed next as background for studies motivated by the SSIRH.

STUDIES EXAMINING THE INTERACTION BETWEEN LINGUISTIC SEMANTICS AND MUSICAL SYNTAX Besson et al. (1998) and Bonnet et al. (2001) had participants listen to sung sentences in which the final word of the sentence was either semantically normal or anomalous, and sung on an in-key or out-of-key note. Besson et al. found that the language semantic violations gave rise to a negative-going ERP (N400), whereas the out-of-key notes gave rise to a late positive ERP, and that a simple additive model predicted the data for combined semantic/music syntactic violations quite well. Bonnet et al. had listeners either perform a single task (judge incongruity of final word or note) or a dual task (judge incongruity of both), and found that the dual task did not result in a decrease in performance compared to the single-task conditions. Thus both studies supported the independence of linguistic semantic versus musical syntactic processing.

Poulin-Charronnat et al. (2005), in contrast, did find an interaction between musical syntactic processing and linguistic semantic processing. They employed the harmonic priming paradigm, using chord sequences of the type introduced in section 5.2.3 (cf. Sound Example 5.2). Recall that these sequences are constructed so that the first six chords establish a musical context that determines the harmonic function of the last two chords. These two chords, which are physically identical in both contexts, form a V-I progression (a perfect cadence) in one context but a I-IV progression in the other. This leads to a sense of closure in the former context but not the latter.

To combine music and language, the chord sequences were sung (in four-part harmony) with one syllable per chord. The words formed sentences in which the final word was semantically either expected or unexpected. For example, the sentence “The giraffe had a very long...” could end either with “neck” (the expected ending) or “foot” (an unexpected ending). Participants were asked to listen to each sequence and decide if the last word was a real word or a nonsense word (in half the cases, the last word was in fact a nonsense word such as “sneek”). The focus of interest was on reaction times to real words. Based on standard psycholinguistic research, a faster RT was predicted for semantically expected versus unexpected words, reflecting semantic priming. The question of interest was whether this semantic priming effect would be modulated by the harmonic function of the final chord (I vs. IV, i.e., tonic vs. subdominant). Indeed, this was what was found. In particular, the RT difference between the expected and unexpected word was diminished when the final chord functioned as a subdominant chord. This result was obtained even for participants without musical training, showing that musical and linguistic processing interacted even in nonmusicians.

This study showed that a musical syntactic manipulation influenced linguistic semantic processing. However, the authors suggested that this effect might be mediated by general attentional mechanisms rather than by shared processing resources between language and music. Specifically, they proposed that the harmonic manipulation influenced linguistic semantics because the different chord endings affected the listeners’ attention to the final word/chord in different ways. For example, ending a chord sequence on a IV chord might draw attention to the music (because it sounds incomplete) and thus distract from language processing. This interpretation is supported by a study by Escoffier and Tillmann (2006), who combined harmonic-priming chord sequences with geometric visual patterns rather than with words (one pattern per chord), and showed that ending on a IV chord (vs. a I chord) slowed the speed of processing of the target pattern at the end of the sequence. Thus ending a sequence on a IV chord appears to have a nonspecific influence on the speed of responding to various kinds of stimuli (cf. Bigand et al., 2001).

The study of Poulin-Charronnat et al. raises an important issue for work on music-language syntactic interactions. Specifically, such work should control for indirect effects of music on language due to general attentional mechanisms (e.g., via the use of nonharmonic but attention-getting auditory manipulations in music).

STUDIES EXAMINING THE INTERACTION BETWEEN MUSICAL SYNTAX We now turn to studies motivated by the SSIRH that combine musical harmonic manipulations with linguistic syntactic manipulations. Currently three such studies exist: one neural study (by Koelsch et al., 2005) and two behavioral studies (Fedorenko et al., 2007; Slive et al., 2007).

Koelsch et al. (2005) conducted an ERP study in which short sentences were presented simultaneously with musical chords, with one chord per word (words were presented visually and in succession, at a rate of about 2 words per second). In some sentences, the final word created a grammatical violation with a gender disagreement. The sentences were in German, in which many nouns are marked for gender. An example of a gender violation used in this study is: Er trinkt den kühl... Bier. (He drinks the cooler... beer). This final word violates a syntactic expectancy in language (cf. section 5.4.3, subsection on expectancy theory in language). The chord sequences were designed to strongly invoke a particular key, and the final chord could be either the tonic...
chord of that key or a (harmonically unexpected) out-of-key chord from a distant key (i.e., a D-flat major chord at the end of a C major sequence). The participants (all nonmusicians) were instructed to ignore the music and simply judge if the last word of the sentence was linguistically correct.

Koelsch et al. focused on early ERP negativities elicited by syntactically incongruous words and chords. Previous research on language or music alone had shown that the linguistic syntactic incongruities were associated with a left anterior negativity (LAN), whereas the musical incongruities were associated with an early right anterior negativity (ERAN; Gunter et al., 2000; Koelsch et al., 2000; Friederici, 2002). (Note that the degree of lateralization is stronger for the LAN than for the ERAN. While the ERAN is strongest over the right anterior hemisphere, it can be clearly observed over the left anterior hemisphere.) For their combined language-music stimuli, Koelsch et al. found that when sentences ended grammatically but with an out-of-key chord, a normal ERAN was produced. Similarly, when chord sequences ended normally but were accompanied by a syntactically incongruous word, a normal LAN was produced. The question of interest was how these brain responses would interact when a sequence had simultaneous syntactic incongruities in language and music.

The key finding was that the brain responses were not simply additive. Instead, there was an interaction: The LAN to syntactically incongruous words was significantly smaller when these words were accompanied by an out-of-key chord, as if the processes underlying the LAN and ERAN were competing for similar neural resources. In a control experiment for general attentional effects, Koelsch et al. showed that the LAN was not influenced by a simple auditory oddball paradigm involving physically deviant tones on the last word in a sentence. Thus the study supports the prediction that tasks that combine linguistic and musical syntactic integration will show interference between the two processes.

Turning to the behavioral study of Fedorenko et al. (2007), these researchers combined music and language using sung sentences. Linguistic syntactic integration difficulty was manipulated via the distance between dependent words. In sentence 5.4a, below, the relative clause “who met the spy” contains only local integrations (cf. section 5.4.3, subsection on dependency locality theory). In sentence 5.4b, the relative clause “who the spy met” contains a nonlocal integration of “met” with “who,” known to be more difficult to process (e.g., King & Just, 1991).

(5.4a) The cop who met the spy wrote a book about the case.
(5.4b) The cop who the spy met wrote a book about the case.

This chord is also known as a Neapolitan 6th chord, and in longer musical contexts (i.e., in which it is not the final chord) it can function as a substitution for the subdominant chord.

The sentences were sung to melodies that did or did not contain an out-of-key note on the last word of the relative clause (underlined above). All the words in the sentences were monosyllabic, so that each word corresponded to one note. The control condition was included for an attention-getting but nonharmonically deviant musical event: a 10 dB increase in volume on the last word of the relative clause. After each sentence, participants were asked a comprehension question, and accuracy was assumed to reflect processing difficulty.

The results revealed an interaction between musical and linguistic processing: Comprehension accuracy was lower for sentences with distant versus local syntactic integrations (as expected), but crucially, this difference was larger when melodies contained an out-of-key note. The control condition (loud note) did not produce this effect: The difference between the two sentence types was of the same size as that in the conditions that did not contain an out-of-key note. These results suggest that some aspect of structural integration in language and music relies on shared processing resources.

The final study described here, by Sleev et al. (2007), manipulated linguistic syntactic integration difficulty via structural expectancy (cf. section 5.4.3, subsection on expectancy theory in language), and also directly compared the influence of musical harmonic manipulations on linguistic syntactic versus semantic processing. In this study, participants read sentences phrase by phrase on a computer screen. They controlled the timing of phrases by pushing a button to get the next phrase. In such studies, the amount of time spent viewing a phrase is assumed to reflect the amount of processing difficulty associated with that phrase. This “self-paced reading” paradigm has been much used in psycholinguistic research. The novel aspect of Sleev et al.’s study was that each phrase was accompanied by a chord so that the entire sentence made a coherent, Bach-style chord progression.

The sentences contained either a linguistic syntactic or semantic manipulation. In the syntactic manipulation, sentences like 5.5a included either a full or reduced sentence complement clause, achieved by including or omitting the word “that.” In sentence 5.5a, for example, omitting “that” results in the reduced complement clause “the hypothesis was being studied in his lab.” (Note: the vertical slashes in 5.5a and 5.5b below indicate the individual phrases used in the self-paced reading experiment.) In this case, when readers first encounter the phrase “‘the hypothesis’,” they tend to interpret it as the direct object of “confirmed,” which causes syntactic integration difficulty when “was” is encountered, as this signals that “the hypothesis” is actually the subject of an embedded clause. In other words, the omission of “that” creates a “garden path” sentence with localized processing difficulty (on “was”) due to violation of a syntactic expectancy. In the semantic manipulation, sentences like 5.5b included either a semantically consistent or anomalous word, thereby confirming or violating a semantic expectancy. The chord played during the critical word (underlined below) was either harmonically in-key or out-of-key. (Out-of-key
chords were drawn from keys 3–5 steps away on the circle of fifths from the key of the phrase.) Because out-of-key chords are harmonically unexpected, the experiment crossed syntactic or semantic expectancy in language with harmonic expectancy in music. The dependent variable of interest was the reading time for the critical word.

(5.5a) The scientist I wearing I thick glasses I confirmed (that) I the hypothesis I was I being I studied I in his lab.

(5.5b) The boss I warned I the mailman I to watch I for angry I dogsbites I when I delivering I the mail.

The main finding was a significant three-way interaction between linguistic manipulation type (syntactic or semantic), linguistic expectancy, and musical expectancy. That is, syntactically and semantically unexpected words were read more slowly than their expected counterparts; a simultaneous out-of-key chord caused substantial additional slowdown for syntactically unexpected words, but not for semantically unexpected words. Thus, processing a harmonically unexpected chord interfered with the processing of syntactic, but not semantic, relations in language. Once again, these results support the claim that neural resources are shared between linguistic and musical syntactic processing.

Taken together, the three studies reviewed above point to shared neural resources underlying linguistic and musical syntactic processing. They also suggest that studies examining concurrent processing of language and music, which have been relatively rare to date, are a promising area for exploring issues of modularity in both domains.

Musical Syntactic Deficits in Aphasia

Remarkably, there has been virtually no work on musical syntactic processing in aphasia in modern cognitive neuroscience. This is particularly striking because an early study by Francès et al. (1973) suggested that aphasic individuals with linguistic comprehension disorders also have a deficit in the perception of musical tonality. The researchers studied a large group of aphasics and had them judge whether two short, isochronous melodies were the same or different. The melodies were either tonal or atonal. Under these circumstances, normal participants (even those with no musical training) show superior performance on the tonal stimuli. Aphasics failed to show this tonal superiority effect, leading the authors to suggest that the perception of tonality “seems to engage pre-established circuits existing in the language area” (p. 133).

This idea has lain fallow for decades, with no further studies of tonality perception in aphasia. Why might this be? Good tools for testing linguistic comprehension in aphasia and for probing the perception of tonal relations have long been available, yet no one has attempted to replicate or extend these results. This is made even more puzzling by the fact that the findings of Francès et al. were somewhat clouded by methodological issues, and naturally called for further work (cf. Peretz, 1993). It is likely that the absence of research on this topic reflects the emphasis on dissociations between aphasia and amusia (cf. section 5.4.1). However, given the caveats about such dissociations raised in section 5.4.1, and the predictions of SSIRH, it is clearly time to revisit this issue.

Patel, Iverson, Wassenaar, and Hagoort (in press) recently examined musical and linguistic syntactic processing in a population of 12 Broca’s aphasics (none of whom had been a professional musician). Broca’s aphasia is a type of aphasia in which individuals have marked difficulty with sentence production, though their speech comprehension often seems quite good. In fact, careful testing often reveals linguistic syntactic comprehension deficits. To check whether the aphasics we studied had such deficits, we employed a standard psycholinguistic test for syntactic comprehension. This “sentence-picture matching task” involves listening to one sentence at a time and then pointing to the corresponding picture on a sheet with four different pictures. Sentences varied across five levels of syntactic complexity. For example, a sentence with an intermediate level of complexity (level 3) was the passive structure: “The girl on the chair is greeted by the man” (Figure 5.17).

Figure 5.17 Example panel from the sentence-picture matching task for the sentence: “The girl on the chair is greeted by the man.”
Determining who did what to whom in such sentences relies on syntactic information (e.g., simple word-order heuristics such as "first noun = agent" do not work). The aphasics performed significantly worse than controls on this test, which established that they did indeed have a syntactic comprehension deficit in language. They were therefore an appropriate population for studying relations between linguistic and musical syntactic deficits.

To test music and language in a comparable fashion, we had the aphasics (and matched controls) perform acceptability judgments on musical and linguistic sequences. The linguistic sequences were sentences (n = 120); Half contained either a syntactic or a semantic error. For example, the sentence “The sailors call for the captain and demands a fine bottle of rum” contains a syntactic agreement error, whereas “Anne scratched her name with her tomato on the wooden door” is semantically anomalous. We tested both syntax and semantics in order to determine if musical syntactic abilities were specifically related to linguistic syntax. The musical sequences were chord sequences (n = 60); Half contained an out-of-key chord, violating the musical syntax (harmony) of the phrase. The musical task was thus comparable to a sour-note detection task, though it used chords instead of a melody of single tones. (The chord sequences were taken from Patel, Gibson, et al.'s 1998 ERP study, and represented the "in-key" and "distant-key" conditions—cf. section 5.4.2 for background on these stimuli, and Figure 5.12 for an example.) We also had the aphasics and controls do an experiment involving same/different discrimination of short melodies, to check if they had any auditory short-term memory problems for musical material.

All aphasics had left-hemisphere lesions, though the locations were variable and did not always include Broca's area. Such variability is well known from studies of Broca's aphasia (Willems & Poeck, 1993; Caplan et al., 1996) and precluded us from addressing issues of localization. We focused instead on cognitive relations between music and language based on performance on tasks in both domains.

Two aphasics and one control performed poorly on the melodic same/different task, and were excluded from further analysis; the remaining aphasics and controls did not differ in their performance on the melodic task, indicating that the groups were matched on basic perception of tone sequences. Turning to the main results, the primary finding of interest was that the aphasics performed significantly worse than controls on detecting harmonic anomalies in chord sequences, indicating a deficit in the processing of musical tonality (Figure 5.18).

They also showed a severe deficit on the linguistic syntactic task, and an impairment on the linguistic semantic task, though this just escaped statistical significance. Figure 5.19 shows the data in a different way, permitting the performance of individuals on the music task to be compared to their performance on the two language tasks.

As can be seen, there is a good deal of overlap between aphasics and controls on the music task, suggesting that linguistic agrammatism was associated with a relatively mild impairment of tonality perception in this group. Indeed, the fact that most aphasics score within the normal range on the music syntax test raises a question. Is the observed aphasic group deficit in tonality perception simply due to a few individuals with lesions that affect separate brain areas involved in language and music (and hence who score poorly in both domains)? Or is the lower performance of the aphasics as a group indicative of some systematic degradation in their perception of tonality, related to linguistic agrammatism? One way to address this question is to look at the correlation between performance on the music task and the language syntax task. For the aphasics, the simple correlation was not significant, but interestingly, when the controls were included in the correlation (via a multiple regression analysis), performance on the music syntax task was a significant predictor of performance on the language syntactic task. This points to some process common to language and music syntax that operates in both the controls and the aphasics, though at a degraded level in the aphasics. Notably, when the same type of multiple regression analysis was conducted on music syntax versus language semantics, performance on the music task did not predict linguistic performance. Hence the putative shared process appears to link music syntax to language syntax rather than to language semantics.

Although the above study employed explicit judgments of tonality, it is also important to test musical syntactic abilities using implicit tasks. This is because research by Tillmann (2005) has shown that individuals with music syntactic
deficits in explicit tasks can nevertheless show implicit access to musical syntactic knowledge. The task we used to tap implicit syntactic abilities is harmonic priming (Bharucha & Stoeckig, 1986). Harmonic priming is a well-studied paradigm in music cognition that tests the influence of a preceding harmonic context on the processing of a target chord (cf. sections 5.2.3 and 5.2.4). Much research has shown that a target chord is processed more rapidly and accurately if it is harmonically close to (vs. distant from) the tonal center created by the prime (Bigand & Pineau 1997; Tillmann et al., 1998; Bigand et al., 1999; Justus & Bharucha, 2001; Tillmann & Bigand, 2001). Importantly, this advantage is due not simply to the psychoacoustic similarity of context and target, but to their distance in a structured cognitive space of chords and keys (Bharucha & Stoeckig, 1987; Tekman & Bharucha, 1998; Bigand et al., 2003). The harmonic priming effect thus indicates implicit knowledge of syntactic conventions in tonal music, and has been repeatedly demonstrated in nonmusician listeners in Western cultures (e.g., Bigand et al., 2003).

Patel, Iversen, and colleagues (submitted) studied harmonic priming in Broca's aphasia using a second group of 9 Broca's aphasics (cf. Patel, 2005). (As in the first study, we first established that the aphasics had a syntactic comprehension deficit using the sentence-picture matching task. As in that study, all aphasics had left hemisphere lesions, but these did not always include Broca's area.) We used the original two-chord version of the harmonic priming task, with a single chord serving as the prime (Bharucha & Stoeckig, 1986). Prime and target were 1 s long each, separated by 50 ms. This places minimal demands on attention and memory and is thus suitable for use with aphasics. The harmonic distance between prime and target was regulated by the circle of fifths for musical keys: Harmonically close versus distant targets were two versus four steps clockwise steps away from the prime on the circle, respectively. This directly pits conventional harmonic distance against psychoacoustic similarity, because the distant target shares a common tone with the prime (Tekman & Bharucha, 1998; Figure 5.20).

The participants' task was to judge whether the second chord was tuned or mistuned (on 50% of the trials, it was mistuned by flattening one note in the chord). The main focus of interest, however, was the reaction time (RT) to

---

**Figure 5.19** Relationship between performance on musical and linguistic tasks for aphasics (black dots) and controls (open circles). Separate best-fitting regression lines are shown for aphasics and controls. (A) shows relations between performance on the music task and the language syntax task, and (B) shows relations between performance on the music task and the language semantics task.

**Figure 5.20** Example of prime and target chords for the harmonic priming task. All chords were major chords, being the principal chord of a key from the circle of fifths. In this case, the prime is a C major chord. The close target is a D major chord, and the distant target is an E major chord.
well-tuned targets as a function of their harmonic distance from the prime. A faster RT to close versus distant chords is evidence of harmonic priming. (Prior to doing the priming study, the aphasics completed two short experiments that showed that they could discriminate tuned from mistuned chords and did not have auditory short-term memory defects.)

The results were clear. Controls showed normal harmonic priming, with faster reaction times to harmonically close versus distant well-tuned targets. Aphasics, however, failed to show a priming effect, and even showed a nonsignificant trend to be faster on distant targets, suggestive of responses driven by psychoacoustic similarity rather than by harmonic knowledge (Figure 5.21).

Thus aphasics with syntactic comprehension problems in language seem to have problems activating the implicit knowledge of harmonic relations that Western nonmusicians normally exhibit. Importantly, this deficit is not a generalized consequence of brain damage, because there are cases of individuals with bilateral cortical lesions who show normal harmonic priming (Tramo et al., 1990; Tillmann, 2003).

Together, these two aphasias studies point to a connection between linguistic and musical syntactic processing. The results are consistent with the SSIRH, which suggests that music and language share neural resources for activating domain-specific representations during syntactic processing. A deficiency in these resources appears to influence both music and language. From the standpoint of neurolinguistics, this supports a “processing view” of syntactic disorders in aphasia, that is, a general problem activating stored syntactic representations (e.g., verbs together with their lexical category and thematic role information), rather than a disruption of these representations (Kolk & Friederici, 1985; Haarmann & Kolk, 1991; Kolk, 1998; Swaab et al., 1998; Kaan & Swaab, 2002). From the standpoint of music cognition, the results are notable in suggesting that left hemisphere language circuits play a role in musical syntactic processing in nonprofessional musicians. Future work seeking to determine the importance of particular left hemisphere circuits to music processing will need to employ aphasias with more tightly controlled lesion profiles. It will be interesting to know if aphasics with a more uniform lesion profile (e.g., agrammatic aphasics with left frontal brain damage) would show even stronger links between performance on language and music tasks than found in the current experiments, which employed aphasias with rather variable lesion profiles.

From a broader perspective, the above results indicate that it is time to reawaken the study of music syntactic processing in aphasia, a topic that has experienced a 30-year hiatus since the pioneering work of Francères et al. (1973). Parallel studies of music and language offer a novel way to explore the nature of aphasic processing deficits (cf. Racette et al., 2006). Such research has clinical implications, and raises the intriguing possibility that it may ultimately be possible to model linguistic and musical syntactic deficits in aphasia in a common computational framework (cf. Tillmann et al., 2000; McNellis & Blumstein, 2001).

5.5 Conclusion

Roughly 30 years ago, Leonard Bernstein’s provocative lectures at Harvard sparked interest in cognitive comparisons between musical and linguistic syntax. Although his own ideas on the subject have not stood the test of time, his intuition of an important link is now being supported by modern research in cognitive neuroscience. This research suggests that although musical and linguistic syntax have distinct and domain-specific syntactic representations, there is overlap in the neural resources that serve to activate and integrate these representations during syntactic processing (the “shared syntactic integration resource hypothesis” [SSIRH]). Exploring this overlap is exciting because it provides a novel way to illuminate the neural foundations of syntax in both domains.

This chapter has attempted to illustrate that comparative research on musical and linguistic syntax should be grounded in a solid understanding of the
important differences between the two systems. Understanding these differences need not deter comparative research. On the contrary, they should guide such research past the pitfalls that trapped earlier thinkers (including Bernstein). Once these traps are avoided, a large and fertile field for investigation is reached, a field that has only just begun to be explored. Such explorations are likely to be the most fruitful when they are hypothesis-driven and based on empirically grounded cognitive theory in the two domains. If this can be achieved, then the power of the comparative method in biology can be brought to bear on the human mind's remarkable capacity for syntax.

Chapter 6

Meaning

6.1 Introduction 300
6.1.1 Music and Translation 301
6.1.2 What Does One Mean by "Meaning"? 303

6.2 A Brief Taxonomy of Musical Meaning 305
6.2.1 The Structural Interconnection of Musical Elements 305
6.2.2 The Expression of Emotion 309
6.2.3 The Experience of Emotion 315
6.2.4 Motion 319
6.2.5 Tone Painting 320
6.2.6 Musical Topics 321
6.2.7 Social Associations 322
6.2.8 Imagery and Narrative 323
6.2.9 Association With Life Experience 324
6.2.10 Creating or Transforming the Self 324
6.2.11 Musical Structure and Cultural Concepts 325

6.3 Linguistic Meaning in Relation to Music 327
6.3.1 Music and Semantics 327
   The Semantics of Leitmotifs 328
   Neural Evidence That Music Can Evoke Semantic Concepts 331
6.3.2 Music and Pragmatics 335
   Cognitive Aspects of Discourse Coherence 337
   Neural Aspects of Discourse Coherence 340

6.4 Interlude: Linguistic and Musical Meaning in Song 342
6.5 The Expression and Appraisal of Emotion as a Key Link 344
6.5.1 Acoustic Cues to Emotion in Speech and Music 345
6.5.2 Neural Aspects of Auditory Affect Perception 347
6.5.3 Cross-Domain Influences 348
6.5.4 Emotion in Speech and Music: Future Directions 349

6.6 Conclusion 350