7 Analysis and cognition of musical structure

Many a curious child has taken apart a mechanical toy to see what parts are inside it, how they interlock, and what makes them work. Listeners often do the same with pieces of music. Careful listening involves determining how a piece of music is constructed, how the parts fit together, and what makes them work as a whole. It is easy to overlook the complexity of this process. Analyzing a thing in motion, with all its dynamic complexities, poses a formidable task. It takes a trained eye to identify patterns in the movements of a gazelle bounding through a forest, or a flock of birds in flight. Similar challenges are true of listening as the listener hears a succession of tones that each only exist for an instant in time, as a composition is unfolding. Furthermore, the systems that regulate perception and memory are limited with respect to the amount and type of information that can be processed (Snyder, 2000).

Although 'analysis' is often equated with musical expertise, it is important to note that any listener, to some degree, can take part in 'analytical' listening. Although not all listeners are musically trained, most at least grasp sufficient basic patterns or relationships to feel an overall sense of coherence in a new piece of music, to form general expectations about what might come next, to feel somewhat puzzled or frustrated when the music strays dramatically from what they expected, to sense when a phrase (or a whole piece) may be nearing the end, and whether the phrase or composition comes to a satisfying close or not. All the while, they are also organizing the incoming tones into some sort of meaningful array and into coherent units, such as melody and accompaniment, and separate phrases. This process may occur implicitly (see chapter 5), without conscious awareness, but implicit or not our understanding of musical structure is integral to our experience of music.

Musicians – and particularly musicologists (music theorists) – develop the listening skills and vocabulary necessary to develop such analyses consciously and explicitly. Sometimes this process involves an examination of the written musical score, note by note, to more fully understand the interlocking parts – identifying the themes and motifs and their variations, the sources of unity, the harmonic underpinnings of a piece of music. This is the process that is typically referred to as 'analysis' of music. Importantly, this explicit understanding of musical structure can lead to theories concerning how listeners come to comprehend musical structure while listening.

This chapter provides a general introduction to two of the most influential analytical approaches that have shaped our understanding of music analysis and cognition. These approaches are particularly interesting for the psychology of music, moreover, because they are predicated on an understanding of the psychological bases of listening, composing, and performing. It is worth noting that an analytical theory of music is not always a theory of how the listener or performer conceptualizes music. Traditionally, music analysis has been used to elucidate the structure of music, which may or may not be part of anyone's experience of that music. More recently an increasing number of musicologists, led by the example of Leonard Meyer (1956), have become interested in the basic perceptual and memory processes that contribute to musical comprehension.

Lerdahl and Jackendoff's (1983) *Generative Theory of Tonal Music* concerns how a listener organizes aspects of the music that are hierarchically structured (such as motifs nested within phrases nested within sections of a composition, or beats and measures set within a metrical structure). Similar to the earlier approach of Heinrich Schenker (1935), Lerdahl and Jackendoff's theory seeks to 'strip away' all the nonessential notes to uncover the very core of a musical piece. Their theory identifies four components of music that are hierarchically structured, and identifies a system of 'rules' that listeners intuitively apply in order to understand the piece. Even though listening itself is a dynamic activity, their theory is concerned only with the 'final state' of the listener's understanding (p. 4). It seeks to capture the observer's lasting impression after the gazelle has disappeared into the woods, after the flock of birds has vanished into the horizon.

Narmour's (1990, 1992) *Implication-Realization model* departs from the hierarchical approach of Lerdahl and Jackendoff and focuses instead on the musical surface. This theory, which finds its roots in Leonard Meyer (1956, discussed more in chapter 13), is concerned with how a listener's expectations are formed, for instance by what two successive tones imply may come next, and how this implication is realized, to explain the analysis and cognition of basic melodic structures. As its name implies, this model seeks to capture both the prospective and retrospective aspects of perceiving, structuring, and comprehending melody. Thus it seeks to capture the listener's strategies in melodic cognition while the piece is in motion, along with listeners' expectations and some retrospective analysis.

Due to the technical nature of the topic, it should be noted that the present chapter is different from all others, as it is the only chapter that assumes previous knowledge of music theory on the part of the reader. Whereas every effort is made to explain musical terms and concepts in all other chapters, it would be an unwieldy task to do so in the present discussion.

The Generative Theory of Tonal Music

A central assumption of mainstream cognitive psychology is that people apply internalized rule systems to guide their interpretation of events. When it comes to the interpretation of organized sequences, like music or language, many have claimed that these rules take the form of implicitly learned *grammars*. Probably the most famous proponent of this idea as applied to language was Noam Chomsky. Chomsky's famous generative model of the cognitive foundations of language competence (Chomsky, 1966) asks us to imagine that the sentence forms we hear or read are associated with abstract rule systems that can be used to generate any possible grammatically correct sentence (including transformations of the sentence you just heard or read).

Chomsky's view was fundamental in the formation of one of the theories we focus on here, the Generative Theory of Tonal Music advanced by Fred Lerdahl and Ray Jackendoff (1983). For example, summing up the basic principles of cognitive psychology of musical composition, Lerdahl (1988, p. 233) suggests that in any process of composition a 'compositional grammar' is involved, which could be represented by systems of rules. These rules may express the knowledge and skills of a composer who creates a sequence of musical events, or of a listener trying to make sense of the same sequence. As Lerdahl (1988) explains, 'the listening grammar then generates the mental representation that comprises the "heard structure" of the piece' (p. 234). In so doing, the listener makes use of whatever body of musical knowledge he or she has so far acquired.

Background: Schenkerian analysis

Before describing the Generative Theory of Tonal Music in more detail it is important to describe a historically antecedent musical theory, that of music theorist Heinrich Schenker. Schenker believed that the deep structures of music could be revealed by a method of 'reduction' (which we describe shortly) to bring out the basic structures which grounded the music, conceived as an elaboration of these primitive forms (Schenker, 1954, p. 29). Though their work is thought to have been independently developed, there is a common spirit in Chomsky's and Schenker's work (see Sloboda, 1985, pp. 11–23). Both are interested in deriving abstract structures that underlie complex sequences.

According to Schenker, every recognizable melody has a deep structure or *ursatz*. The *ursatz* is complex, having a melodic line (the *urlinie*) and a bass line (the *bassbrechung*). Schenker proposed that the *urlinie* is built from the descending sequence of third, second, and tonic (E D C in C major), whereas the *bassbrechung* is built on the tonic, dominant, tonic sequence (C G C in C major). Schenker suggested that these deep structures are natural and universal (see chapter 15 for further discussion of universality). As such, the listener's capability of melody recognition may rely on the generic *ursatz*. The tonic triad (first, third, and fifth tones) pattern can also be in the repertoire of *ursatzen* since it is clearly the basis on which the whole scheme is built. If we take the suggestion that the 'circle of fifths' (figure 5.4) is the underlying structure stemming from physics and physiology and on which the tonic triad itself rests, the simple 'tonic/dominant' pair would seem to be the deepest *musical* pattern of all. Melody then must consist in departures from and returns to the *ursatz* defined by the basic Schenkerian structures. This is the basis of anticipation, the most important psychological state for the listener's experience of music.

The reduction of a piece of music to reveal the Schenkerian deep structure involves several steps by which the deep structure is revealed as we strip away or 'reduce' the intervening notes which constitute elaborations of base structures. By proposing that all tonal music compositions are generated from a simple fundamental structure (to which they can subsequently be 'reduced'), Schenker provided a way of isolating the distinctive features of a composition that set it apart from features that characterize all tonal music. Figure 7.1 shows an example of a pitch reduction of the Bach chorale *O Haupt voll Blut und Wunden* in the spirit of Schenker but taken from Lerdahl and Jackendoff (1983, p. 108). We should note, however, that the primary goal of Schenker's theory was to identify a universal *ursatz*, from which all melodies could be generated, a kind of 'top-down' approach. The use of Schenkerian theory to create 'reductions,' as we have described, is closer to the goal of most theories in music psychology and constitutes a kind of 'bottom-up' approach (Lerdahl, 2009).

An overview of GTTM

Similar to the ideas of Schenker and Chomsky, music composer and theorist Fred Lerdahl and linguist Ray Jackendoff (1983) set out to describe principles through which an underlying 'structural description' of music can be derived in their *Generative Theory of Tonal Music* (hereafter referred to as GTTM). Put simply, their theory aims to identify the specific rules by which a listener generates a 'heard structure' from the 'musical surface.' Lerdahl and Jackendoff (1983–84)[1] define *musical surface* broadly as 'the physical signal of a piece when it is played' (p. 229) and *heard structure* as 'all the structure a listener unconsciously infers when he [or she] listens to and understands a piece, above and beyond the data of the physical signal' (pp. 229–230). In the broadest sense, therefore, GTTM offers a 'musical grammar' in the form of an elaborate set of rules that explain how listeners come to hear sounds such as pitches of different durations and loudness, as organized music. The applicability of the theory is restricted to tonal music, and the musical grammar focuses only on hierarchically organized aspects of music. The theory consists of four components, each devoted to one of four types of hierarchical structure (referred to as *Grouping Structure, Metrical Structure, Time-span Reduction, and Prolongational Reduction*), each subject to a set of rules.

Figure 7.1 Schenkerian-style reduction of the Bach chorale *O Haupt voll Blut und Wunden* (Lerdahl and Jackendoff, 1983).

Source: Lerdahl, Fred, and Ray S. Jackendoff, *A generative theory of tonal music*, figure 5.4 page 108 © 1983 Massachusetts Institute of Technology, by permission of The MIT Press.

GTTM's reductions (and in particular, prolongational reductions) are broadly similar to Schenker's in that an elaborate musical 'surface' is 'reduced' to its core constituents. However, the two theories differ in their aims. Lerdahl and Jackendoff view GTTM as a 'psychological' theory that aims 'to find principles of music cognition' and Schenker's as an 'aesthetic' theory whose purpose is 'to illuminate musical masterpieces' (1983–84, pp. 248–249). The theories also differ in their focus. For instance, GTTM focuses far more on rhythmic elements than does Schenker's approach.

A musical grammar must identify the kinds of structures that are 'preferred' or 'not preferred,' and 'acceptable' or 'unacceptable,' within a given rule system. Thus, the rules associated with each of the components occupy an important role in GTTM. In Lerdahl and Jackendoff's (1983) words, 'each rule of musical grammar is intended to express a generalization about the organization that the listener attributes to the music he [or she] hears' (preface, p. x). In other words, their rule systems are designed to suggest a kind of 'structural description' that the listener forms while listening to music. There are two primary rule categories they propose. *Well-formedness rules* characterize a fairly fixed set of rules that distinguish 'acceptable' from 'unacceptable' structures. By contrast, *preference rules*, which are more flexible, characterize the kinds of descriptions that a listener will gravitate to when more than one description is possible. In the original theory there was some ambiguity concerning which preference rules would 'win' in specific situations. More recently, Temperley (2001) has developed an extension of the preference rule approach that features a quantitative rule system that can be used to determine which combination of preference rules best matches the interpretation of a listener.

A notational system is needed to visually represent the 'heard structure' of the music, and the general features of this symbol system are also shown in two reductions of Bach's *C Major Prelude* in Figures 7.2 and 7.3. (These figures are only provided as an illustration of the basic features of GTTM's notation system, and a detailed analysis of this composition can be found in Lerdahl and Jackendoff's [1983] book on pages 260–264.)

Grouping Structure 'describes the listener's segmentation of the music into units of various sizes' (1983–84, p. 231), each nested within the other (thus hierarchically structured) such as motifs within short phrases, within longer phrases, within sections of music, and finally within the whole composition. In GTTM's notational style, these are represented by a series of slurs beneath the musical notation, with the broadest grouping (in which all smaller groups are nested) shown at the lowest layer as shown in Figure 7.2. According to GTTM's formal grammar, listeners group the musical surface into segments or chunks consistent with two sets of rules ('well-formedness' and 'preference' rules) that are specified for each of the components.

Here, for example, are five 'well-formedness' rules for groupings of notes (1983, pp. 37–39).

(1) Any contiguous sequence of pitch-events, drum beats, [etc.] can constitute a group, and only contiguous sequences can constitute a group.
(2) A piece [of music] constitutes a group.
(3) A group may contain smaller groups.
(4) If a larger group contains a smaller group, all the elements of the smaller group must be elements of the larger group.
(5) If a larger group contains a smaller group, the larger group must be exhaustively partitioned into smaller groups.

Figure 7.2 Time-span reduction of Bach's *C Major Prelude* (Lerdahl and Jackendoff, 1983).

Source: Lerdahl, Fred, and Ray S. Jackendoff, *A generative theory of tonal music*, figure 10.10 page 262 © 1983 Massachusetts Institute of Technology, by permission of The MIT Press.

Though these principles may sound abstract, a simple example shows that they are in fact highly intuitive. Imagine the first line of the tune 'Twinkle Twinkle Little Star':

Twinkle twinkle little star, how I wonder what you are.

This line when sung may be considered to be a single musical phrase, and thus a group. At the same time, we can hear two smaller groups within this phrase, one functioning as a kind of musical 'question' ('Twinkle

twinkle little star') and the other functioning as an 'answer' ('how I wonder what you are'). Note that these two smaller groups are bounded by the larger group. It simply does not make sense conceptually to have a smaller group (a subphrase) cross a boundary that is formed by a larger group. Furthermore, this single phrase is one of several that together make up the entire tune, and the two subphrases discussed above can likewise be carved into yet smaller groups, the smallest of which would comprise individual notes.

Preference rules are greater in number and more complex than well-formedness rules, and so we will not enumerate all grouping preference rules here but instead focus on two examples. Generally speaking, grouping preference rules suggest where a listener will typically, but not always, hear a group boundary. One such rule is the rule of *proximity*. As in Gestalt psychology (see chapter 5) this rule suggests that a relatively long time-span between notes will often be heard as a group boundary. Going back to our example, the lengthened note associated with the first singing of 'star,' in comparison to the shorter notes associated with 'twinkle, twinkle little . . .,' would lead a listener to (most likely) hear a group boundary where the lengthening has occurred.

But proximity clearly is not sufficient. One could extend the rule of proximity to the point that every single note is its own group; after all there is a 'boundary' after each note in a melody (disregarding chords). To guard against such counter-intuitive 'preferences' Lerdahl and Jackendoff (1983) also propose a rule stating that listeners prefer groups that are not too short in length (certainly not just one event).

Metrical Structure refers to 'the hierarchy of beats that [the listener] attributes to the music' (Lerdahl & Jackendoff, 1983, p. 231), for instance organizing the heard structure into strong and weak metric beats. These form a hierarchy as they may be organized into several increasing broader layers or levels (for instance if in compound time: beats at the individual eighth note [quaver] level, dotted quarter note [dotted crotchet] level, and dotted half note [dotted minim] level). This metrical hierarchy is represented by two or more rows of dotted markings (with the smallest units [i.e., beats] at the topmost level), which appear above the grouping hierarchy in GTTM's notation system as seen in Figure 7.2. (See chapter 6 for an extended discussion of metrical structure and what constitutes a 'beat.')

The well-formedness rules for identifying metrical structure include (Lerdahl & Jackendoff, 1983, p. 69):

(1) Every attack point must be associated with a beat at the smallest level of metrical structure.
(2) Every beat at a given level must also be a beat at all smaller levels.
(3) At each metrical level, strong beats are spaced either two or three beats apart.[2]
(4) Each metrical level must consist of equally spaced beats.

Again, these seemingly abstract principles lead to more intuitive results, in that they lead to metrical structures that are both regular (with respect to spacing of beats) and alternating (with respect to the prominence of beats). For the listener, the result is that meters frame music with respect to alternately stronger and weaker points in time, an effect that we discuss in more detail in chapter 6. For instance, the melody 'Twinkle twinkle little star' yields the strong sense of alternating strong and weak beats on different syllables, with the first syllable of each 'twinkle' being associated with a strong beat.

Preference rules for meter address the thorny issue of which events are interpreted as strong versus weak as well as the link between these patterns of accenting on the musical surface and the underlying meter that a listener interprets from that music. The link from musical surface to meter is not always straightforward, as we discussed in chapter 6. As with grouping, GTTM suggests many metrical preference rules from which we select two examples. Perhaps the simplest preference rule is referred to as the *event* rule. It stipulates that strong metrical accents should coincide with the onset of notes. Thus a listener is unlikely to hear a strong beat in the middle of the sustained note for 'star.' Another, complementary rule suggests that stronger beats are associated with the onset of relatively longer notes. Thus, one is unlikely to hear a strong beat in the middle of 'star' and is instead likely to hear a strong beat at the beginning of that note.

Time-span Reduction 'establishes the relative structural importance of pitch-events within the heard rhythmic units of a piece' (Lerdahl & Jackendoff, 1983, p. 231). Formally, this is notated as a tree diagram that 'expresses the way in which pitch-events are heard in the context of hierarchically organized rhythmic units' (p. 237). In other words, these reductions allow us to identify the relatively important pitch events (tones) in each rhythmic group, which determines the underlying 'deep' pitch structure. Finding Schenker's notation 'not explicit enough' (p. 112), GTTM borrows the 'tree' notation of linguistics to express the hierarchical nature of reductions. However, they point out that the similarity with linguistic trees is cosmetic; theirs are 'purely musical trees, having nothing to do with linguistic trees except that both express hierarchical structures with precision' (1983, p. 112). The tree notation appears above the musical notation, and can be quite elaborate for complex compositions. Underneath the musical notation, the reduction appears on a separate series of staves as seen in the time-span reduction of Bach's *C Major Prelude* in Figure 7.2.

The basic procedure for notating a time-span reduction takes several steps. One starts with a hierarchy of rhythmic groups using principles sketched in the previous sections. In each time-span, for each level of the hierarchy, one identifies a single chord or pitch event that functions as 'the head' (i.e. is prominent) according to whether it acts as a structural point of resolution within that time-span. For instance, in the first group of four chords in Figure 7.2, the initial C-major chord functions as the head; this is evidenced by the fact that the 'branch' from this chord extends over the branches from all other chords, thereby encompassing the rest. As one moves up the hierarchy to broader time scales, tones that functioned as 'heads' at lower levels become subservient to other 'heads' at higher time scales. Extending our previous example, the 'head' of the first four chords ultimately is itself subordinated by the final chord of the sequence, as evidenced by the fact that the branch extending from the final chord reaches over the branch from the initial chord. The result is a hierarchical structure of dominance/subservience in tonal relationships as segmented by time-spans, with branches showing elaborations of pitch events. (The well-formedness and preference rules associated with time-span and prolongational reductions are complex, and will not be specified here.)

Prolongational Reduction. The function of prolongational reduction is to express 'the sense of tension and relaxation involved in the ongoing progress of music . . . the incessant breathing in and out of music in response to the juxtaposition of pitch and rhythmic factors' (p. 179). Movement away from the tonal center, dissonance, a climactic point of a phrase, are all examples of musical tension while return to the tonic, consonance following dissonance, arrival at the end of a phrase, are usually experienced as relaxing. Diagrammed in 'tree' notation appearing above the musical notation, prolongational branching is represented by tensing motions depicted by right branches, and relaxing motion by left branches. The prolongation reduction for Bach's *C Major Prelude* is shown in Figure 7.3.

Unlike the time-span reduction, prolongations (as the name implies) are not constrained by group boundaries. Thus, whereas the resulting time-span units in a time-span reduction are fairly evenly spaced, units in a prolongational reduction can be of varying lengths and can extend across group boundaries. Similar to Schenker's 'prolongation,' a note or chord may remain in effect across measures or even a whole section, even when not physically sounded at each moment throughout them. For instance, a tonic triad can be heard as suspending across an entire section, including several contiguous time-spans. This is represented by the 'slur' notation in the reductions shown on the separate staves at the bottom of Figure 7.3. Note that in many respects the hierarchical structure of time-span and prolongational reductions are similar, as can be seen at higher hierarchical levels in Figures 7.2 and 7.3. However, at lower levels of the hierarchy these structures diverge; one sees less influence of metrical and grouping structure – hence less influence of 'time' – in prolongational than in time-span reductions.

More recently, Lerdahl (2001) has developed the theory of *Tonal Pitch Space* as a means of quantifying the intuitions behind prolongational reductions. This new theory generates distance metrics among pitches in a tonal system (cf. Krumhansl, 1990) and thus can be used to predict the degree of 'stability' associated with certain pitches (and chords), as well as the degree to which certain pitches 'attract' other pitches to them, as in the way that the tonic note 'attracts' the leading tone to it.

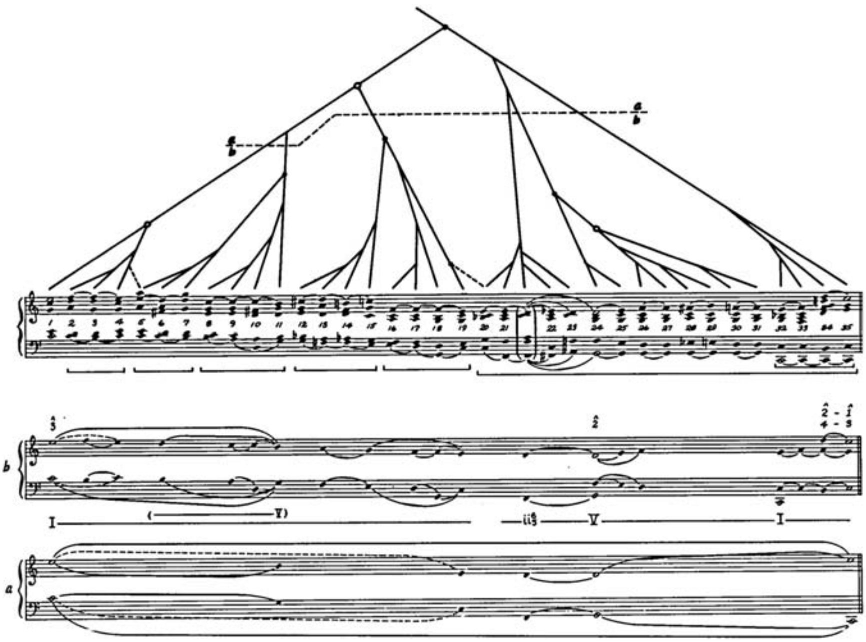

Figure 7.3 Prolongational reduction of Bach's *C Major Prelude* (Lerdahl and Jackendoff, 1983).

Source: Lerdahl, Fred, and Ray S. Jackendoff, *A generative theory of tonal music*, figure 10.11 page 263 © 1983 Massachusetts Institute of Technology, by permission of The MIT Press.

Our brief overview of the theory serves only as an introduction, and has barely captured the complexity of GTTM. In describing the notational system, it is important to remember that it is merely a convenient way of expressing what the theory aims to do – which is to provide a formal description of the musical intuitions of a listener, sorting the 'raw material' at the musical surface of a piece of tonal music into organized 'heard structures.' It is not a theory that focuses on the listener's moment-by-moment response to the tone-by-tone unfolding of a musical work in motion, but rather one that seeks to convey the listener's lasting impressions after a phrase or section or even a whole composition has been presented (i.e., the 'final state' of the listener's understanding [Lerdahl & Jackendoff, 1983, p. 4]).

Empirical support

GTTM has been tremendously influential on the study of music cognition. The website 'Google Scholar' reports well over 1200 citations of Lerdahl and Jackendoff (1983). In some respects, though, its influence has been felt more as an overall perspective than as a specific set of predictions. In comparison to the myriad citations of the theory, relatively few papers have rigorously tested its principles, although the authors have invited empirical tests of their work. This may be due in part to the fact that the original theory was not designed to frame quantitative predictions (Frankland & Cohen, 2004).

Of the empirical work on GTTM, most has been devoted to its claims concerning preference rules. One of the earliest attempts in this regard was Deliège (1987), who demonstrated that perceived segment boundaries are generally well predicted by GTTM rules for music drawn from the standard repertory. More recently, Frankland and Cohen (2004) verified grouping rules with artificially constructed stimuli designed to measure quantitatively the degree to which different principles predict grouping. A major advancement in the quantification of GTTM preference rules is described in Temperley (2001) who instantiated (and extended) these principles in an artificial intelligence system. The theory of Tonal Pitch Space, which evolved out of GTTM's concept of prolongational reduction, has likewise been shown to predict the experience of tension and relaxation in music listening (Lerdahl & Krumhansl, 2007).

The idea that the mental representation of music may be 'reduced' in the sense predicted by GTTM has also seen some support. For instance, influential work by Bharucha (e.g., 1984) has shown that when people hear a tonally unstable note followed by a stable note (see chapter 5), they hear the unstable note as if it were 'anchored' to its more stable neighbor. In other words, the unstable note is treated as somehow incidental. In other research, reductions have been found to predict perceptions of tension and relaxation (e.g., Krumhansl, 1996), closure (e.g., Palmer & Krumhansl, 1987), and memory (Large, Palmer, & Pollack, 1995). At the same time, a problematic result was reported by Cook (1987). He recomposed classical pieces so that they ended in a different key from where they began at the beginning. According to classic Schenkerian views (consistent with the spirit of GTTM), such alterations should profoundly alter the large-scale structure of the piece. Listeners, however, were apparently insensitive to these large-scale alterations. One possibility is that listeners are only sensitive to large-scale structures at a subconscious level that is hard to measure.

Melodic expectations and the Implication-Realization model

We now turn to a different approach to the cognitive representation of music. Whereas the ultimate goal of GTTM is to elucidate the 'deep' hierarchical structure of music, the next model seeks to explain expectations for events at the musical 'surface.' The emotional experiences that listening to melodies brings about are intimately connected with the way in which a sequence of notes fulfills or violates our expectations (see also chapter 14). Although an earlier theory of how these expectations work was suggested by Leonard Meyer (1956), a more fully formalized and widely tested version was proposed by one of Meyer's students, Eugene Narmour (1990, 1992) in his *Implication-Realization model*.

The structure of expectations

The central question Narmour asks is: 'What are the specific, note-to-note principles by which listeners perceive, structure, and comprehend the vast world of melody?' (1990, p. 3). Narmour's Implication-Realization model (hereafter referred to as the I-R model) assumes that a listener's experience of melodic structure is shaped by *expectations of how a melody will continue*. These expectations are influenced by two perceptual systems: one following 'bottom-up' principles that are not dependent on learning and are thought to be universal, and the other applying 'top-down' processes which are influenced by learning and may be specific to certain broad styles or genres of music, or even more specifically to the works of a particular historical period or to a single composer.

Narmour begins with melodic expectations of the broadest sort: the hypotheses that when presented with two musical events of some sort (e.g., two interval patterns, two pitches in succession, etc.), a listener will subconsciously or consciously infer one of two general outcomes:

(a) *Similarity* (or *process*): From the sameness or similarity in successive musical events, a listener comes to form an expectation of more similarity, which can be expressed in the form of the hypothesis (where → stands for 'implies'):

$$A + A \rightarrow A$$

(b) *Differentiation*: Change in successive musical events leads to the expectation of more change, expressed as follows:

$$A + B \rightarrow C$$

What are the dimensions along which sameness or differentiation may occur? Almost any musical dimension, in fact. According to Narmour (1992), these could be 'pitches and durations,' 'melodic intervals or durational patterns,' and musical forms as 'repeated or differentiated units.'

With respect to more specific melodic implications, on its simplest tone-to-tone level, the I-R model accounts for what we expect the third note to be after hearing two notes before it, assuming that the first two notes do not imply a completion. In this context, the first two notes form the *implicative interval*, and the second and third notes form the *realized interval*. (Note that in this context, the word 'interval' should be taken to refer to a *melodic* interval rather than a harmonic interval that is simultaneously played.) Realized intervals that match expectations suggest closure and release, whereas realized intervals that deviate from expectations suggest continuation and tension. According to Narmour, 'it is the dynamic, non-reductive, individuated melodic motion on the note-to-note level that captures music lovers' (1992, p. 331).

Given two melodic tones that are not perceived as completed, the listener generates expectations for the third tone, but not every possible pitch is as strongly implied for the continuation of the melody. The expectancies arising from the implicative interval are, in part, influenced by five 'bottom-up' principles of continuation that have roots in Gestalt ideas about grouping. In our discussion of Gestalt principles of perception in chapter 5, we saw that the Gestalt pioneers were mainly concerned with the visual perception of static figures. However, as Narmour explains, 'in invoking principles of proximity, similarity, and common direction in a temporal art, I hypothesize that rules governing such similarity apply not only retrospectively to perceptual things already seen or heard but also *prospectively* to *expectations* yet to happen' (1990, p. 73, emphases added). Formally, these implicative principles can be set out as follows:

(1) *Intervallic difference*: This principle suggests that after a small interval (five semitones or fewer, a semitone being equivalent to a half-step), the realized interval is expected to be similarly sized; whereas after a large interval (seven or more semitones [or half-steps]), the realized interval is expected to switch to a smaller interval. (Intervals of six semitones, which is an augmented fourth or diminished fifth, are considered neither large nor small, thus lead to neutral expectations.) In other words, intervals are typically expected to be equal to or smaller than a 'perfect fourth' (five semitones), and so when one encounters a large interval ('perfect fifth' or larger) the listener expects the following interval to revert back to the small interval standard.

(2) *Registral direction*: Following a small interval, the listener expects a realized interval that continues in the same pitch direction. That is, an ascending pitch interval that is small generates the expectation of continued ascents to follow. However, after a large interval, the listener expects the next interval to reverse its direction. Thus, a large ascending pitch interval generates the expectation that the next interval will go down.

(3) *Registral return*: A more extreme version of registral direction, registral return suggests that one expects the second interval to bring the pitch back (within two or less semitones [or half-steps]) to where it began at the beginning of the implicative interval.

(4) *Proximity*: Small realized intervals are favored, because these result in pitches that are close (i.e. proximal) to the end point of the implicative interval.

(5) *Closure*: Narmour emphasizes that this principle does *not* refer to the Gestaltist's idea of 'closure' and need not occur at the end of a phrase or section (see footnote 3, Narmour, 1990, p. 151), but refers more broadly to 'syntactic events whereby the termination, blunting, inhibiting, or weakening of melodic implication occurs' (1990, p. 102). Specifically, the implication of the initial interval is inhibited when there is a change in

registral direction, when a large interval is followed by a smaller-sized interval, or both.

Of the principles listed above, the principles of 'registral direction' and 'intervallic difference' play important roles in the I-R model. Moreover, the conjunction of these rules leads to the prediction of a melodic archetype identified by Leonard Meyer as the '*gap-fill principle*' (Rosner & Meyer, 1982). That is, if an implicative interval is smaller than five semitones, a listener expects a continuation of pitch direction. However, if an implicative interval is large the listener expects a reversal of pitch direction, as well as a smaller interval than was heard implicatively. A good example of the latter situation is the opening of 'Somewhere Over the Rainbow,' which begins with an ascending octave that is directly followed by a one-semitone downwards interval.

Note that in certain cases the principles may conflict with each other. For instance, registral return may predict a large interval at the same time as other principles (proximity, intervallic difference) predict a small interval. According to the model, such clashes are not problematic but rather contribute to the tension created by a melodic structure. Further, the I-R model does not restrict melodic implication to aspects of pitch and interval alone. For instance, with respect to closure, factors such as tone duration, meter, and harmony may influence both the location of the implication. These may also determine how weak or strong the expectancy for melodic continuation will be; for instance, Narmour points out that melodic implication can be suppressed by a rest pause, metric stress, or stable tonality.

Bottom-up and top-down expectancies

In Narmour's view, there are two independent 'tracks' operating simultaneously in melodic implication: one which operates on a more automatic and subconscious level (*bottom-up* processes) and one which interprets incoming data more flexibly within the framework of previously learned knowledge (*top-down* processes). So far, we have focused mainly on the 'bottom-up' principles, and in particular the application of elementary Gestalt-like principles to tone-to-tone developments of melodic forms. However, 'top-down' differentiations overlay and modify the bottom-up processes described above. These are the influences of one's knowledge and familiarity regarding musical style as a result of exposure or formal training, and consist of *intra-opus knowledge* (of the particular musical work) and *extra-opus knowledge* (of the style of music). Familiarity or formal training in a specific genre of music or a particular historical period or composer serve as sources from which 'top-down' principles shaping an individual listener's expectations may be derived. Each listener possesses a unique body of style knowledge that influences melodic expectations, and this in part may account for variations among the melodic expectancies of listeners.

Ultimately bottom-up and top-down principles jointly determine musical expectations, the former operating subconsciously and on a local level (e.g., an implicative interval), the latter operating consciously and on a more global level (e.g., pre-existing note patterns). Narmour insists on the independence of the two processes as they shape melodic expectancies in separate and distinct ways:

> In terms of *top-down cognition* . . . a lowly diatonic scale set in quarter notes is a style structure in that, once the listener consciously recognizes its inception, he can easily project its registral, intervallic, metric, and rhythmic continuation and also map out its scale-step conclusion on the basis of his prior knowledge. In terms of *bottom-up perception*, however, the listener also subconsciously, simultaneously, and locally processes each intervallic motion and each registral direction . . . separately since his initial global projection may be mistaken: the major scale he initially envisions from, say, an ascending C-D-E-F might, for instance, actually continue in a quarter-tone fashion after the F.
>
> (1990, p. 9)

We conclude our discussion with two examples of the application of Narmour's model. Let us first illustrate the application of the basic 'bottom-up' principles, or cognitive primitives of the Implication-Realization model, in a simple example shown in Figure 7.4 (from Narmour, 1991, Figure 5, p. 7). The focus here is on the first four notes, or the first three intervals. Intervals 1 and 2 are 'small' according to the I-R theory, and as such one expects a continuation of both registral direction and interval size. However, a surprise comes with the third interval (signified by the exclamation point). Here we have a large interval where a small one is expected. In addition, the large interval continues registral direction and thus we have *differentiation* on one dimension (interval size) and *process* on another (registral direction). This conflict adds to the surprise and – according to Narmour – enhances the listener's aesthetic experience.

Through repeated exposure, the listener can form expectations by way of

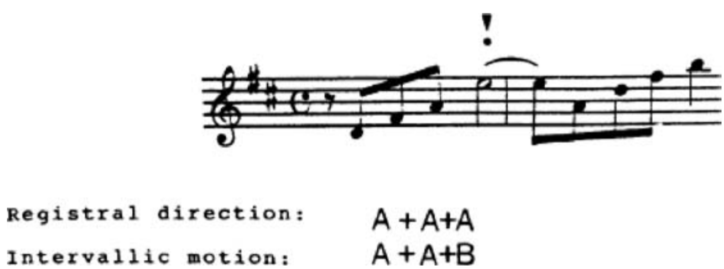

Figure 7.4 'Bottom-up' analysis using the I-R model (Narmour, 1991, figure 5).

Source: Reprinted with permission from the author and the University of California Press.

the top-down system that counteract bottom-up tendencies. This leads to a curious prediction of the I-R theory: Music can surprise our top-down expectations while at the same time fulfilling our bottom-up expectations. An example is shown in Figure 7.5. Here we see a motif that is repeated twice; based on *bottom-up* principles, the listener should expect *process* (a repetition of the motif). However, a different motif follows these repetitions. Narmour argues that, over time, a listener comes to expect the differentiation A + A → B within the *top-down* system in a way that counteracts the A + A → A expectation typical of the *bottom-up* system. The example here does just this. In this example, Narmour points to a clever embedding of themes within themes. Within the 'B' segment identified at the top of the graph there is an embedded theme, B + B → C that mirrors the A + A → B theme at the higher hierarchical level.

The I-R model is complex and consists of many principles to account for expectations, and the above introduction to the I-R model is necessarily simplified. A full explication of the model requires an eight-page glossary of symbols, of which we have not included any. Narmour's various books and articles have elaborated the model we have sketched in several dimensions. Subsequent research has confirmed and extended Narmour's ideas in significant ways, as discussed in the next section.

Empirical support

Like GTTM, empirical tests of the I-R model have primarily focused on the simpler claims of the model, that is, its 'bottom-up' components. These principles appear to account for melodic expectancies quite well, particularly when combined with a component based on the tonal hierarchy that arises from Krumhansl's (1995) research. For instance, when given two tones and asked to rate how well a third tone serves as a continuation of the melody, the responses were in line with Narmour's predictions (e.g., Cuddy & Lunney, 1995). Similarly, when participants were given two notes and asked to

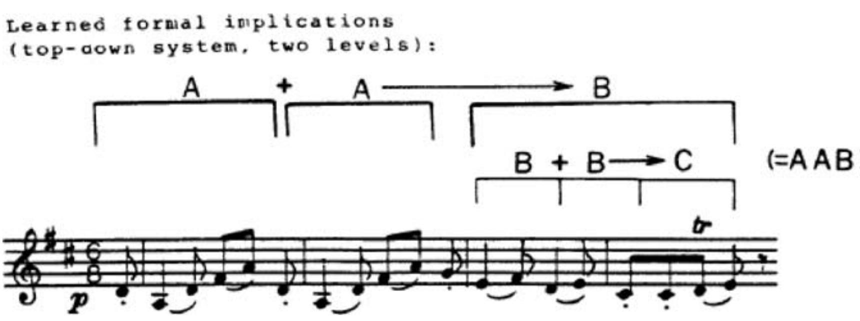

Figure 7.5 The interplay between 'bottom-up' and 'top-down' implications (Narmour, 1991, modified figure 15a).

Source: Reprinted with permission from the author and the University of California Press.

compose the rest of the melody, the third tone they chose in creating their own melodies was well predicted by the I-R model (Thompson, Cuddy, & Plaus, 1997). The level of musical training of participants did not make a difference in these studies, which is consistent with Narmour's assumption that the application of 'bottom-up' principles is not dependent on formal musical training. Further, Krumhansl (1995) and Schellenberg (1996) have shown that these 'bottom-up' expectations can transcend cultural boundaries, and can apply to expectancies for Chinese music and – unlike Krumhansl's tone profile – they can also account for expectances in atonal music (Krumhansl, 1995).

However, in one recent cross-cultural study the I-R model was shown to be limited (Krumhansl et al., 2000). This research investigated expectancies for Sami Yoiks, a style of music originating in Northern Europe and characterized by large pitch leaps. As one might expect, the I-R model, based largely on an assumption that small intervals are favored, did not provide as strong an account of expectancies for Yoiks as it has for other musical styles. Interestingly, the fit of the model was culture-specific. Listeners who were unfamiliar with Yoiks made expectancy implications that were better predicted by the model than did listeners who were familiar with Yoiks. In other words, Yoik-listeners had modified the schemata they used to form expectations for melodies.

Narmour's model has also stimulated further theoretical work, refining and expanding models of melodic expectation. Schellenberg (1996, 1997), for instance, demonstrated that the I-R model could be simplified by removing some predictor variables without loss of its predictive power. His parsimonious 'two-factor model' proposes that melodic expectancy is determined by the principle of pitch proximity and pitch reversal, two ideas articulated by Narmour. Drawing on the work of Narmour (1990, 1992) and Lerdahl (2001), Margulis (2005) proposed a model of melodic expectation, which assigns expectancy ratings ('predictions about degrees of expectedness' [p. 664]) to melodic events and associates them with listeners' experiences of tension across melodies. Aside from providing a rubric for computing the expectedness of melodic events, the model also 'enables the graphical representation of moment-to-moment expectancy-based fluctuations in affect across a melody's course' (p. 665), capturing the dynamic and real-time aspects of listening to music.

Coda

There are many different ways to take apart a mechanical toy or a clock, a poem, or a musical work. The guiding questions we have in mind, the way we inspect the internal structure, the tools we use, and what we choose to focus our attention on, may all lead to different but often complementary discoveries. In this chapter, a few different approaches to understanding musical structure were explored. The Generative Theory of Tonal Music and the Implication-Realization model capture different aspects of the way listeners perceive and comprehend a piece of music, and both have contributed significantly to our understanding of music cognition. However, none of the theorists claim to have set forth a 'complete' or 'finished' theory. Further, the theories do not attempt to 'explain music' but to capture the complex combination of innate and learned responses in listeners' perception and cognition of a musical work. Drawing from an analogy by psychologist Paul Fraisse, Narmour explains: 'gravity does not explain architecture, but architecture is subject to its law; likewise, perceptual laws do not explain music, but music cannot escape their influence' (1990, p. 4).

The theories described here offer a good starting point toward understanding the perception and cognition of a piece of music. However, many questions ensue: How do musical structures give rise to emotions in listeners? And how does music come to have *meaning* for the listener? These questions are addressed later, in chapters 13 and 14. But first, having explored the perception and cognition of melody, rhythm, and musical structure in the previous and present chapters, we turn our attention to the emergence and development of these and other musical abilities in the next chapters.

Notes

1 The publication date for Lerdahl and Jackendoff's paper, 'An overview of hierarchical structure in music,' is cited in this chapter as 1983–84. Note that this is a different publication from their 1983 book entitled *A Generative Theory of Tonal Music*.

2 This rule would lead to the privileged choice of the time signatures 4/4 or common time and 3/4, waltz time.