

Chapter 10

Music perception

W. Jay Dowling

10.1 Introduction

When we listen to music, as well as to other patterns of sound, many of the same principles of perception apply as in other domains like vision. We immediately encounter figure-ground relationships between the focal aspect of the pattern we are attending to, and the rest of the sounds that reach our ears. As an example, suppose we hear the piece shown in Fig. 10.1, the slow movement from Beethoven's 'Spring' sonata for violin and piano (*Sonata in F Major, op. 24*). When the piano begins to play, it stands out as a figure in contrast to other sounds in the room. You follow the undulating accompaniment pattern in the left-hand piano part, but this recedes into the background, like wallpaper, when the melody starts in the upper voice. The dynamic shape of the melody dominates the auditory scene, and though you may notice the quiet entry of the violin in the second measure, your attention reverts quickly to the melodic shape.

These are very simple observations, but already we encounter a distinct difference between audition and vision. At first our attention is drawn to the piano as a *sound source*, in contrast to all the other sound sources in the environment. But then the sound of the piano splits into two *sound objects*—the melody and the accompaniment—both coming from the same source. The piano as a sound source is located in physical space, along with the violin, the rattling of page turns, coughs in the audience, etc. The melody as a sound object is, in contrast, located in a kind of 'virtual space' of pitch and time, as is the background accompaniment. Melody and accompaniment have the same location in physical space, but different locations in pitch and time. Kubovy and Van Valkenburg (2001) provide a stimulating discussion of this contrast, pointing out that the closest auditory analog of shape perception in vision is the perception of melodic contours traced in pitch across time.

When the melody shifts to the violin in measure 10, melody and accompaniment come from different locations in physical space as well as occupying different pitch ranges (and contrasting in timbre—their instrumental tone color). These contrasts help clarify the musical structure; being present at a concert changes what we hear and understand of the music. As Vines *et al.* (2006) have documented, the gestures of the musicians while playing communicate aspects of the musical message, as well as reinforce the differentiation of sound sources. (This could be especially important in listening to a string quartet, where all the instruments have similar timbres, to help differentiate what each instrument is doing, or a rock band in which the instruments tend to mask one another's sound. The visual message helps us focus attention on a particular instrumental line as figure.)

The contour of the melodic line has what Mari Jones has called 'dynamic shape' in emphasizing its motion in time through the pitch space (Jones, Summerell, and Marshburn, 1987). The dynamic motion of the melody conveys feelings of tension and relaxation. The first phrase is felt as very relaxed. Then very quietly the tension increases, reaching a peak in the third two-bar phrase. Then it recedes as the eight-measure melody comes to an end on the very stable tonic

The image shows the beginning of the second movement of Beethoven's Sonata for violin and piano, op. 24. The score is in 3/4 time and B-flat major. It features a piano introduction with a steady eighth-note accompaniment in the left hand and a melodic line in the right hand. The tempo is marked 'Adagio molto espressivo'. The score is divided into measures 1 through 17, with some measures containing 'cresc.' markings. The piece concludes with a final cadence in measure 17.

Fig. 10.1 The beginning of the second movement of Beethoven's *Sonata for violin and piano op. 24*.

pitch (Bb). Much of this increase and decrease in tension arises from the relative stability and instability of pitches in a tonal context. In tonal music (which includes the pitch-organizing systems of virtually all the world's folk and art music before the explorations of twentieth-century Europe), the pitches are arranged in a hierarchy, with the tonic (the first degree of the scale) as the central and most stable pitch. The less stable pitches are attracted in greater or lesser degree toward the more stable pitches. You encounter this kind of instability very strongly if you sing a nursery song like 'Twinkle, Twinkle' and stop before you reach the last note: 'How I wonder what you ...'. Stopping on the very unstable second degree of the scale makes you feel how strongly it needs to resolve to the stable tonic. We shall explore this further in looking at the work of Krumhansl (1990) who found that listeners are quite consistent in their judgments of the stability of pitches in a tonal context, and are even able to track their position in a virtual space of tonal pitch while listening to a piece that shifts from key to key (that is, from one tonal scale to another). The virtual space of pitch is well represented in the human brain (as Janata *et al.*, 2002, have shown).

Pitch and time as dimensions of the virtual space of audition have a feature that sets them apart from the physical space of vision: these dimensions are calibrated in terms of cognitive frameworks that help listeners keep track of the positions of the musical notes in the space. As Helmholtz

(1877/1954, p. 252) pointed out, ‘Melody has to express a motion, in such a manner that the hearer may easily, clearly, and certainly appreciate the character of that motion ... This is only possible when the steps of this motion, their rapidity and their amount, are ... exactly *measurable* by immediate sensible perception.’ As we shall see, pitch is heard in terms of motion along the musical scale framework that underlies the melody, and time and rhythm (the ‘rapidity’ of the motion) are heard in terms of a framework of beats. We feel the tension and stability of a melody in terms of its departures from and returns to the stable tonic region of the scale, and the temporal solidity of being on the beat.

Before we look in detail at how music is perceived, there is one more general feature of perception to consider. Perception is intimately bound up with memory, and is very difficult to conceive of as a separate process. In order for us to experience it, information from the outside world must be held in a memory buffer, and that information is registered in such a buffer as soon as it enters our nervous system. We do not experience raw, uninterpreted sensory data. The stimuli we experience have already been encoded in terms of our customary categories of perception (‘assimilated’, in Piaget’s terms (see Inhelder and Piaget, 1958)). When we hear speech in our familiar language, we hear meaningful words, not uninterpreted speech sounds or babbling, and when we listen to tonal music we hear pitches in terms of their places in the tonal framework, and we hear them occurring in relation to a metrical framework of beats. The notes of a melody have already been encoded into our memory buffers by the time we experience them. Furthermore, what we acknowledge having heard depends on how long after the fact our memory system is queried; seconds make a difference in the answer.

There is another way in which memory has an immediate impact on our perception. We experience a piece of music as extended in time, and not present all at once in our auditory field (the way a picture is in the visual field). This means that as we continue listening we are continually evaluating (usually implicitly) the familiarity of what we are hearing, and understanding how it fits what preceded it. For each new phrase, we are implicitly asking, echoing William James (1890), ‘Is that thingumabob again?’ We are tracking the degree to which the present phrase is like what went before, and how it fits its context. Too much new and different material and the music becomes incomprehensible; too much similarity, and it becomes boring. There is good reason to believe that this tracking relies on our sense of familiarity with the new material, and not on explicit recollection of the previous phrases. (See Yonelinas, 2002, for the distinction between familiarity and recollection.)

It is clear from this discussion that music perception is multidimensional. Perceiving music leads the brain to combine patterns of features varying on several sensory dimensions at once. In this chapter we will look at those dimensions one at a time, and see how they fit together into a meaningful overall pattern. It is also clear that we hear music embedded in a particular culture. Our perceptual habits have been shaped by the regularities of the music we are used to hearing. Those habits lead us to expect certain continuations of what we are hearing, and so the things that surprise us are determined by what is usually done. The cognitive frameworks for pitch and time, so useful as aids in our understanding of the music we hear, are largely culture-specific. It takes considerable perceptual learning with the music of a new culture before we automatically encode the relevant features and contrasts that are built into it, just as it does with language.

10.2 Perceptual frameworks

When we listen to music from our own cultural tradition, what we experience is not a stream of undifferentiated sound, but sounds that are encoded in terms of the cultural system of categories. This is similar to what happens when we listen to speech in a language we know: the sounds enter

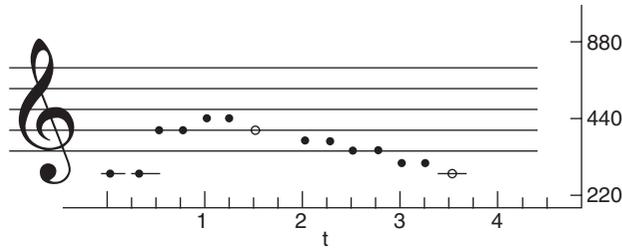


Fig. 10.2 The tune ‘Twinkle, Twinkle, Little Star’ presented at 4 beats/s (MM = 240 beats/min) displayed on a grid showing pitch and frequency in Hz (f) vs. time in s (t), with the musical staff notation superimposed on the y-axis.

our conscious experience already encoded as syllables, words, and sentences. To convince yourself that we hear musical pitches in terms of scale categories, sing ‘Twinkle, Twinkle, Little Star’ and stop on the next-to-last note: ‘... How I wonder what you.’ The pattern sounds incomplete when it ends on *re* in the *do, re, mi* scale; it needs to get back to *do*, the tonic, to end in a stable place and sound finished. Traditional Chinese music is based on pentatonic scales, and Lao-Tse (1972) warns us that ‘the five tones deafen the ear’—that we should try to hear things as they really are, and not in terms of conventional categories. In the twentieth century composers tried a variety of ways of breaking through our automatic encoding of pitches: by giving all 12 pitches in the octave equal weight so there was no longer a stable tonic to return to (Schönberg, 1967), by constructing scales with unfamiliar patterns of dividing the octave (Partch, 1974), or by constructing musical patterns out of sound objects found in nature and their transformations (*musique concrète*, Schaeffer, 1952). But the overwhelming preponderance of music we hear is constructed in our familiar tonal system, and our ears are well practiced in encoding it in terms of its categories.

The category systems of pitch and time divide our auditory space into a two-dimensional grid such as that shown in Fig. 10.2. The y-axis, pitch, represents a logarithmic scale of frequency along which the frequency doubles with every octave (going from an A at 220 Hz to an A an octave higher at 440 Hz, to an A another octave higher at 880 Hz). The x-axis, time, is divided into beats and their subdivisions. Time perception organized in terms of beats is much more precise, with a much smaller just-noticeable difference between similar time intervals, than time perception not so organized (Bharucha and Pryor, 1986). In our Western European tradition we are so used to beats occurring at absolutely regular intervals (such as in ‘Twinkle, Twinkle’ where beats are grouped by twos) that we don’t often consider the possibility of unevenly timed divisions of the measure into groupings of 3 + 2, for example, that are characteristic of some Eastern European songs. Grouped in this way, ‘Twinkle, Twinkle’ would have a hitch in its rhythm: ‘Twi-uncle, twinkle, lih-uh-tle star; howow I wonder wha-hat you are.’ Hannon and Trehub (2005) showed that whereas American adults had difficulty noticing structural violations in the more complex 3 + 2 patterns, Bulgarian adults coped equally well with their familiar 3 + 2 patterns and the simpler 2 + 2 patterns (such as in Fig. 10.2). Six-month-olds, not yet acculturated to one or another pattern, also coped equally well. Just as the pitch categories on the y-axis of Fig. 10.2 are the result of acculturation (for example, in traditional Indonesia the pitch categories of the five- and seven-tone scale systems can even vary from village to village), so are the time categories on the x-axis.

10.2.1 Pitch

The pitches of tonal music, in whatever culture, are specified in terms of a hierarchical series of constraints, illustrated for European music in Fig. 10.3 (Dowling, 1978; Dowling and Harwood, 1986). We start with the non-musical *psychophysical scale*, which assigns perceived pitches to physical frequencies (the frequencies of sine waves, or the fundamental frequencies of harmonic

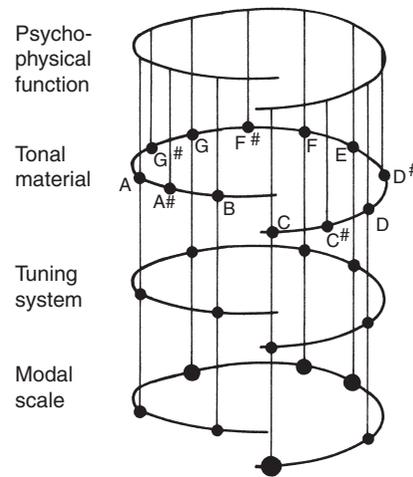


Fig. 10.3 Levels of analysis of the pitch material of music, illustrating the selection of culturally defined pitch categories in Shepard's helical psychophysical scale for pitch. Tone height in octaves is represented on the y-axis, and tone chroma (tonal scale value) is represented around the helix. (From Dowling and Harwood, 1986.)

complex tones—see Chapter 4). As a first approximation the psychophysical scale could be represented as a logarithmic scale relating pitch in octaves to frequency—doubling the frequency increases pitch by 1 octave. However, especially where music is concerned, pitches an octave apart are perceptually similar to each other, and serve the same functions in the tonal structure. When men and women sing together, they sing the same song but with the pitches an octave apart, and cultures that name their pitch categories give the same (or similar) names to pitches an octave apart. As Shepard (1982) pointed out, this octave equivalence means that the helix shown in Fig. 10.3 is a natural way to represent the relation of pitch and frequency. Pitches an octave apart are close to one another, and for every octave that the pitch ascends, it goes through a cycle of the helix, returning an octave higher to a point on the helix just above the point representing that pitch class in the lower octave. This allows us to classify each pitch in terms of its *pitch height* or octave level, the y-axis of the graph, and its specific functional pitch quality or *chroma*, which changes as we go around the curve of the helix.

The helix also captures another important fact about pitch perception, that what defines a melody is the pattern of relations among the pitches, and not the absolute pitch levels themselves. You can start 'Twinkle, Twinkle' on any pitch, and as long as you follow the same pattern of pitch intervals from note to note, it will still be 'Twinkle, Twinkle'. Those pitch intervals can be represented as movements back and forth around the helix. Shifting the tune to a new starting point, a new 'key', means translating that pattern of intervals in a screw-like motion along the helix. The helix has the very useful property that a tune (its interval pattern) can be shifted anywhere along it in this way without altering its shape.

Out of the infinity of pitches represented on the helix, the psychophysical scale, each culture selects a set of pitches for use in its music, which I have called the *tonal material*. In European music the tonal material consists of the chromatic scale that divides the 2/1 frequency ratio of the octave into twelve logarithmically equal steps, called *semitones* or half steps. Ascending 1 semitone along the chromatic scale involves multiplying the frequency by $2^{1/12} = 1.058463 \dots$; note that doing this twelve times lands us exactly 1 octave higher: $(2^{1/12})^{12} = 2$. This method of generating the tonal material in European music is called *equal temperament* because of the uniformly equal semitones that constitute the building blocks of all its tonal structures. Equal temperament was borrowed from China around 1720, and makes possible a shift in pitch of the tonal center (*tonic*)

to any of the twelve pitches in the tonal material. Such a shift of the tonal center along with all of the tonal scale pitches related to it is called *modulation*.

Note that in defining the tonal material and the subsequent levels of analysis shown in Fig. 10.3, what is important is the *relative*, and not the absolute, pitches involved. This is easy for us to forget in the European tradition, where we think of the pitch A in the middle of the keyboard as having a fundamental frequency of 440 Hz. But this standardization of pitch levels is only about 100 years old (see Ellis's appendices to Helmholtz, 1877/1954). Unless one is among the small minority of the population with absolute pitch (the ability to name pitches out of context), small variations in the anchor points of the tonal material amid the infinity of possibilities in the psychophysical scale will not be noticeable.

Out of the tonal material, cultures often (but not universally) select a more restricted set of pitches, called a *tuning system*, as the basis for the formation of modal scales to be used in music. For example, in European music we could select the seven white keys on the piano (C, D, E, F, G, A, B) as a tuning system, one that could provide the basis for the keys of C major and A minor. In some cultures the tuning system is simply identical with the tonal material. For example, many Native American cultures use a tuning system (and tonal material) essentially the same as the pentatonic tuning systems of China and Europe (European white notes C, D, E, G, A, or black notes F#, G#, A#, C#, D#) from which modal scales and melodies can be formed.

The final step in this process, of moving from the abstract to the concrete in the selection of pitches with which to make a melody, is to form a *modal scale* from the tuning system. This involves establishing a tonal hierarchy (Krumhansl, 1990) on the pitches of the tuning system. For example, selecting C as a tonic in the set of white notes establishes the modal scale of C major, with a tonal hierarchy in which the most stable pitches are C, E, and G, and pitches D, F, A, and B are less stable (that is, have strong tendencies that 'pull' them toward the more stable pitches. To take the example of 'Twinkle, Twinkle' again, if the tune is in C major and begins on C, then the next-to-last note D has a strong tendency pulling it toward the C, that we feel when we stop short without resolving that tendency. Similarly, if we sing a *do-re-mi* scale beginning on C, and stop on the seventh pitch: *do, re, mi, fa, sol, la, ti* ... (C, D, E, F, G, A, B ...), we feel a very strong tendency pulling the B toward the upper C. These tonal tendencies have implications for how we perceive the pitches in the music we hear. Francès (1988, Experiment 2), for example, showed that if the pitches of a piece of music are altered in the direction of their tendencies in the tonal context, they are much less noticeable as 'out-of-tune' notes than pitches altered in the opposite direction.

The expectations we have of what pitches are coming next as we listen to music are engendered by our experience at the level of modal scales and the tonal hierarchy. The violation and resolution of these expectancies have strong emotional effects (Sloboda, 1998; Bharucha, 1999). Meyer (1956) called attention to this rise and fall of tension as an important source of our emotional response to music. If we return to the Beethoven sonata movement in Fig. 10.1, we can see examples of tonal tendencies and their resolution in the melody that starts in measure 2. The melody begins on D in the tonic chord Bb-D-F. The chord is stable, but the D is a little less stable than the tonic Bb, and this instability is resolved in the next measure to the Bb. The piece continues with pitches in the stable tonic chord until measure 5, where the harmony shifts to the dominant seventh chord F-A-C-Eb, in which the least stable pitch, Eb, is found in the melody. A very standard practice would be to resolve the dominant seventh to the tonic chord, with the Eb following its strong tendency downward to the D. However, at this point Beethoven defies gravity and takes the melody line to G in measure 6—an even less stable pitch in the context of the underlying F-A-C dominant chord, and one that generates acoustic dissonance with the neighboring scale notes F and A. The melody outlines the subdominant chord Eb-G-Bb against the dominant chord

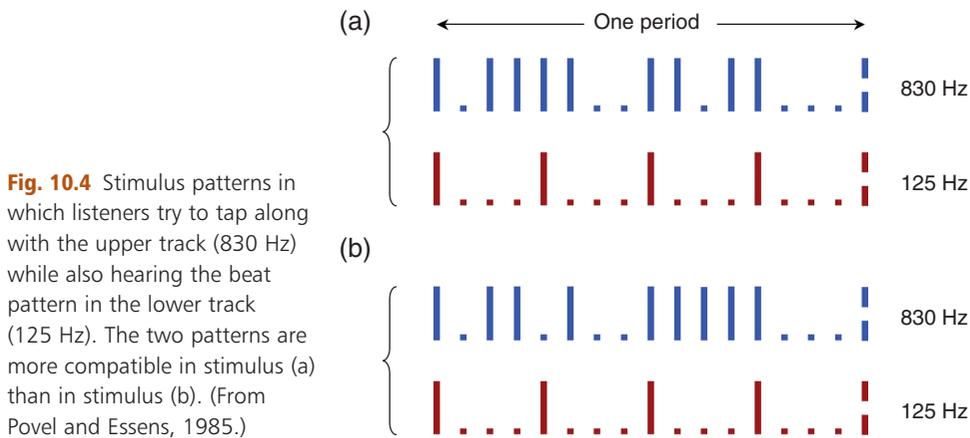
in the accompaniment, landing on the Bb in measure 7. This Bb, though the tonic, is now very unstable in the context of the continuing dominant chord (F-A-C), and needs to resolve downward to the A, which it does. We are now back to an unambiguous dominant harmony in both melody and accompaniment, which resolves peacefully to the tonic chord in measure 9. Thus in the course of eight measures Beethoven traces an excursion into very unstable tonal regions, and then back again to the stable tonic. This excursion gives the melody emotional energy that is immediately felt by the listener. These considerations of the tonal hierarchy show in detail how this is done, and there is clear evidence that musicians track the relationship of the pitches they hear in a piece to the tonal center (Toiviainen and Krumhansl, 2003) and that non-musicians as well as musicians track the rise and fall of tension in chord sequences (Bigand *et al.*, 1996). Of course Beethoven has at his disposal techniques for even greater excursions; this is just a brief example in an eight-measure melody. In a movement lasting several minutes complications can be added by modulating to other keys (moving the pitch level of the tonic) and then back again, and by introducing contrasting melodic and rhythmic material. Musically trained listeners track such modulations, and Janata *et al.* (2002) have shown how that tracking is represented in the brain. And non-musicians provide evidence of knowing implicitly what key they are in during the middle of a continuously modulating piece, in that they respond quickly and accurately to out-of-key pitches when they occur (Janata *et al.*, 2003). Listeners have an implicit knowledge of musical structure, and that knowledge, applied through the fulfillment and violation of expectancies, leads them to experience the ebb and flow of tension in the music, as Meyer (1956) theorized.

10.2.2 Time

Our perceptual and memory systems impose constraints on the temporal framework in terms of which we hear music. The sequence of sounds must not go too fast—somewhere between 10 and 20 notes/s it becomes a blur in which we lose track of individual pitches—and it must not go too slow, either. As a practical matter, we have difficulty recognizing familiar melodies when they go much faster than 6 notes/s (167 ms/note) or slower than about 0.6 note/s (1670 ms/note; Warren *et al.*, 1991; Dowling *et al.*, 2008). The notes simply do not hang together in a meaningful pattern at slower tempos. Furthermore, much music, especially songs and folk music, is organized into phrases, and the phrases are generally no longer than about 5 s, which is about the duration of material that can be stored in our auditory sensory memory buffer (Dowling and Harwood, 1986; Winkler and Cowan, 2005).

The temporal framework (Fig. 10.2) is organized into beats. We have a strong tendency to hear a beat in the music we listen to. Even if we hear just a series of evenly spaced clicks, we tend to group them perceptually into patterns of twos or threes or fours, set off by strong and weak beats (Fraisse, 1982). We find it easy to tap along with the beat in a piece of music, but only if there are sufficient cues in the music to where the beat is. Conversely, we can tap a complex rhythm along with a beat, but only if they are going at the same speed (that is, if they coincide periodically at the same spot in the pattern; Povel and Essens, 1985). Povel and Essens also showed that complicating the relationship between the rhythmic pattern and the beat makes the pattern more difficult to follow. They had listeners tap along with the upper line in Fig. 10.4; this was much easier for the simpler pattern (a) than for the more complex pattern (b).

A problem arises concerning how our internal beat is represented in the nervous system. The internal representation must be precise enough so that we can succeed in tapping along with the music, but it cannot be so rigid that it will fail to follow variations in tempo. Music is full of slight and large alterations in tempo from moment to moment, often introduced in the interest of emotional expression, aesthetic impact, and naturalness (see Gabrielsson and Lindström, 2001). A piece in



which the beats and notes occur with rigid temporal regularity sounds dull and mechanical. Hence the grid illustrated in Fig. 10.2 is an oversimplification of what occurs in practice. How do we track the sequence of beats, and remain synchronized with the continually varying music? Large and Jones (1999) have proposed a solution in their theory of internal oscillators. They propose that we tune internal oscillators to temporal regularities in the environment. However, unlike a rigid beat structure, the oscillators continually update their relationship to the rhythmic events. They generate expectancies, that are then checked against reality and the oscillators are reset to continue tracking the events. An oscillator produces a periodic signal controlled by two parameters which can be reset: its tempo and its phase. If the oscillator is set to correct tempo to match the event sequence, but is out of sync with the events, its phase can be reset to achieve synchronization. This resetting is typically automatic, but can also be done intentionally to anticipate a hitch in the rhythm (Repp, 2002). The system copes with expressive timing changes such as a *ritardando* at the end of a phrase by gradually resetting the tempo.

10.3 Attention

The frameworks of pitch and time assist the direction of attention to important aspects of the musical pattern. When we are familiar with a style of music, our attention is directed automatically to regions in pitch and time where important events are likely to occur. This is similar to what happens when we learn a language: through practice we learn to direct our attention to important details in the stream of speech, so that what had formerly been an undifferentiated stream of sound becomes a comprehensible stream of words and phrases. We can see the effects of such learning with musical patterns in the examples of Fig. 10.5. In Fig. 10.5(a) we see two familiar melodies interleaved in time: 'Frère Jacques' (odd notes) and 'Twinkle, Twinkle' (even notes) played in two different octaves. It is easy to direct our attention to one or the other and perceive it clearly (Bregman, 1990); see also Chapter 8. In Fig. 10.5(b) the two melodies are interleaved in the same pitch range. Now it is very difficult to hear either one. However, with less than an hour's practice most people, musicians and non-musicians alike, can come to hear one or the other target melody when they are interleaved as in Fig. 10.5(b), provided they know what melody to listen for (Dowling, 1973). Knowing the target melody makes it possible to aim our attention at points in pitch and time where notes of that melody are likely to occur, and verify whether they did. As we would expect from Large and Jones's (1999) oscillator theory, it is easier to discern an

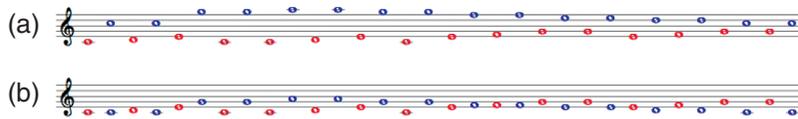


Fig. 10.5 (a) The tunes 'Frère Jacques' (red notes) and 'Twinkle, Twinkle' (blue notes) interleaved in time in separate octaves, presented at 8 notes/s. (b) The same two tunes interleaved in the same pitch range.

interleaved target when its notes occur on the beat (odd notes) than when they are off the beat (even notes; Dowling *et al.*, 1986). Our attentional system is apparently able to aim at a series of little windows in the pitch and time grid, and select the information that occurs there. This is illustrated by another experiment of Dowling *et al.* (1987). On each trial the pattern shown in Fig. 10.6 was presented at about 6 notes/s. The target note was the E in the middle of the pattern. When the pattern was repeated, it was interleaved with distractor notes, and the E moved to another pitch. Finally a probe tone was presented, and the listener judged whether the probe was higher or lower or the same in pitch as the pitch of the target. When the target moved within two semitones up or down, responses were relatively accurate; but on the few occasions when the target moved outside the limits of the pattern (to a higher or lower A), the listeners completely lost track of the target and performance fell to chance. Within the focus of attention, pitch judgment was affected by expectancies based on the system shown in Fig. 10.3. As long as the target landed on a pitch in the tonal material (D, D#, F, F#), judgment of the probe was relatively accurate. If the target landed on a quarter step (0.5 semitone) between those pitches, however, the pitch judgments were assimilated to neighboring tonal scale steps. That is, if the target was a pitch midway between D and D#, the probe D was judged equal to it. This indicates that when pitch encoding is hurried (because of the immediately following distractor note), the pitch encoding system takes the nearest scale note as a default value. This suggests that in listening to music, our auditory system will 'clean up' slight intonation errors in rapid passages.

There is considerable evidence that the perception of expected musical events is faster and more accurate than for unexpected events, due to the preparation for processing that expectancy sets in motion. For example, Bigand *et al.* (1999) presented chord sequences followed by target chords

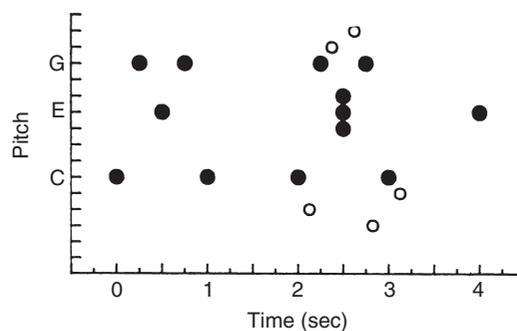


Fig. 10.6 Stimulus pattern in which the target pattern is first played alone at 4 tones/s (filled symbols), and then with interleaved distractor notes (open symbols). The center tone of the target is presented at a new pitch, higher or lower (multiple filled circles). Finally, a probe tone is presented, and the listener judges the pitch of the probe relative to that of the target tone. (From Dowling, 1992).

that were in-tune or out-of-tune (with the fifth of the chord raised by a semitone—a relatively obvious mistuning). Both musicians and non-musicians were better at detecting the mistunings in expected chords than in unexpected chords, with correct responses around 10 percentage points greater and 100 ms faster. Our experience with a particular musical style develops our attentional habits, which facilitate the processing of musical patterns within that style.

10.4 What we perceive

We have the subjective impression of perceiving things in the world just as they are, but as Treisman (2006, p. 317) points out, we are able to maintain this impression because we are seldom tested at those awkward moments when our finished perceptions—the world as we will remember it—are still under construction. Brunswik (1956, 2001) developed a ‘lens’ model of perception (Fig. 10.7), characterizing the coherent representation of the world we experience as achieved across a seemingly chaotic complex of sensory processing. Not only does our mental representation of events in music depend on what we select for attention, but it depends on the continued processing of a phrase even after subsequent phrases have entered the auditory sensory buffer. We can see this in the results of Dowling *et al.* (2001) who presented listeners with the beginnings of classical minuets. One of the initial phrases would be a target phrase to be tested later. The music continued, and a high-pitched signal indicated the occurrence of a test phrase, which was either a replica of the target phrase, a similar lure with the same melodic contour and rhythmic pattern as the target but at a different pitch level, or a different phrase. Figure 10.8 shows examples of these types of test in Beethoven’s Minuet in G. Listeners had to say whether the test phrase was exactly the same as the target they had heard. In general, they found it easy to say that target test phrases were the same and different test phrases were different, no matter how long the delay between target and test (up to 30 s). However, the responses to the similar lures varied with delay. When interrupted after a 5-s delay, listeners confused the similar lures with the targets, shown in a high proportion of ‘same’ responses (false alarms). But when posed the same

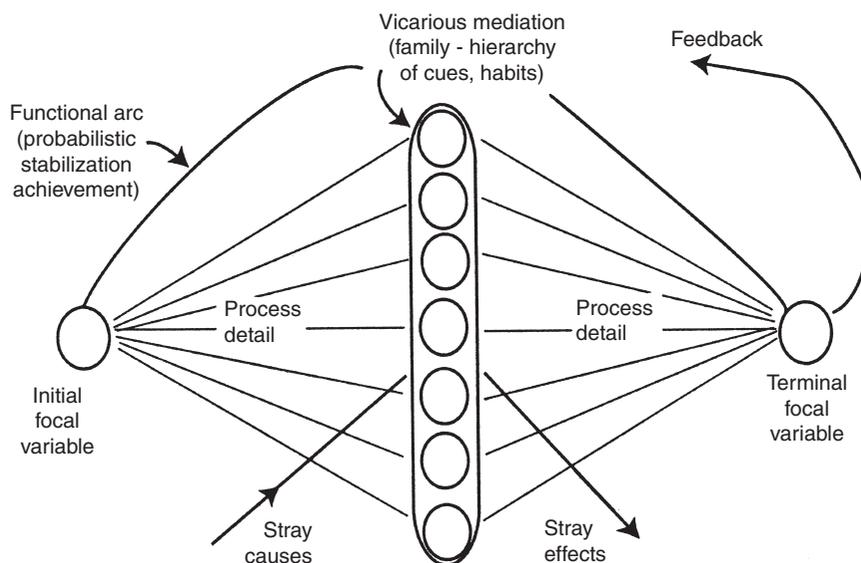


Fig. 10.7 Brunswik’s lens model of perception. (From Brunswik, 2001.)



Fig. 10.8 The start of Beethoven's *Minuet in G* illustrating the structure of stimuli from Dowling *et al.* (2001), with a possible target phrase (bracket 1), a possible test item testing the target with a similar lure (bracket 3), and a possible different lure (bracket 7). (From Dowling *et al.*, 2001.)

question after 15 s their confusion had disappeared, and listeners no longer thought they had recently heard the similar lure. In the 10 s that had elapsed since the initial question was posed, it seems likely that the memory representation of the target phrase had gained in coherence and detail. For example, if we take the initial phrase (bracket 1) in Fig. 10.8 as a target, it could be tested after a 5-s delay with the phrase in bracket 3 as a similar lure. The melody in the lure has the same contour and rhythm as in the target, the accompaniment patterns are very similar, and both are in the same key. The main difference is that the melody in bracket 3 is attached to the tonal scale in a different place from the melody in bracket 1; the former centers on the third degree of the scale (B), whereas the latter centers on the fifth degree of the scale (G). Otherwise these two phrases match in key, melodic contour, and rhythm. It appears that the response given after 5 s is based on those feature matches, and that the response after 15 s takes account of the relationship of the contour to the scale. I think the development during the additional delay of a richer and more detailed representation of the phrase is responsible for the change in responses. This is an example of the answer to the question 'What did you perceive?' depending on when the listener is asked.

Once a coherent, meaningful representation of a phrase is formed, it can be stored in a relatively permanent form, to be called up into working memory when needed for comparison with other phrases, manipulated, or used in thinking about the piece and its structure (Winkler and Cowan, 2005; Treisman, 2006). Proust, an avid student of the psychology in the 1890s, has provided an excellent introspective account of the initial stages of cognitive processing of a melodic phrase. He begins by describing the initial experience of vague fleeting shapes that 'have vanished before these sensations are well-enough formed in us to avoid being submerged by following ... notes.' He continues: 'It was as if memory, like a worker striving to erect a solid foundation in the midst of a flood, while making facsimiles of these fleeting phrases, would not allow us to compare them to those that follow, and to differentiate them.' But shortly afterward, the listener's 'memory gave him a provisional and summary transcript of it, even while he continued listening. He took a good enough look at the transcript while the piece continued, so that when the same impression suddenly returned, it was no longer impossible to grasp' (Proust, 1999, p. 173, as cited in Dowling *et al.*, 2001, pp. 272–3).

10.4.1 The perceptual/memory representation

The content of our experience is a memory representation at some stage of encoding. Hence the question arises concerning the content of that representation: what features are encoded and

brought to mind? Some 20 years ago our answer would have been that the representation of a melody is fairly abstract. Clearly what allows us to identify ‘Twinkle, Twinkle’ is a pattern of pitches that can be transposed to start on any note; when we sing ‘Happy Birthday’ it is rarely in the same key as the time before. We thought that what we remember consists of a melodic and rhythmic contour attached to a tonal scale in the appropriate mode (major or minor; Dowling, 1978). However, we have now learned that when the tune is one that we have always heard in the same key, our memory is quite literal, and the tune is represented much as we have always perceived it. Levitin (1994) discovered this in an experiment in which he brought students into a lab with a display of currently popular CDs on the wall. He had them pick out their favorite album, and then think of their favorite song on that album, and then sing it. Levitin found that they sang the song within a semitone or two of the key of the original. Moreover, they sang it at more or less the original tempo (Levitin and Cook, 1996). It seems likely that in the early stages of encoding the pop songs, the contour and tonal scale were encoded as separate features, and then combined, as in the Dowling *et al.* (2001) experiment described above. Levitin’s result suggests that the scale in question is not just the general pattern of the major scale, for example, but fixed at least to some degree in its pitch level. It may take many repetitions for the scale to be absolutely fixed, however, since Dowling and Tillmann (in preparation), in a replication of Dowling *et al.* (2001), varied the key of the test item up or down 1 semitone without a noticeable effect on recognition performance. We do know that varying the pitch of the test item by much more than a semitone is disruptive of performance in recognizing novel melodies (Dowling and Fujitani, 1971).

It is tempting to suppose that the pitch pattern of a melody is initially encoded in terms of its contour plus the exact sizes of intervals from note to note. This view is widely expressed in the music cognition literature. In this view the results of Dowling *et al.* (2001) described above would mean that when tested after a 5-s delay listeners base their responses on contour alone, but at 15 s respond in terms of both contour and interval sizes. However, when the features of interval pattern and tonal framework are separately controlled at brief delays, the tonal framework has strong effects, which would not be the case if the initial encoding were simply in terms of contour, or of contour plus interval sizes, without regard to the scale framework (Bartlett and Dowling, 1980). Furthermore, there is converging evidence that melody recognition occurs in situations where interval sizes have been eliminated as cues, as in the interleaved melodies of Fig. 10.5(b) (Dowling, 1973), and in a similar experiment in which the pitches of a familiar melody are randomly assigned to different octaves (that is, ‘Twinkle, Twinkle’ would still consist of C-C-G-G-A-A-G as in Fig. 10.2, but the Cs and Gs, etc., would all be in different octaves; Dowling and Hollombe, 1977). There is also evidence that intervals are very difficult to identify out of context, and are remembered in terms of familiar tunes that feature them, rather than the tunes being remembered in terms of the intervals (Smith *et al.*, 1994). And the tonal hierarchy, including the dynamic tendencies of pitches in the scale noted above, operates in terms of pitch classes, and not intervals. That is, in Fig. 10.2, when we stop on the next-to-last note of ‘Twinkle, Twinkle’ the pull we feel toward the tonic is due to the fact that the next-to-last note is the second degree of the scale, and not to the fact that it is 2 semitones above the tonic and 2 semitones below the third degree. The seventh degree of the scale pulls upward to the tonic, but not because it is 1 semitone lower; the third degree is 1 semitone below the fourth, and doesn’t pull upward in the same way. The expectancies we have of where a melody is going are based on the tonal values of the pitches, and not on the interval pattern. I believe the evidence shows that we remember melodies in terms of a contour attached to a scale framework, and the contour and scale are among the principal perceptual features that we use in our early encoding of the melody into the representation we experience.

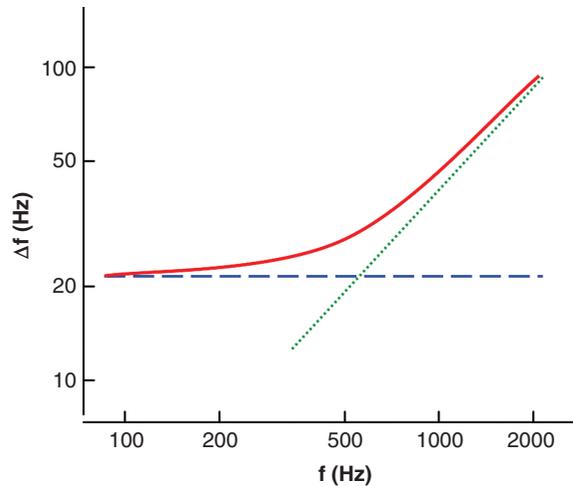
10.5 Consonance and dissonance

The issue of consonance and dissonance is closely related to the rise and fall of musical tension discussed above. We usually distinguish two types of consonance and dissonance in music: tonal or acoustic, and aesthetic. Acoustic dissonance is produced by the interaction of two or more simultaneous tones when they enter the auditory system, and creates a state of tension that can be resolved by moving to a more acoustically consonant combination of tones. Acoustic dissonance can characterize a single simultaneous tone cluster presented out of context. Aesthetic dissonance depends on musical context, and involves the very sources of instability and tension discussed above in relation to the tonal hierarchy. A single tone (like the next-to-last note of ‘Twinkle, Twinkle’) can be aesthetically dissonant if it is unstable and requires resolution to another tone, but it cannot be acoustically dissonant. In practice, European composers use acoustic dissonance and aesthetic dissonance in tandem, so that the least stable points in a progression of chords are also usually the most acoustically dissonant. For example, in measure 6 of the Beethoven sonata movement (Fig. 10.1), the harmony contains the pitches F-A-C-Eb-G—including four adjacent pitches in the scale (Eb-F-G-A)—which is a very acoustically *and* aesthetically dissonant combination. This occurs at the peak of tension in his melody, which gradually subsides into the cadence reaching the tonic in measure 9.

The acoustic dissonance of a pair of tones is not directly related to the musical interval between them. If it were, the function relating the size of the most dissonant interval to the frequency level of the tones would be a straight line with a positive slope representing the proportional frequency change of the interval, like the dotted line in Fig. 10.9. (Note that all equal musical intervals in equal temperament represent equal frequency ratios—1.05946 for the semitone, 2.0 for the octave, etc.) A second hypothesis, suggested by Helmholtz (1877/1954), is that dissonance is caused by beats between adjacent tones with a frequency difference around 40 Hz, which produces a very rough sensation. In that case, the most dissonant intervals would be represented by a constant difference between frequencies, shown in the horizontal dashed line in Fig. 10.9. (Note that all these considerations apply to dissonance between pairs of sine waves—pure tones. We will consider complex tones below.) Plomp and Levelt (1965) found that neither of these hypotheses is correct. The curve they found (the solid line in Fig. 10.9) follows the constant musical interval rule in the upper register, but flattens out in the lower register. It follows the curve for one-quarter of the critical bandwidth, a parameter of the auditory system summarizing a large amount of converging evidence concerning the interaction and mutual interference of adjacent tones (see Chapter 2). One consequence of the shape of the curve is that in the middle and upper registers of music it makes sense to think of some intervals (like perfect fourths and fifths) as being quite consonant across a wide range of frequencies, and of other intervals (like those of 1 and 2 semitones) as being fairly dissonant, but that this rule will not hold in the lower register. As we go lower, the proportionate size of the most dissonant frequency interval becomes greater, so that even thirds and fourths become quite dissonant. (You can imagine that if we transpose Bach chorales for a tuba quartet to play in their register, the result will be horrendous, even with excellent tuba players, because of the normally consonant thirds and fourths that abound in those pieces.) In practice, this leads composers to write with generally larger intervals between simultaneous bass notes than between the upper notes, as Plomp and Levelt documented in their report (and which is evident in Fig. 10.1).

Musical sounds are generally not pure tones, but rather harmonic complex tones, with a fundamental sine-wave frequency (which corresponds to the pitch) and a series of sine-wave harmonics (or ‘overtones’) at frequencies that are integer multiples of the fundamental. When two complex tones sound together, the dissonance produced is the sum of the dissonance produced

Fig. 10.9 The relationship of the most dissonant frequency interval (Δf) between two tones and the mean frequency (f) of the tones (solid line), which is approximately one-fourth of the critical bandwidth. The dotted line shows the results to be expected on the basis of a constant musical interval, and the dashed line shows a constant frequency difference (beat frequency).



by all the adjacent pairs of harmonics in the combination. There are two consequences of this that we should consider. First, combinations of complex tones will be less dissonant if their fundamental frequencies are small integer multiples of some common divisor, than if their relationships are more complicated. For example, consider two complexes in a $3/2$ ratio: an E with components at 330, 660, 990, 1320, and 1650 Hz, and an A with components at 220, 440, 660, 880, 1100, and 1320 Hz. Either the adjacent components are sufficiently far apart not to interfere with each other (220, 330 and 440, etc.), or they coincide (at 660 and 1320). Not much dissonance is generated, and the combination, a perfect fifth, is consonant. But if we substitute an A# at 233, 466, 699, 932, and 1165 Hz, we get numerous clashes at 660–699, 1100–1165, etc., and the combination is quite dissonant. This is why *a capella* choirs and string quartets in slow passages tune their chords with simple whole number ratios rather than in equal temperament (where the perfect fifth has a ratio of 1.498 rather than the 1.5 of our example). The coincidence of the harmonics gives the chords a clarity and brilliance that they do not have in equal tempered tuning.

The second consequence of the interaction of adjacent harmonics concerns the dissonances in measures 9 and 10 of Fig. 10.1. Those dissonances (with pitches F-A-C-Eb-G-Bb, six of the seven pitches in the scale, simultaneously present or implied) do not stand out for the listener in the same way that they would if they all occurred clustered together in one octave, where they would have generated very strong acoustic dissonances. Because of the voicing of the chords (the distribution of pitches across several octaves), the harmonic components of the clashing notes are spread out, and not as much acoustic dissonance is generated as might have been. The aesthetic dissonance is subtle by comparison, and the eight-measure melody goes through changes in tension, but the effect is not at all jarring.

10.6 Timbre

Timbre involves qualities of sound apart from pitch and consonance. It is often called ‘tone color’ (*Klangfarbe* in German), and can be used to distinguish musical instruments and speech sounds. We can identify our favorite recordings within a tenth of a second on the basis of their timbre and texture (which instruments are playing in which registers, and how fast; Schellenberg *et al.*, 1999).

The psychophysics of timbre is multidimensional; that is, sound qualities cannot be arranged on a single dimension going from less to more. The cues involved in timbre perception consist of steady-state cues such as those that distinguish vowel sounds in speech, and transient cues, mostly

in the first 50 ms after the onset of a sound, that distinguish initial consonants in speech (Dowling and Harwood, 1986). The steady-state cues depend largely on frequency regions in which the harmonics are strong due to resonances in the vocal tract or musical instrument, and are themselves multidimensional (Ladefoged and Broadbent, 1957; Slawson, 1968). Instruments are easily confused, however, if all the listener has are the steady-state cues. Saldanha and Corso (1964) demonstrated this by eliminating the onset transients from recordings of musical instruments, and found that discrimination among instruments as different as violins and trombones became very difficult when they were all playing the same pitch. Iverson and Krumhansl (1993) refined this result, finding that for the most part the same kinds of transient and steady-state cues are present in the onset of a tone and in the continuation of a tone as are present in the complete stimulus from beginning to end.

A major puzzle in the study of timbre perception has been the phenomenon of timbre constancy. The clarinet, for example, has different resonances and hence different steady-state cues in its upper and lower register, and yet it still sounds like a clarinet. But it seems likely that this is mainly true for musicians who have had years of experience hearing (or playing) clarinets in bands or orchestras. Steele and Williams (2006) had listeners distinguish a bassoon and a French horn, two instruments that Iverson and Krumhansl (1993) had found were easily confusable when they were playing in different octaves as well as playing the same pitch. Non-musicians had difficulty with this task, and their performance fell to chance when the pitch separation approached 2 octaves, whereas musicians' performance remained above 80% correct at 2.5 octaves separation. Timbre constancy clearly improves with perceptual learning.

10.7 Other topics

Several topics in the area of music perception have not been explored here for reasons of space. For issues concerned with loudness and sound localization the reader is referred to Dowling and Harwood (1986) and chapters in Deutsch (1999), as well as Bregman's (1990) book on auditory scene analysis that puts those issues into broader context. For music and the emotions, chapters in Juslin and Sloboda's (2001) volume provide broad, detailed coverage.

10.8 Summary

Music presents us with patterns of sound in a virtual space of pitch and time. Salient points on those dimensions—the tonal scale in pitch and the beat in time—give us a framework to track the organization of what we hear. We sense the degree of tension or relaxation in the music in relation to the stability of the pitches in the tonal and temporal frameworks. That instability can be enhanced by the addition of acoustic dissonance. Our familiarity with a genre of music provides expectancies by which we guide our attention to important aspects of the musical pattern. The distinctive timbres of the various voices and instruments also provide cues to musical organization; we can follow a particular melodic line more easily if its timbre contrasts with those in the background. The complexity of music and the extensive perceptual learning and acculturation involved in listening to and understanding it make music a fertile domain in which to study human cognition.

References

- Bartlett, J. C. and Dowling, W. J. (1980). Recognition of transposed melodies: A key-distance effect in developmental perspective. *Journal of Experimental Psychology: Human Perception and Performance* 6:501–15.
- Bharucha, J. J. (1999). Neural nets, temporal composites, and tonality. In *The Psychology of Music* (ed. D. Deutsch), pp. 413–40. San Diego: Academic Press.

- Bharucha, J. J. and Pryor, J. H. (1986). Disrupting the isochrony underlying rhythm: An asymmetry in discrimination. *Perception and Psychophysics* 40:137–41.
- Bigand, E., Parncutt, R., and Lerdahl, F. (1996). Perception of musical tension in short chord sequences: The influence of harmonic function, sensory dissonance, horizontal motion, and musical training. *Perception and Psychophysics* 58:125–41.
- Bigand, E., Madurell, F., Tillmann, B., and Pineau, M. (1999). Effect of global structure and temporal organization on chord processing. *Journal of Experimental Psychology: Human Perception and Performance* 25:184–97.
- Bregman, A. S. (1990). *Auditory Scene Analysis*. Cambridge, MA: MIT Press.
- Brunswik, E. (1956). *Perception and the Representative Design of Psychological Experiments*. Berkeley: University of California Press.
- Brunswik, E. (2001). *The Essential Brunswik* (ed. K. R. Hammond and T. R. Stewart). Oxford: Oxford University Press.
- Deutsch, D. (ed.) (1999). *The Psychology of Music*. San Diego: Academic Press.
- Dowling, W. J. (1973). The perception of interleaved melodies. *Cognitive Psychology* 5:322–37.
- Dowling, W. J. (1978). Scale and contour: Two components of a theory of memory for melodies. *Psychological Review* 85:341–54.
- Dowling, W. J. (1992). Perceptual grouping, attention and expectancy in listening to music. In *Gluing Tones: Grouping in Music Composition, Performance and Listening* (ed. J. Sundberg), pp. 77–98. Stockholm: Royal Swedish Academy of Music.
- Dowling, W. J. and Fujitani, D. S. (1971). Contour, interval, and pitch recognition in memory for melodies. *Journal of the Acoustical Society of America* 49:524–31.
- Dowling, W. J. and Harwood, D. L. (1986). *Music Cognition*. Orlando, FL: Academic Press.
- Dowling, W. J. and Hollombe, A. W. (1977). The perception of melodies distorted by splitting into several octaves: Effects of increasing proximity and melodic contour. *Perception and Psychophysics* 21:60–4.
- Dowling, W. J. and Tillmann, B. Memory improvement while hearing music: Effects of structural continuity on feature binding. (In preparation.)
- Dowling, W. J., Lung, K. M.-T., and Herrbold, S. (1987). Aiming attention in pitch and time in the perception of interleaved melodies. *Perception and Psychophysics* 41:642–56.
- Dowling, W. J., Tillmann, B., and Ayers, D. (2001). Memory and the experience of hearing music. *Music Perception* 19:249–76.
- Dowling, W. J., Bartlett, J. C., Halpern, A. R., and Andrews, M. W. (2008). Melody recognition at fast and slow tempos: Effects of age, experience, and familiarity. *Perception and Psychophysics* 70:496–502.
- Fraisse, P. (1982). Rhythm and tempo. In *The Psychology of Music* (ed. D. Deutsch), pp. 149–80. New York: Academic Press.
- Francès, R. (1988). *The Perception of Music* (transl. W. J. Dowling). Hillsdale, NJ: Erlbaum. [Original work published in 1958.]
- Gabrielsson, A. and Lindström, E. (2001). The influence of musical structure on emotional expression. In *Music and Emotion: Theory and Research* (ed. P. N. Juslin and J. A. Sloboda), pp. 223–48. Oxford: Oxford University Press.
- Hannon, E. E. and Trehub, S. E. (2005). Metrical categories in infancy and adulthood. *Psychological Science* 16:48–55.
- Helmholtz, H. L. F. von. (1877/1954). *On the Sensations of Tone* (transl. A. J. Ellis). New York: Dover.
- Inhelder, B. and Piaget, J. (1958). *The Growth of Logical Thinking: From Childhood to Adolescence* (transl. A. Parsons and S. Milgram). New York: Basic Books.
- Iverson, P. and Krumhansl, C. L. (1993). Isolating the dynamic attributes of musical timbre. *Journal of the Acoustical Society of America* 94:2595–603.
- James, W. (1890). *Principles of Psychology*. Boston: Henry Holt.

- Janata, P., Birk, J. L., Van Horn, J. D., Leman, M., Tillmann, B., and Bharucha, J. J. (2002). The cortical topography of tonal structures underlying Western music. *Science* 298:2167–70.
- Janata, P., Birk, J. L., Tillmann, B., and Bharucha, J. J. (2003). Online detection of tonal pop-out in modulating contexts. *Music Perception* 20:283–305.
- Jones, M. R., Sommerell, L., and Marshburn, E. (1987). Recognizing melodies: A dynamic interpretation. *Quarterly Journal of Experimental Psychology* 39A:89–121.
- Juslin, P. N. and Sloboda, J. A. (eds) (2001). *Music and Emotion: Theory and Research*. Oxford: Oxford University Press.
- Krumhansl, C. (1990). *Cognitive Foundations of Musical Pitch*. New York: Oxford University Press.
- Kubovy, M. and Van Valkenburg, D. (2001). Auditory and visual objects. *Cognition* 80:97–126.
- Ladefoged, P. and Broadbent, D. E. (1957). Information conveyed by vowels. *Journal of the Acoustical Society of America* 29:98–104.
- Lao-Tse (1972). *Tao Te Ching* (trans. G.-F. Feng and J. English). New York: Vintage.
- Large, E. W. and Jones, M. R. (1999). The dynamics of attending: How we track time-varying events. *Psychological Review* 106:119–59.
- Levitin, D. J. (1994). Absolute memory for musical pitch: Evidence from the production of learned melodies. *Perception and Psychophysics* 56:414–23.
- Levitin, D. J. and Cook, P. R. (1996). Memory for musical tempo: Additional evidence that auditory memory is absolute. *Perception and Psychophysics* 58:927–35.
- Meyer, L. B. (1956). *Emotion and Meaning in Music*. Chicago: University of Chicago Press.
- Partch, H. (1974). *Genesis of a Music*. New York: Da Capo.
- Plomp, R. and Levelt, W. J. M. (1965). Tonal consonance and critical bandwidth. *Journal of the Acoustical Society of America* 38:548–60.
- Povel, D. J. and Essens, P. (1985). Perception of temporal patterns. *Music Perception* 8:411–40.
- Proust, M. (1999). *À la Recherche du Temps Perdu* (edn in 1 vol.). Paris: Gallimard. [Original ref.: Proust, M. (1913). *Du côté de chez Swann*. Paris: Grasset.]
- Repp, B. (2002). Automaticity and voluntary control of phase correction following event onset shifts in sensorimotor synchronization. *Journal of Experimental Psychology: Human Perception and Performance* 28:410–30.
- Saldanha, E. L. and Corso, J. F. (1964). Timbre cues and the identification of musical instruments. *Journal of the Acoustical Society of America* 36:2021–6.
- Schaeffer, P. (1952). *À la Recherche d'une Musique Concrète*. Paris: Éditions du Seuil.
- Schellenberg, E. G., Iverson, P., and McKinnon, M. C. (1999). Name that tune: Identifying popular recordings from brief excerpts. *Psychonomic Bulletin and Review* 6:641–6.
- Schönberg, A. (1967). *Fundamentals of Music Composition*. New York: St Martins Press.
- Shepard, R. N. (1982). Musical pitch. In *The Psychology of Music* (1st edn; ed. D. Deutsch), pp. 343–90. San Diego: Academic Press.
- Slawson, W. (1968). Vowel quality and musical timbre as functions of spectrum envelope and fundamental frequency. *Journal of the Acoustical Society of America* 43:87–101.
- Sloboda, J. (1998). Does music mean anything? *Musicae Scientiae* 2:21–31.
- Smith, J. D., Nelson, D. G. K., Grohskopf, L. A., and Appleton, T. (1994). What child is this? What interval was that? Familiar tunes and music perception in novice listeners. *Cognition* 52:23–54.
- Steele, K. M. and Williams, A. K. (2006). Is the bandwidth for timbre invariance only one octave? *Music Perception* 23:215–20.
- Toiviainen, P. and Krumhansl, C. L. (2003). Measuring and modeling real-time responses to music: The dynamics of tonality induction. *Perception* 32:741–66.
- Treisman, A. (2006). Object tokens, binding, and visual memory. In *Handbook of Binding and Memory: Perspectives from Cognitive Neuroscience* (ed. H. D. Zimmer, A. Mecklinger, and U. Lindenberger), pp. 315–39. Oxford: Oxford University Press.

- Vines, B. W., Krumhansl, C. L., Wanderley, M. M., and Levitin D. J. (2006). Cross-modal interactions in the perception of musical performance. *Cognition* **101**:80–113.
- Warren, R. M., Gardner, D. A., Brubaker, B. S., and Bashford, J. A., Jr (1991). Melodic and nonmelodic sequences of tones: Effects of duration on perception. *Music Perception* **8**:277–90.
- Winkler, I. and Cowan, N. (2005). From sensory to long-term memory: Evidence from auditory memory reactivation studies. *Experimental Psychology* **52**:3–20.
- Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language* **46**:441–517.