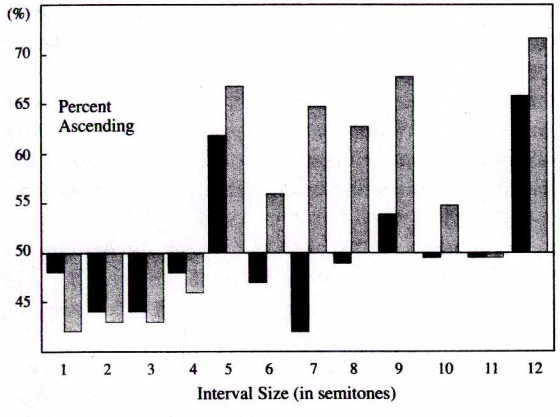

Figure 5.2
Proportion of non-unison melodic intervals that ascend in pitch. Dark bars: sample of thirteen Western composers. Light bars: sample of Albanian, Bulgarian, Iberian, Irish, Macedonian, Norwegian, and African-American folk songs. From Vos and Troost 1989.

In my laboratory we have replicated this asymmetry for many other cultures. The tendency for large intervals to ascend is evident in Australian aboriginal music, Chinese folk songs, traditional Korean music, Ojibway, Pondo, Venda, and Zulu songs. Of the many musical samples we've examined over the years, only a small sample of north Indian classical music failed to display this feature.[8]

There is another way of interpreting this asymmetry between ascending and descending large intervals. Melodies tend to meander around a central pitch range—what goes up must come down (and vice versa). If most large intervals go up, then it necessarily follows that most small intervals will go down. And, as we have already seen, most melodic intervals are small rather than large. So instead of saying that most large intervals ascend, we might simply say that the majority of melodic movements are descending small intervals.

Ethnomusicologist Curt Sachs long ago noted the tendency of certain cultures to produce what he called *tumbling* melodies. Tumbling melodies are dominated by phrases that typically begin with an initial ascending leap and are followed by a series of steplike descending tones. Sachs speculated that the tumbling phrase derives from a sort of impassioned "howl" or "wail." These sorts of descending lines are plainly evident in a number of repertories—for example, in Russian laments, in Australian aboriginal music, and in Lakota (Sioux) music.[9]

The tumbling strain is very similar to descending pitch patterns that are commonplace in speech. Researchers who study the pitch contours of speech ("intonation") have observed that the initial part of an utterance tends to ascend rapidly, and then the pitch of the voice slowly drops as the utterance progresses. There are a few exceptions, such as the "up-speak" apparent in the speech of some American adolescents, as well as in the final syllables of some interrogative forms. But in general, the pitch of the speaking voice tends to descend. In the field of linguistics, this ubiquitous prosodic pattern is known as *declination.*[10] The origin of pitch declination in speech is thought to be the drop in subglottal air pressure as the air supply in the lungs is exhausted.

Since singing requires sustained control of pitch height, it is unlikely that the tumbling melodic phrase pattern originates in the loss of subglottal air pressure. However, given the similarity between the pitch patterns in speech and the prevalence of descending small intervals, we might borrow the linguistic term and dub the musical phenomenon *step declination.*

Having identified this widespread asymmetry of occasional ascending leaps and frequent descending steps, we might ask whether experienced listeners come to expect this pattern. We'll postpone discussing the pertinent experimental evidence until our discussion of melodic regression (see below).

3 Step Inertia

Leonard Meyer suggested that small pitch intervals (1 or 2 semitones) tend to be followed by pitches that continue in the same direction. Paul von Hippel has coined the useful term *step inertia* to refer to this property of melodic organization. Music theorist Eugene Narmour (a former student of Meyer) has suggested that listeners form such "step inertia" expectations, and has even suggested that these expectations might be based on innate dispositions.[11]

The first question to ask is whether melodies themselves are actually organized according to step inertia. Is it the case that most small pitch intervals tend to be followed by pitch contours that continue in the same direction? The answer to this question is a qualified yes. Von Hippel examined a large sample of melodies from a broad range of cultures. His results are shown in table 5.1. Von Hippel found that only descending steps tend to be followed by a continuation in the same direction. Roughly 70 percent of descending steps are followed by another descending interval. However, in the case of ascending steps, no trend is evident. When an ascending step occurs, melodies are as likely to go down as to continue going up.

But what about listeners' expectations? Do listeners expect a step movement to be followed by a pitch movement in the same direction? Two critical experiments were

Table 5.1
Probabilities for step–step movements in a large sample of Western and non-Western musics.

	Followed by ascending step	Followed by descending step
Initial descending step	30%	70%
Initial ascending step	51%	49%

carried out in my laboratory—the first by Paul von Hippel and the second by Bret Aarden.[12] Von Hippel's experiment used randomly contrived pitch sequences whereas Aarden's experiment used actual melodies. Von Hippel's listeners heard a randomly generated twelve-tone row. This was done to minimize the possible confounding influence of tonality-related expectations. After hearing the twelve tones, listeners were asked to indicate whether they expected the next (13th) note in the sequence to be higher or lower than the last pitch heard. Since the sequences were random, there is no "correct" response to this question; von Hippel simply looked at the interval formed by the last two notes in the sequence. If the last two notes formed a descending step, were listeners more likely to say the ensuing note would be lower? If the last two notes of the sequence formed an ascending step, were listeners more likely to expect the ensuing note to be higher? Von Hippel's results showed that listeners do indeed expect descending steps to be followed by another descending interval. Surprisingly, listeners also expect ascending steps to be followed by another ascending interval.

In a subsequent reaction time study by Aarden, listeners were asked to judge whether the pitch in a folksong melody went up, down, or remained the same. Aarden's results fully replicated von Hippel's earlier study. After hearing a step interval, listeners respond more quickly and accurately when the ensuing note moves in the same melodic direction. It doesn't matter whether the pitch sequence is ascending or descending.

These results are a nice vindication of Meyer's and Narmour's intuitions about step inertia. In light of Narmour's suggestion that step inertia might be innate, the results also seem to go against the statistical learning theory of expectation. Real melodies exhibit a tendency for step inertia *only* for descending intervals. If expectations are formed by apprehending statistical regularities in the music, then why do listeners expect step inertia for both ascending and descending contexts?

In response to this problem, Paul von Hippel has suggested that listeners tend to overgeneralize in forming their melodic expectations. Notice that since ascending steps have a fifty–fifty chance of going in either direction, there is no penalty for (wrongly) assuming that ascending steps should typically continue to go up. That is, for ascending contours, the expectation for step inertia is no worse than chance. Since

the strategy of expecting step inertia pays off for descending intervals, listeners who always form a step-inertia expectation will still, on average, have more accurate expectations than a listener who has no step-inertia expectation.

Furthermore, if ascending and descending steps were equally prevalent, then a step-inertia expectation would prove correct in just over 60 percent of cases. But ascending steps account for only about 42 percent of all step motions. This further reduces the penalty for wrongly expecting that an ascending step is likely to continue in the same direction. On average, a step-inertia expectation will prove correct in roughly 62 percent of cases. Interestingly, if listeners relied on the "correct" heuristic and expected step inertia only for descending intervals, then the proportion of correct predictions would be the same—62 percent. That is, there is no practical difference between expecting step inertia only in the descending case, and expecting step inertia in both the ascending and descending cases. It would seem that listeners who form expectations based on the step-inertia heuristic are performing near the optimum level, even though they are employing the wrong rule.

Notice, moreover, that the same predictive accuracy would occur if listeners simply assumed that melodies tend to descend. (Rule: Always expect the next pitch to be lower.) So why do listeners form a step-inertia expectation rather than a pitch-descent expectation? One plausible answer goes as follows: Incorrect heuristics are most likely to be revised or discarded when falsifying instances are obvious. Melodic leaps are more perceptually salient or noticeable than steps. Most large intervals ascend in pitch. Therefore, each occurrence of an ascending leap represents a salient observation that contradicts the general inference that pitches tend to descend whether by steps or by leaps.

A possible objection that can be levied against this account is that it is contradicted by another plausible scenario. Most intervals are descending steps. Most large intervals ascend in pitch. On average, this means that most ascending large intervals are preceded by a descending step. Melodic leaps are more perceptually salient. Therefore, each occurrence of an ascending leap preceded by a descending step represents a salient observation that contradicts the inference that descending steps are followed by another descending interval.

Notice, however, that in the first scenario, *all* ascending leaps are falsifying observations, whereas in the second scenario, only the *majority* of ascending leaps are falsifying observations. Although this difference might not seem convincing, it might nevertheless explain why listeners come to favor the step-inertia heuristic over the descending-pitch heuristic.

There is one further finding from von Hippel's experiment that must be mentioned. Von Hippel tested both musician and nonmusician listeners and found step-inertia expectations only for the musician participants. The nonmusicians had no discernible pattern related to step-interval antecedents. This result raises problems for the idea

that step-inertia may be innate, since if it were innate, one might expect to see it operating in all listeners. Although other explanations might account for this finding, the greater musical experience of musicians provides a plausible source for this difference—with the implication that learning plays the formative role.

4 Melodic Regression

We have seen that listeners expect melodies to consist mostly of small pitch intervals. Experienced listeners also expect that small intervals tend to be followed by pitches that preserve the melodic direction—although melodies exhibit step inertia only for descending intervals. What about expectations for what follows large intervals?

Since at least the sixteenth century, music theorists have observed that large intervals tend to be followed by a change of direction. Most of the theorists who have commented on this phenomenon have further suggested that large intervals tend to be followed by step motion in the opposite direction. Since most pitch intervals are small, any interval should tend to be followed by step motion. The important part of the claim is the idea that large leaps should be followed by a *change* of direction. Following Paul von Hippel, we can call this purported tendency *post-skip reversal.*[13]

Once again, the first question to ask is whether actual melodies conform to this principle. Do most large leaps tend to be followed by pitches that change direction? In 1924, Henry Watt tested this idea by looking at melodic intervals in musical samples from two different cultures: Lieder by Franz Schubert and Ojibway songs. Watt's results for Schubert are shown in figure 5.3. For intervals consisting of 1 or 2 semitones, roughly 25 to 30 percent of contours change direction. That is, the majority of small intervals continue in the same direction. However, as the interval size increases, the graph tends to rise upward to the right. For octave (12 semitone) intervals, roughly 70 percent of intervals are followed by a change of direction. (There is no data point corresponding to 11 semitones because there were no 11-semitone intervals in Watt's sample.) Watt found similar results for the Ojibway songs.

Paul von Hippel and I carried out further tests of this idea using a broader and more diverse sample of melodies from cultures spanning four continents: traditional European folk songs, Chinese folk songs, South African folk songs, and Native American songs. For each of these repertories we replicated Watt's finding: the majority of large intervals are indeed followed by a change of direction.[14]

Paul and I proposed a rather unexciting reason for the existence of post-skip reversal, namely, *regression to the mean.* Statisticians have shown that whenever a distribution exhibits a central tendency, successive values tend to "regress toward the mean." That is, when an extreme value is encountered, the ensuing value is likely to be closer to the mean or average value. For example, when rolling a pair of dice, the highest possible combination is twelve (two sixes) and the lowest combination is two ("snake eyes"). The most common outcome is the number seven. If you roll a pair of dice with an outcome of, say, ten, the likelihood is that the next roll will be lower than ten. Similarly, if you roll an outcome of three, the next roll is likely to be higher. That is, successive values tend to "regress" toward the mean.

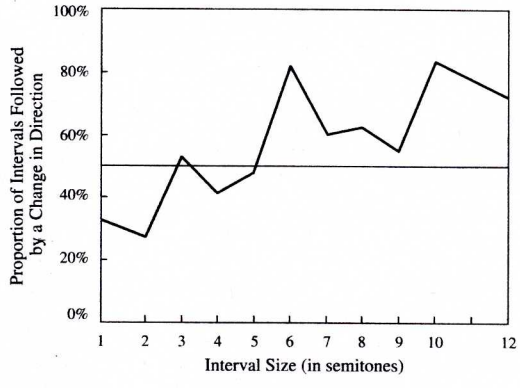

Figure 5.3
Watt's (1924) analysis of intervals in Schubert Lieder. Larger intervals are more likely to be followed by a change of melodic direction than small intervals. Watt obtained similar results for Ojibway songs. No data point corresponds to eleven semitone intervals because of the absence of such intervals in Watt's sample. From von Hippel and Huron 2000.

A similar phenomenon could conceivably occur with melodies. In general, most large intervals tend to take the melody toward the extremes of the melody's range. For example, a large ascending leap has a good probability of placing the melody in the upper region of the tessitura or range. Having landed near the upper boundary, a melody has little choice but to go down. That is, most of the usable pitches lie below the current pitch. Similarly, most large descending leaps will tend to move the melody near the lower part of the range, so the melody is more likely to ascend than to continue descending.

Another analogy might help to illustrate this point: When you encounter a very tall person, the next person you encounter is likely to be shorter. But the shorter person is not "caused" by the previous encounter with a tall person. It is simply a consequence of the fact that most people are near average height. There is no "force" or "magnet" drawing values toward the mean. Regression to the mean is simply a numerical artifact—a necessary consequence of the fact that most values lie near the center of some distribution.

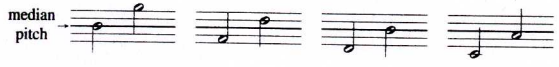

Figure 5.4
Four hypothetical interval relationships relative to the median (or average) pitch (represented by the bold central line): (1) median-departing leap, (2) median-crossing leap, (3) median-landing leap, and (4) median-approaching leap. See also figure 5.5.

Like human heights, the distribution of pitches in melodies exhibits a central tendency. Melodies do not simply wander around in an unbounded pitch space. Melodies also display a stable range or tessitura. The most frequently occurring pitches in a melody lie near the center of the melody's range. Pitches near the extremes of the range occur less commonly. This makes melodies a candidate for regression to the mean.

If post-skip reversal were merely a consequence of regression to the mean, then we ought to see a difference for leaps depending on where they occur in the range. Consider the ascending intervals shown in figure 5.4. In this schematic illlustration, the mean or median pitch for the melody is represented by the bold center line in the staff. The first ascending leap takes the contour above the median. Both regression to the mean and post-skip reversal would predict a change of direction to follow. In the second case, the ascending leap straddles the median pitch. Once again, both regression to the mean and post-skip reversal predict a change of direction. In the third and fourth cases, the two theories make different predictions. In the third case, the leap lands directly on the median pitch. Post-skip reversal continues to predict a change of direction, whereas regression to the mean predicts that either direction is equally likely. Finally, in the fourth case, the leap lands below the median pitch. Here regression to the mean predicts that the contour should tend to continue in the same direction (toward the mean), whereas post-skip reversal continues to predict a change of direction. So how are real melodies organized? Are they organized according to post-skip reversal? Or according to regression to the mean?

In order to answer this question, we studied several hundred melodies from different cultures and different periods. For each melody we calculated the median pitch and then examined what happens following large leaps. Our results are plotted in figure 5.5 for the case where a "skip" is defined as intervals larger than two semitones. The black bars indicate instances where an interval is followed by a change of direction. The white bars indicate instances where an interval is followed by a continuation of the melody in the same direction.[15]

If post-skip reversal is the important organizing principle of melodies, then we would expect to see taller black bars than white bars in each of the four conditions.

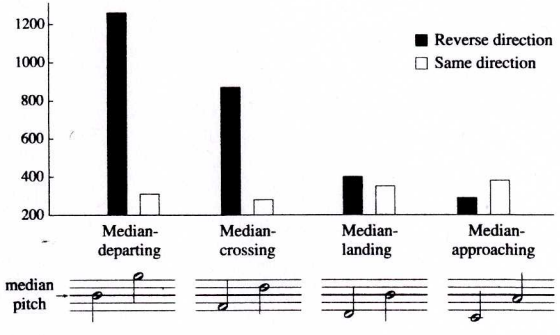

Figure 5.5
Number of instances of various melodic leaps found in a cross-cultural sample of melodies. Most large intervals that approach the median pitch continue in the same melodic direction. Large intervals that land on the median pitch are as likely to continue in the same direction as to reverse direction. Results support the phenomenon of melodic regression and fail to support post-leap reversal. From von Hippel and Huron 2000.

By contrast, consider regression to the mean. This would predict that black bars should be taller than white bars for the median-departing and median-crossing conditions (which is the case). For skips that land on the median pitch, regression to the mean would predict roughly equivalent numbers of continuations and reversals (that is, we would expect the black and white bars to be roughly the same height). Finally, in the case of median-approaching skips, regression to the mean would predict that melodies ought to be more likely to continue in the same direction toward the mean (that is, we would expect the white bar to be taller than the black bar). The results shown in figure 5.5 are clearly more consistent with regression to the mean than with post-skip reversal.

Paul von Hippel carried out further statistical analyses that reinforce the above result. With regard to large intervals, melodies behave according to regression to the mean and are not consistent at all with the idea of post-skip reversal. The further the leap takes the melody away from the mean pitch, the greater the likelihood that the next pitch will be closer to the mean. If a leap takes the melody toward the mean, then the likelihood is that the melody will continue in the same direction. Incidentally, we tried a number of different definitions of "large" leap. The results are the same no matter how a leap is defined in terms of size. We also looked for possible "delayed" resolutions. That is, we looked to see whether the second or third note following a large leap tended to change direction. Once again, the aggregate results always conformed to regression to the mean, and not to post-skip reversal. This was true in

Schubert, in European folk songs, in Chinese folk songs, in sub-Saharan African songs, and in traditional Native American songs.

We also entertained the possibility that melodies might be organized according to some combination of regression to the mean and post-skip reversal. Perhaps regression to the mean accounts for (say) 70 percent of the resolution following large intervals, and post-skip reversal accounts for the remaining 30 percent. To test this possibility we carried out a multiple regression analysis. After the effect of regression to the mean is removed, we found that post-skip reversal accounted for *none* of the residual. Zero. The results were unequivocal.

For musicians, this finding is an attention-getter. For hundreds of years musicians have been taught that it is good to resolve a large leap with a step in the other direction. Surely at least some composers followed this advice? The statistical results from von Hippel and Huron imply that for each passage where a composer had intentionally written according to post-skip reversal, then they must have intentionally *transgressed* this principle in an equivalent number of passages. Otherwise the statistics would not work out. This led von Hippel on a quest to see if any composer's music showed genuine evidence of post-skip reversal. Using our extensive database of scores, he found one: Palestrina. Palestrina's music does exhibit evidence of post-skip reversal above and beyond the effect of regression to the mean. However, the magnitude of the effect is small. Even in Palestrina's music, the overwhelming share of contour change following a large leap is accounted for by regression to the mean. Since Palestrina himself promoted the idea of post-skip reversal, we shouldn't be surprised that he sometimes took his own advice.

Palestrina notwithstanding, these studies simply refute the idea that post-skip reversal is an organizational principle in melody. This is true not just in Western music, but also in music in many (perhaps all) of the world's cultures.

It bears reminding that most large intervals are indeed followed by a change of direction. (For skips of 3 semitones or greater, roughly 70 percent are followed by a reversal of contour.) But this is only because most large intervals tend to take the melody away from, rather than toward, the mean pitch for the melody. When looking at notated music, one finds that the most noticeable leaps are precisely those where the melody moves to an especially high or low pitch. Unless one remains aware of the relationship of the interval to the tessitura, it is easy to see how theorists might have been deceived.

Having investigated the organization of actual melodies, we might now turn to the question of what listeners expect. Even if melodies are not organized according to post-skip reversals, might it not be the case that listeners *expect* large intervals to be followed by a change of direction? Or do listeners expect the next pitch to move in the direction of the average pitch?

Once again consider my earlier analogy to people's heights. When we encounter a tall person, do we (1) expect the next person to be of average height (the "real" phenomenon) or (2) expect the next person to be shorter—an artifact of (1)? This question was answered experimentally by Paul von Hippel in my laboratory.[16] Paul played large intervals in a variety of melodic circumstances, and asked listeners to predict whether the melody would subsequently ascend or descend. The melodic contexts were arranged so that some large intervals approached the mean and other large intervals departed from the mean. If listeners' expectations are shaped by post-skip reversal, then they ought to expect all large intervals to be followed by a change of direction. However, if listeners' expectations are shaped by regression to the mean, then they ought to respond according to the register of the interval: intervals in the low register (whether ascending or descending) should be followed by higher pitches while intervals in the higher register (whether ascending or descending) should be followed by a lower pitch.

The results were clear: the register or tessitura of the interval doesn't matter. Listeners typically expect large intervals to be followed by a change of direction without regard to the location of the median pitch. That is, listeners' expectations follow the post-skip reversal principle, rather than regression to the mean. As before, these results apply only in the case of musician listeners. Von Hippel's nonmusician listeners showed no discernible pattern of responses. Although the difference between musicians and nonmusicians might suggest some sort of genetic or innate difference, a more plausible possibility is a difference due to learning, either formal training or through passive exposure.

But why would musicians' expectations follow post-skip reversal over regression to the mean? A quick glance at figure 5.5 reminds us that roughly 70 percent of all large intervals are followed by a change of direction. If listeners adopt the simple post-skip reversal heuristic, their expectations will be correct 70 percent of the time. A regression-to-the-mean heuristic would be more accurate. However, in order to use a regression rule, the listener would need to constantly be inferring the tessitura or distribution of the pitches in the melody, in order to judge whether the current pitch is relatively high or relatively low. Post-skip reversal provides a simple and efficient heuristic that serves well enough to keep the listener's expectations on track.

5 Melodic Arch

To this point we have only been considering the note-to-note organization of music. What about larger structures such as phrases or whole melodies? Does music exhibit stereotypic phrase-related patterns? And if so, do listeners form expectations that reflect such patterns?

Earlier, I mentioned Curt Sachs's notion of a *tumbling* melody where phrases tend to start on a relatively high pitch and then slowly descend via small intervals. Such falling phrases are commonplace in Australian aboriginal songs and in many Native American songs. Another popular phrase-related pattern is the so-called *melodic arch*. For centuries, music scholars have observed a general tendency for phrases to rise upward and then descend in pitch, forming an arch-shaped contour. Examples of such melodic arches include the initial phrases of "Twinkle, Twinkle, Little Star," "On Top of Old Smoky," "Itsy Bitsy Spider," and "We Wish You a Merry Christmas." The phenomenon isn't limited to Western music. In Tuvan throat singing (from central Asia), for example, nearly every phrase rises and then falls in pitch.

In Western music, not all phrases are arch-shaped. For example, both "Joy to the World" and "The Star Spangled Banner" begin with a marked descending-then-ascending contour. Is there any truth in the notion of the melodic arch? In 1996 I published a comprehensive study of phrase contours in a collection of over six thousand European folk songs. Using well-defined criteria, I had a computer classify each phrase into one of nine types: ascending, descending, concave, convex, horizontal (hovering), horizontal-ascending, horizontal-descending, ascending-horizontal, and descending-horizontal. Nearly 40 percent of the roughly ten thousand phrases analyzed were classified as convex (i.e., arch-shaped)—the most common classification. Convex contours were four times more common than concave contours, even though the classification criterion was exactly symmetrical. Ascending and descending phrases were the next most common contour types, accounting for nearly 50 percent of all phrases between them.

Interestingly, further analysis showed that ascending and descending phrases tend to be paired together (thus forming an "arch" over two phrases). Moreover, while ascending phrases tend to be followed by descending phrases, the reverse is not true: descending phrases are not more likely than chance to be followed by an ascending phrase.

The arch tendency within phrases is illustrated graphically in figures 5.6 and 5.7. Figure 5.6 shows what happens when six thousand seven-note phrases are all averaged together. The first value shows the average pitch of all the first notes in the phrases; the second value shows the average pitch of all the second notes, and so on. (Pitch heights are given in semitones above middle C.) The arch shape evident in this graph also occurs for phrases containing different numbers of notes, from five-note phrases to seventeen-note phrases. However, for phrases longer than twelve notes, a central dip tends to appear—what I like to call the "McDonald's effect." This dip might appear because two shorter phrases were inadvertently notated as a single long phrase, or because long phrases exhibit subphrase structures.

Incidentally, I found that if one represents each phrase by the average pitch-height of all the notes within the phrase, then whole melodies also tend to exhibit an arch

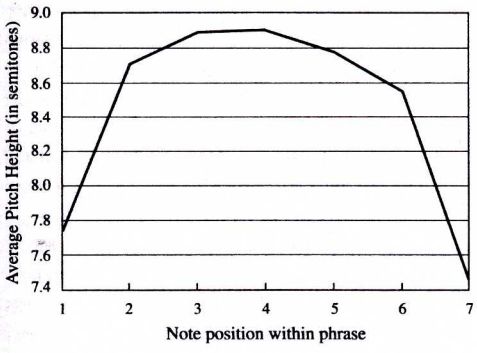

Figure 5.6
Average contour for 6,364 seven-note phrases taken from *The Essen Folksong Collection* (Schaffrath 1995). The graph shows the average pitch height (measured in semitones above middle C) according to serial position in the phrase. This arch shape contour is present for 5-note, 6-note, 7-note, 8-note, 9-note, 10-note, and 11-note phrases. From Huron 1996.

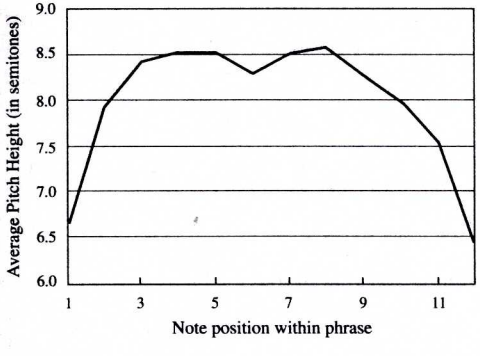

Figure 5.7
Average contour for 1,600 twelve-note phrases taken from *The Essen Folksong Collection*. The graph shows the average pitch height (measured in semitones above middle C) according to serial position in the phrase. The dip in the middle of the arch ("McDonald's effect") is evident in 12-note, 13-note, 14-note, and 15-note phrases and suggests possible subphrase structures. From Huron 1996.

contour. Over 40 percent of all melodies have successive phrases that tend to increase in overall pitch and then descend in overall pitch. However, no hierarchical relationship exists between convex phrases and convex melodies. That is, a melody that is convex overall is not more likely than other melodies to exhibit convex phrases.[17] In subsequent unpublished research I have found the same arch tendency in samples of other Western music. The effect is robust; however, it appears to be slightly less marked in instrumental music than in vocal music. To be sure, not all phrases in Western music exhibit an arch contour. Nevertheless, it is a clearly notable tendency.

Since Western phrases tend to exhibit arch-shaped contours, the next question is whether experienced listeners form appropriate "arch" expectations. Do listeners expect pitches to rise and then fall over the course of a phrase? Using music students at the Ohio State University, Bret Aarden was able to test this hypothesis. In a reaction-time study, listeners heard a number of folksong melodies. After each tone they indicated as quickly as possible whether the tone went up, down, or stayed the same. Aarden divided all of the reaction-time data into two sets: responses that occurred in the first half of phrases and responses that occurred in the second half of phrases. He then looked to see whether listeners are faster at recognizing an ascending interval in the first half of the phrase and whether they are faster at recognizing descending intervals in the second half of phrases. The results showed a clear pattern: musician listeners are no faster at identifying ascending versus descending intervals in the first half of phrases. However, in the second half of phrases, musicians are much faster in responding to descending intervals than to ascending intervals. In short, the results are consistent with the notion that listeners expect descending intervals in the latter half of phrases.

These results are difficult to explain. When a musical phrase begins, the listener is in no doubt that they are listening to the first half of the phrase. Surely, this should make it easy for listeners to expect that successive pitches will tend to rise upward in pitch. In contrast, it is much less clear how a listener determines when she is listening to the second half of a phrase. The number of notes in musical phrases varies considerably; surely, it is more problematic for a listener to infer that she is listening to the latter half of any given phrase. Nevertheless, it is the final descent of the phrase that listeners seem to expect, not the initial pitch ascent. The experimental results seem to contradict intuition.

Reprise

In this chapter I have highlighted five robust melodic tendencies that are evident in (notated) Western music: *pitch proximity, step declination, step inertia, melodic regression,* and *melodic arches*. Many of these organizational patterns can be found in non-Western musics as well. With the exception of yodeling and Scandinavian yoiks, *pitch proximity*

is pervasive throughout the world's music. With the possible exception of Hindustani music, *step declination* also appears to be a common musical pattern. *Melodic arches,* by contrast, are clearly limited to certain cultures. *Melodic regression* may well be universal, while the extent of *step inertia* awaits cross-cultural study.

In identifying musical patterns, we have only scratched the surface. In chapter 7 we will discuss some further pitch-related statistical patterns, and in chapter 10 we will look at some patterns related to rhythm and timing. In chapter 11 we will consider patterns that are style- or genre-specific.

Many of the patterns discussed in this chapter have long been known to music theorists. But the precise nature of these patterns has not always been recognized. As we will discover in the next chapter, it can be instructive to review how various music scholars have interpreted these musical patterns.

7 Mental Representation of Expectation (I)

When I answer the phone, I expect to hear a human voice reply to my "hello." When I turn on the faucet, I expect to hear the sound of flowing water. When I hear footsteps just outside my door, I expect to hear more footsteps, continuing on their way. If we ask "what do listeners expect?" the natural answer is "sounds of a certain sort"; we expect particular sound events with specific properties at given moments in time.

A problem with this answer is that brains do not store sounds per se. Auditory images are not organized in the brain like phonograph recordings. Instead, brains *interpret*, *distill*, and *represent* sounds. As noted earlier, expectations imply some sort of mental representation. The *what*, *when*, and *where* of expectation exist as mental codes.[1] These mental codes are not disembodied abstractions. They exist as real biological patterns that have taken up residence somewhere inside people's heads.

When a listener correctly anticipates a sound, that *real* sound will appear as two time-variant pressure fluctuations—one for each of the left and right eardrums. A listener ultimately anticipates a pair of signals like those shown in figure 7.1. The core question is: What is it about these signals that was correctly anticipated by the listener?

Suppose for the moment that you know nothing about sound or human audition. Suppose that you are a resourceful Mother Nature, attempting to build a biological organ that can predict two pressure signals such as the ones shown in figure 7.1. Our hypothetical sound-prediction organ (SPO) might anticipate that the fluctuations will begin at a particular moment. Or perhaps the SPO anticipates that the squiggles will exhibit upward spikes (rather than downward ones). Or perhaps it anticipates that the upper squiggle (the left ear) will begin slightly before the lower squiggle (right ear), or that the upper squiggle will be slightly taller. Perhaps our SPO correctly anticipates that the squiggles will grow large, then diminish, grow large again, and then diminish. Or perhaps it anticipates that there will be a repeated pattern of three little humps followed by one big spike.

First, let's dispense with the idea that there is a "right" way to describe such pressure signals. It is a common misconception to suppose that mathematics tells us how best

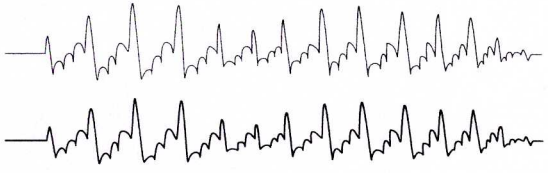

Figure 7.1
Any anticipated sound appears as two time-variant pressure signals, one for each of the left and right ears. What is it about these signals that is correctly anticipated by a listener?

to characterize such waveforms. The French mathematician Jean Baptiste Joseph Fourier famously devised a method for describing a periodic function as the sum of a series of harmonic sinusoids of various amplitudes and phases. The details here are not important. Suffice it to say that Fourier's theorem is a masterly bit of mathematics that is justly celebrated. But we should not be blinded by the formal beauty of Fourier analysis. One can similarly create a method that describes any function as the sum of harmonic *square waves* of various amplitudes and phases. Or one could use triangle waves. Or, for that matter, the facial profile of Napoleon Bonaparte (in whose army Fourier served with distinction). There are infinitely many such methods. Moreover, few sounds are precisely periodic, so any unadulterated Fourier analysis is strictly impossible for most real sounds. Finally, there are entire universes of nonlinear and nonharmonic ways of describing a function. With a moment's thought, it becomes plain that the number of ways of describing a sound function is unbounded. It is true that Fourier analysis holds some special mathematical properties, but that doesn't mean Fourier analysis provides the single "right" way for a biological system to represent a sound.

If the possibilities seem mind-boggling, the existence of functioning auditory systems indicates that real minds aren't completely boggled. The brain does somehow represent these squiggles (or, at least, properties of these squiggles) and is able to use these representations to form expectations about the properties of future squiggles.

Localization

Hearing scientists, psychoacousticians, and auditory neurologists have been able to identify many aspects of sound functions that brains represent. A useful place to begin is with the seemingly lowly phenomenon of sound localization. Location is biologically important. A priority for any auditory system is to establish the location of possible prey and predators. In seeking food and avoiding danger, biology follows the real-estate agent's maxim: *location, location, location*. The biological importance of

localization is evident in comparative auditory physiology. The neural structures that decipher sound location are among the oldest in the brain. Localization represents an evolutionarily ancient neurophysiological adaptation. Not all animals experience pitch, but all animals that have a hearing organ can localize sounds.

Several factors are known to influence the perception of auditory location. Two factors figure prominently. One is the relative time difference between pressure fluctuations in the left and right ears. A second is the relative difference in the magnitude or amplitude of these signals. Both of these factors rely on differences between the ears—so-called *interaural* differences. These interaural squiggle-properties provide useful correlates with the physical *location* of a sound source.

When localizing sounds, interaural time and amplitude differences reflect genuinely stable physical relationships between sound sources and pairs of listening devices (like ears). In a homogeneous medium the speed of sound is constant, so even in the most reverberant of environments the onset of a sound will always reach the nearer ear prior to the more distant ear. Similarly, the inverse-square law describes a simple geometric fact whose consequence is that sounds lose energy with increasing distance from the source. Given the width of the head, the magnitude of these interaural differences will be systematically related to the horizontal direction or *azimuth* of a sound. If a sound originates from your left, the pressure disturbance will arrive at your left ear prior to your right ear; it will also be more energetic in your left ear. As noted in chapter 4, when an environment remains stable, then evolution is likely to favor the creation of an innate mechanism.[2] In the case of azimuth, there are some stable physical principles that make it possible for Mother Nature to build such innate mechanisms.

The perception of the horizontal azimuth of sounds contrasts notably with the perception of their vertical position or *elevation*. Whereas two-eared ("binaural") organisms can infer a sound's azimuth, they cannot infer the elevation of a sound from universal geometric facts such as interaural time or amplitude differences. Instead, the perceived elevation of a sound is known to be facilitated by the distinctive shape of the external visible part of the ear known as the *pinna*. The outer ear acts as a complex resonant filter, changing the tone color of sounds depending on their vertical projection onto the surface of the pinna. A sound coming from above will have a slightly different timbre than the same sound coming from straight ahead, or from below.

Notice that the ability to infer the elevation of a sound presupposes that an organism "knows" (in some sense) what the unfiltered timbre should sound like, and is then able to use timbral modifications as cues for elevation. Since the world is full of lots of different sounds of various elevations and timbres, this implies that elevation is probably a *learned* aspect of localization—unlike horizontal azimuth. Current auditory research provides good support for this view. Paul Hofman and John Van Opstal at

the University of Nijmegen fitted volunteers with binaural plastic molds that modified the shape of their pinnas. These molds immediately abolished the ability of listeners to judge the elevation of sounds but did not disrupt their azimuth (left–right) judgments. However, after wearing the molds for periods of up to five weeks, all listeners had adapted to the plastic molds and had regained their ability to judge elevation. That is, listeners were able to learn how the novel spectral transformations relate to elevation.[3]

In a contrasting experiment, Hofman and his colleagues outfitted participants with full-spectrum "in-the-ear-canal" hearing aids that were contralaterally wired: the sounds arriving at the left ear were shipped over to the right ear and vice versa. The immediate effect of this manipulation was to swap the horizontal azimuth: listeners judged sounds from the right as originating on the left, and vice versa. With the passage of many weeks, listeners failed to learn to adapt to this new arrangement. Listeners never stopped hearing the sounds as laterally displaced, even though the auditory cues were constantly contradicted by what they were seeing.[4]

In short, *azimuth* perception appears to be innate, whereas *elevation* perception appears to be learned. Notice that these differences are in accord with the Baldwin effect discussed in chapter 4. Recall that the Baldwin effect says that innate mechanisms are more effective when the environment is relatively unchanging, while learned mechanisms are more effective when the environment is more variable. Those aspects of localization that can be linked to stable physical properties in the world (like principles of geometry) appear to lead to innate coding mechanisms. By contrast, those aspects of localization that rely on variable factors (such as the familiarity of different timbres) lead to mental codings that appear to be learned.

Research suggests that there are several *levels* of representation involved in what we call "localization." Neurophysiologists have identified specific regions in the brain where neurons "code" interaural time and amplitude differences. We might say that location is represented using the codes of interaural time and amplitude differences. But auditory processing does not stop with these representations. Other parts of the auditory system translate these (and other) sensory codes into the more useful representations of *azimuth*, *elevation*, and *distance*. These perceptual codes are much more useful to an organism. When we experience a sound, we have no conscious access to the neural representations of interaural time difference or interaural amplitude difference. Instead, we experience a mental representation that says "over to the left," "up high," or "pretty far away."

Of course, the auditory system doesn't just attempt to resolve the current location of a sound. It also endeavors to predict the location of future sounds. In the realm of expectation, *where next* is probably the first predictive mechanism to have evolved in the primordial auditory system.

The *where-next* function requires yet another set of mental representations. Since the perceiving animal may be in motion, it is useful to translate the relative relationships of azimuth, elevation, and distance to an absolute location—namely, the representation of *place* rather than *bearing*. By creating a mental code for place, we can better predict a sound location that compensates and eliminates the possible effects of our own movements. In addition, a place-based mental representation might integrate information from the visual system to provide a more robust coding. Finally, a mental representation of *speed* and *trajectory* can help us better predict location when the sound source itself is in motion.

By way of summary, there exist several levels of mental representation for localization. The lowest (unconscious) representational level includes interaural time, amplitude differences, and spectral shape. A higher (subconscious) representational level includes azimuth, elevation, and distance. With further processing, conscious representations emerge, such as place, speed, and trajectory.

Having described some functionally useful representations, let's return for a moment to consider the problem of predicting sound functions such as those shown in figure 7.1. In the case of localization, only some of the features evident in these sound functions are pertinent. The most important features are the relative time relationship and the difference in amplitudes.[5] From the perspective of predicting location, it is important to predict that the upper sound function begins slightly before the lower one, and that the upper function also exhibits a slightly higher amplitude. Note that most of the features one could describe about these two sound functions are simply not germane to the task of localizing a sound—or predicting its future location. The mental representation is strongly shaped by the physiological function.

The case of sound localization illustrates an important fact about mental representation. One might think that an ideal mental representation would be one that accurately reflects the organization of the real world. However, the business of biology is survival and procreation, not truth or accuracy. In order to understand the nature of mental representations, we must consider what biological functions they serve and what types of pertinent information are available in the world.

Notice that we have not resolved the question of which *level* of localization-related mental codes might form the basis for expectation. Suppose, for example, that a listener anticipates that the next sound will come from a more central location. How is this expectation represented? Does the listener's brain anticipate the shortening of the *interaural time difference* between the ears? Or does the listener's expectation make use of a higher level representation, such as anticipating that the *azimuth* will shift toward the body's midline? Which mental code or codes are used in anticipating the future? In what "language" are predictions expressed?

The existing auditory research doesn't yet have an answer to this question. It is a difficult question because there are lots of possibilities. For example, a higher-level

representation (such as *trajectory*) might be used to "calculate" an expectation, but that expectation might be concretely expressed as, say, an anticipated interaural amplitude difference. Alternatively, a low-level representation such as interaural amplitude difference may play no role in the forming of an expectation, except that interaural amplitude difference is used as an intermediate stage in calculating the higher-level codes that *are* expected. Yet another possibility is that all of the mental representations related to localization are concurrently involved in independent predictions about what will occur next.[6] These are empirical questions that can be answered only through future experimental research.

The problem of predicting auditory location might seem to have little pertinence to music.[7] But our discussion highlights some useful lessons about mental representation in general. First, just because we can identify (mathematically) certain features in a sound does not necessarily mean that these features are germane to brains. Mental representation is shaped by physiological function. Second, there may be several representations pertinent to a particular function—as in the case of interaural time differences and interaural amplitude differences. Third, representations can build on one another. For example, auditory "bearing" probably depends on underlying representations such as interaural time and amplitude codes. Fourth, mental codes rely on features of stimuli that maintain a stable relationship to the environment—either the long-term environment of evolutionary history, or the short-term environment of individual learning. A change in the attended feature should reliably reflect a pertinent change in the real world.

Representing an Expected Pitch

Consider now the more musically pertinent problem of pitch representation. How is the pitch of an expected tone mentally represented? We might begin by pointing to the usual suspects. Theoretically, a listener might expect a particular absolute pitch, such as the pitch B3. Or the listener might expect a particular pitch-class—for example, the set of all Bs including B2, B3, B4, B5, and so on. Or the listener might expect a particular contour; for example, the listener might simply expect the next pitch to be higher than the current pitch. Alternatively, the listener might expect a particular interval, such as the interval of a rising major second. Or the listener might expect a particular scale degree, such as the seventh scale step or leading tone. The listener might expect an ensuing pitch to be a member of some specific chord; for example, any of the pitches of the G major chord G3, B3, or D4. More subtly, rather than expecting particular notes of a particular chord, the listener might expect the pitch to be a member of "the dominant chord." Or the listener might expect the pitch to be a member of any dominant function, such as the tones of a dominant minor-ninth chord or an augmented dominant triad.

There are also lower-level representations that must be considered. For the psychoacoustician and hearing scientist, *pitch* is already a remarkably high-level representation. At the level of the sensory organ, the cochlea encodes nothing that resembles pitch. At the cochlear level, all sounds appear to be represented in two forms. In the first instance, different frequencies cause different places along the basilar membrane to be excited. Each place is associated with the firing of different sensory neurons. This sound–place mapping is referred to as a *tonotopic* representation, and this tonotopic coding of sound can be observed at many places throughout the auditory system—including on the auditory cortex.

A second representation relates to the rate and pattern of firing of various auditory neurons. Throughout the auditory system, neurons exhibit a tendency to fire in synchrony with the frequencies driving the ear drum. Neurons cannot fire faster than about 1,000 times per second. Since humans are able to hear frequencies well above 10,000 Hz, there are limits to this synchronous firing. Nevertheless, the rate of firing is known to play a role in the representation of pitch. The idea that frequency is mentally represented using a combination of tonotopy and neural firing was first proposed in the 1950s by Joseph Licklider. This two-component theory is referred to as the "duplex theory" of pitch.

As in the case of localization, it is possible that an expected sound is represented by a combination of several underlying representations. For example, a listener might expect a tone to be from a particular pitch-class, *and* to activate particular tonotopic points. Or a listener might expect that the pitch will be very close to the current pitch, but will also be a member of the dominant seventh chord. Or a listener might expect that the contour will ascend to either the fourth or sixth scale degrees. In addition, it is possible for combinations to involve non-pitch-related representations. For example, a listener might expect a combination of long duration and particular pitch-class.

So how precisely are musical pitches represented in the brain? What precisely do listeners expect? First, let's consider why a profusion of different representations might be useful for a listening brain.

Neural Darwinism

As I have already emphasized, when a phenomenon is stable over a long period of time it becomes possible to evolve specialized innate representations such as interaural time and amplitude differences. However, when an environment is highly variable, then it is better to evolve the capacity for learning. When learning proves to be the most adaptive strategy, it is appropriate not only for the specific expectations to be learned, but also for the underlying representation itself to be learned rather than innate. A good mental representation would be one that captures or approximates some useful organizational property of an animal's actual environment.

In the case of audition, we know that learning plays a dominant role. If the real world of sound is organized according to scale degrees, for example, then scale degree might be an appropriate mental representation for expressing pitch-related expectations. If the real world is organized according to a combination of (say) pitch contour, metric position, and diatonic interval, then an appropriate mental representation might echo this mixed organization.

But how does a brain know which representation is the best? How can an auditory system learn to discard one representation in favor of another? Here expectation may play a defining and perhaps essential role. Expectation is an omnipresent mental process; brains are constantly anticipating the future. Moreover, we have seen that there is good evidence for a system of rewards and punishments that evaluates the accuracy of our unconscious predictions about the world. A defective mental representation will necessarily lead to failures of prediction. Conversely, a mental representation that facilitates useful predictions is likely to be retained. In effect, our mental representations are being perpetually tested by their ability to usefully predict ensuing events.

This claim carries an important implication. It suggests that the auditory system is spontaneously capable of generating several representations from which the less successful can be eliminated. This in turn suggests that *competing concurrent representations* may be the norm in mental functioning.

Neurophysiologists have posited precisely such competitive concurrent neural structures. The foremost advocate of this view has been the Nobel laureate Gerald Edelman, who has dubbed this theory *neural Darwinism*. Other advocates include William Calvin.[8] Edelman and Calvin have assembled strong neurophysiological evidence supporting this view. According to the neural Darwinism theory, "cortical real estate" is the resource over which different functions and representations compete. "Unsuccessful" functions atrophy and their cortical resources are taken over by the more "successful" functions.[9] (This competition is restricted primarily to the cortex and is less evident in subcortical brain areas.) It has been suggested that this competitive process accounts for much of the flexibility or plasticity that is evident in the brain.

According to Edelman and Calvin, representations compete with each other according to Darwinian principles applied to patterns of neural organization. Those representations that prove most useful in predicting future events are preserved and reinforced, while less useful representations atrophy. Such neural competition is possible only if more than one representation exists in the brain. That is, in forming expectations, the normal brain would maintain multiple concurrent representations. Relying on a single representation would mean either that the brain had achieved near perfection in forming predictions about the world, or that the representation is genetically ordained, or that the brain has become pathologically structured.

Viewed from a functional perspective, something like the prediction response would be a logical necessity in order to provide feedback concerning the relative success of different representations. The whole process involves a sort of feedback loop: representations are used to form expectations, and the accuracy of these expectations is used to select among the various alternative representations.

Notice that this feedback loop would be highly sensitive to the type of auditory environment in which a listener resides. Given different environments, we would expect people to differ in their mental representations for sound. For example, if the prevailing music in some culture were dominated by contour-related regularities, then we would expect contour-related representations to be foremost in the mental functioning of members from that culture. In a different culture, the prevailing musical organization might favor some other pitch-related mental representation among experienced listeners.

As we will see later in this chapter, recent research on absolute pitch will support this interpretation. That is, we will see evidence of multiple concurrent music-related representations, of competition and atrophy, and evidence that the preeminence of one or another representation depends on listener-specific circumstances and environments.

Expectation serves at least three functions: *motivation*, *preparation*, and *representation*. First, by anticipating future events, we may be able to take steps now to avoid negative outcomes or increase the likelihood of positive outcomes. That is, expectations have the capacity to motivate an organism. Second, even if we are unable to influence the course of future events, expectations allow us to prepare in appropriate ways. For example, accurate expectations allow us to adopt a state of arousal that is better suited to what is likely to happen next. Accurate expectations also help us orient in the direction of an anticipated stimulus, and so increase the speed and accuracy of future perceptions. That is, expectation allows us to prepare suitable motor responses and craft suitable perceptual strategies. Finally, expectation provides the test-bed for evaluating various mental representations. That is, expectational success and failure provides the "natural selection" mechanism for the neural competition that may underlie mental representation.

If it is true that mental representations are in competition according to neural Darwinism, then two consequences follow. First, there must be a mechanism for generating new representations. It is possible that this mechanism is only or at least predominantly active early in life and becomes inactive in adulthood. But at some point in a listener's life, there must be a process that generates alternative auditory representations. Second, selection can take place only if all of the competing representations are simultaneously engaged in generating expectations. Selection can take place only if there are competing (and diverging) predictions. This implies that normally more than one mental representation is involved in activating expectations.

These points bear on our earlier question: Do listeners expect a particular pitch, or a particular scale degree, or a particular interval, or a particular contour? If I am correct, for any given auditory stimulus, the listener's brain is generating predictions using several representations. However, the predictions arising from the different representations are by no means treated equally. Each listener will have a distinctive listening history in which some representations have proved more successful than others. A typical listener will hold a combination of differently favored expectations. Informally, we might regard a listener's expectation as some sort of weighted sum, say, 70 percent scale degree, 15 percent harmonic function, and 10 percent pitch contour.

Acquiring a Representation—The Case of Absolute Pitch

In chapter 4 we briefly discussed the phenomenon of absolute pitch (AP)—the ability to identify the pitch of tones without the use of any external reference. Absolute pitch is one of the most studied phenomena in music perception. Experimental research concerning AP has been carried out for more than a century. The intense curiosity about absolute pitch arises from its relative rarity. If everyone had AP we wouldn't give it a moment's thought. What makes it a compelling topic is the fact that only a few people develop this skill. Why?[10]

Some correlational evidence suggests that the capacity to develop AP may involve a genetic predisposition.[11] But even if genetics plays a role, the existing research suggests that a critical learning period is involved. One of the best generalizations one can make about "perfect pitch" is that its possessors typically begin musical instruction or involvement at a comparatively early age—often before the age of six or seven years.[12]

Many musicians would like to acquire perfect pitch and so there is a ready market for "methods" that purport to help adults develop it. Occasionally, an adult does succeed in developing AP, but the methods produce inconsistent results. One of my former students, Peter Sellmer, spent a year attempting to develop AP. He worked at it daily and regularly tested his progress. At one point he was even convinced that he was improving. But systematic pre- and post-tests established that his year-long effort had been fruitless. It turns out that those individuals who succeed in acquiring absolute pitch as adults typically began their music involvement at an early age. Adults who began musical involvement as teenagers almost never acquire perfect pitch.[13]

An often overlooked precondition for the development of absolute pitch is that the pitch environment remain stable. A person cannot learn to name pitches if the pitch of nominally identical tones keeps changing. Some musical instruments, such as the guitar and the violin, must be frequently retuned. If there is no stable external pitch

reference for tuning such instruments, then it will be impossible for a person to develop absolute pitch. In many cultures, there is no fixed tuning system. For example, the instruments of a traditional gamelan are typically tuned to one another, but the overall tuning will change from ensemble to ensemble. Since most gamelan instruments are metallophones, they retain a fairly constant pitch. This means that a musician may develop absolute pitch related to an individual instrument, or to a specific group of instruments. But this skill may not transfer to the gamelan in a neighboring village.

The human voice is perhaps the most obvious example of an "instrument" without fixed tuning. Unless a culture maintains instruments with stable tuning, a predominantly vocal culture is anathema to the development of absolute pitch.

The phenomenon of pitch instability or variable tuning might explain why not all people acquire absolute pitch. In 1901, Otto Abraham proposed a theory that has come to be known as the "pitch unlearning" theory. A century ago this theory was largely regarded as an implausible curiosity, but today the theory is regarded much more favorably.[14] Abraham pointed out that in many music-making environments, there is no fixed tuning system. Consider, for example, the sort of musical activity commonly found in a nursery school or kindergarten. Most such environments don't have a piano; the music-making is often voice-only. Today, the children might sing "Happy Birthday" for Jill in the key of E major; tomorrow, the children might sing "Happy Birthday" for Ken in the key of G major. As long as individual musical works are sung in different keys, there is no possibility of "coding" the tune as a sequence of absolute pitches. In such an environment, there is simply no advantage for a brain to represent events in terms of absolute pitch. Said another way, in a pitch-variable environment, an absolute pitch encoding provides no predictive advantage whatsoever.

Abraham suggested that people might have a natural tendency to acquire absolute pitch. However, when exposed to an auditory environment characterized by high pitch variability, any latent perfect pitch capacity is likely to atrophy. It is not quite correct to say that people "unlearn" perfect pitch. But it is an appropriately colorful name for Abraham's theory.

Yet there is an even more compelling reason why brains should be indifferent to perfect pitch: there are better possible mental representations. If one of the purposes of a mental representation is to facilitate prediction, then a *relative pitch* representation will prove much more useful for music-related pitch. In the singing of "Happy Birthday," a relative pitch representation will prove consistently more accurate in predicting ensuing tones.

Given the greater value of relative pitch representations, one might ask why *anyone* would retain an absolute pitch ability. What possible advantage does absolute pitch

confer? Musicians will note that absolute pitch is useful for tuning without a reference pitch, that it helps performers maintain accurate intonation, that it facilitates recall of musical works from memory, and that it aids in transcribing music by ear. But however much these skills might be useful for professional musicians, they hold only negligible value from the perspective of mental functioning. Indeed, researchers have shown that there are significant *disabilities* that attend perfect pitch (see below). Later we will see why a brain might develop this skill, despite its limited utility.

Recall that in chapter 4 we reviewed some of the evidence in support of absolute pitch as a learned phenomenon. One of the strongest pieces of evidence is to be found in reaction-time studies for pitch identification. Recall that people with absolute pitch are faster when identifying the most commonly occurring pitches, and that the reaction-time measures are consistent with the Hick-Hyman law.[15] As we noted in chapter 4, this finding implies that AP is learned through simple exposure.

In the past decade, researchers have become more aware of how absolute pitch interferes with certain musical tasks. A seminal study was carried out by Ken'ichi Miyazaki at Niigata University.[16] Miyazaki demonstrated that in certain circumstances, absolute pitch possessors perform worse than non-AP possessors. Consider, for example, the rising major sixth interval from C4 to A4. Both AP and non-AP possessors could readily identify this interval—although AP possessors responded slightly faster than non-AP musicians. Now consider the rising major sixth interval from C#4 to A#4. Here non-AP possessors identified this interval just as quickly as the interval C4 to A4. However, the AP possessors responded significantly slower. Miyazaki's study demonstrates that many AP possessors "calculate" the interval from the pitch names. That is, many AP possessors hear the pitches C4 and A4, and then determine that the interval must be a major sixth. However, when the pitches are C#4 and A#4, the interval calculation becomes more complicated—especially since many AP possessors tend to identify the A sharp as a B flat instead. Because of the enharmonic change, musicians are slower to identify the interval between C sharp and B flat.

In effect, Miyazaki showed that for many AP-possessors, the ability to identify pitches impedes their ability to learn intervals by ear. Rather than identifying intervals from the relative distance separating the tones, AP-possessors often rely on deriving intervals from pitch names. Miyazaki's work suggests that many possessors of AP have no native mental representation for intervals.

For music teachers this situation would seem highly problematic. Identifying intervals by ear is considered a basic musicianship skill. But we shouldn't be too quick to judge. First, why do musicians attempt to learn to identify pitch intervals? The answer is to be able to reproduce heard or imagined sounds. That is, we learn to recognize intervals so that we can notate pitch sequences, or reproduce them when playing on an instrument. Notice, however, that both of these tasks could be carried out perfectly

well if we could always recognize the pitch of a sound. What incentive is ther person with absolute pitch to ever use interval-recognition when notating mus the relative pitch user, once the key is established, intervals are translated bac pitches anyway. Why not just represent the pitches directly?

Music teachers have a ready answer for this. It is often useful to be able to tra music into a different key. For example, singers have different vocal ranges ar important to be able to perform the music at a different absolute pitch height singer has absolute pitch, this is hard to do. If the notation reads "F#," it is h the AP-possessor not to imagine and sing F#.

It may well be that developing brains begin by assuming a simple represer (such as absolute pitch). If the world is not organized in a manner consister absolute pitch (as in the persistent singing of "Happy Birthday" in different then some other representation (such as interval or scale degree) will becom appropriate. However, any latent absolute pitch representation will be prese the extent that it retains some value in predicting the future.

Let's review, then, the basic facts about absolute pitch as they are cu understood.

1. Not everyone develops absolute pitch.
2. If absolute pitch emerges, the basis for it tends to be laid in early life.
3. Reaction time data shows that AP is acquired by exposure to the environ faster reaction times happen for those pitches that are encountered most ofte
4. Possession of AP doesn't mean that the person can *only* code pitches this v
5. Nevertheless, possession of absolute pitch can retard the development of or intervalic pitch coding.
6. Absolute pitch proves useless in situations where there is no standard tuni
7. Absolute pitch never develops in sound environments where it is not usef

Notice that all of these facts are consistent with the principles of neural Dar all of these facts are consistent with the notion that representations that pr most useful in predicting future events are preserved and reinforced. Represer that fail to provide accurate predictions about the world are less likely to dev short, the case of absolute pitch is consistent with three theoretical claims:

1. There are competing mental representations for sound.
2. Representations are shaped by exposure to the environment.
3. Representations are differentially favored depending on their predictive su

The evidence in support of this view is not extensive; further research is Nevertheless, the phenomenon of absolute pitch is consistent with my earlier tion that one of the functions of the prediction response is as an engine of s for mental representation.

Correlated Representations

Over the years, perceptual research in music has shown that listeners are sensitive to many different experimental manipulations. Reading this research literature, one might conclude that there exists a multitude of concurrent representations that are involved in music listening. The research implies that at least some listeners are able to code sounds as absolute pitches, pitch chromas (or pitch classes), scale degrees, intervals, scale-degree dyads (successive scale degrees), contours, durations, relative durations, metric positions, harmonic functions, chord qualities, and spectral centroids—to name a few. Some mental representations are apparent only in certain tasks. Other representations (like absolute pitch) are available only to a minority of listeners. Some research explicitly suggests that musical coding involves more than one concurrent representation.[17] There is, however, a problem that should make us wary of accepting the research at face value. The problem becomes evident only when we directly compare several different possible representations.

Figure 7.2 plots the flow of information for the tune "Pop Goes the Weasel." Information is plotted (in bits) for five different representations. The uppermost plot shows information according to the probabilities of different scale degrees. Less common scale degrees (such as $\hat{6}$—pitch E) convey more information compared with frequently occurring scale degrees (such as $\hat{1}$ and $\hat{5}$—pitches G and D). The second plot shows information according to scale-degree successions or scale-degree dyad. Here high probabilities (low information) events include the succession $\hat{1}$ followed by $\hat{1}$ and $\hat{4}$ followed by $\hat{3}$. Low probability (high information) events include $\hat{1}$ followed by $\hat{6}$ and $\hat{6}$ followed by $\hat{2}$. The third graph plots metric position for 6/8 meter. High probability events include notes whose onsets coincide with the downbeat of each measure. For 6/8, a low probability (high information) event occurs for notes whose onsets coincide with the second eighth of the measure. The fourth graph plots interval. Common intervals include the unison repetition and the rising major second. Uncommon intervals include the rising major sixth and the descending perfect fifth. Finally, the fifth graph plots interval dyad (pairs of successive intervals). High probability events include a unison followed by an ascending major second and an ascending major second followed by a unison. Low probability events include the rising major sixth followed by the descending perfect fifth, and the falling perfect fifth followed by the ascending minor third. The probabilities used in figure 7.2 were derived from an analysis of some six thousand European folk songs.

Notice that the information for both scale-degree and melodic interval representations peak at the word "pop." For scale degree dyad and interval dyad the word "pop" coincides with the second highest information value—with the maximum value following immediately after the word "pop." There appears to be an element of musical "surprise" at this point that is echoed in the lyrics. As a children's action

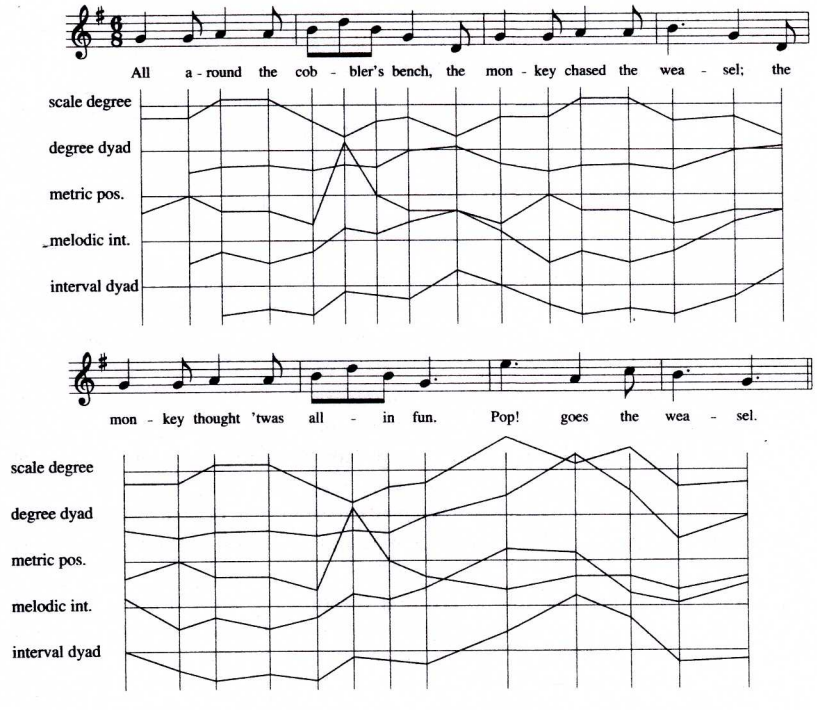

Figure 7.2
Information theoretic analysis of "Pop Goes the Weasel" showing changes of information (in bits) as the piece unfolds. Plotted information includes scale degree, successive pairs of scale degrees (degree dyad), metric position, melodic interval, and successive pairs of melodic intervals (interval dyad).

Table 7.1
Correlation matrix showing similarities in information fluctuation for five representations used to analyze "Pop Goes the Weasel." Pitch-related representations are highly correlated.

	degree	degree dyad	metric position	interval	interval dyad
degree	+1.00				
degree dyad	+0.45	+1.00			
metric position	–0.31	–0.05	+1.00		
interval	+0.17	+0.74	–0.00	+1.00	
interval dyad	+0.30	+0.90	+0.02	+0.77	+1.00

song, this point is usually accompanied by some abrupt action, also suggestive of surprise.

Note, however, that there is no comparable peak in information for metric position. That is, the interval/pitch/scale-degree may be relatively surprising, but the *moment* of its occurrence is not surprising. This highlights the independence of the *what* and *when* of surprise. Here, the *what* is relatively surprising but not the *when*. In other musical situations, the *what* is expected, whereas the *when* may be relatively unexpected. (More musical examples will be considered later in chapter 14.)

With the exception of the metric position information, all of the pitch-related information values are broadly similar. That is, they are positively correlated. Table 7.1 shows a correlation matrix for the information content (measured in bits) for the various representations used in the above analysis of "Pop Goes the Weasel." Three representations are closely intertwined: interval, interval dyad, and scale-degree dyad. The average correlation between these three representations is a remarkable +0.80 (of a theoretical maximum of +1.0). These positive correlations are not limited to Western music. I have carried out analyses of some 200 melodies from American, Chinese, Dutch, Pawnee, and Xhosa sources and found that these positive correlations are typical.

The fact that different musical representations are positively correlated is both good and bad news. An advantage of this correlation is that it suggests that probabilistic analyses of music are fairly insensitive to the choice of pitch-related representation. For some analytic tasks, knowledge of the precise nature of the mental representation may not be important. The disadvantage of this high correlation is that it invites onerous mistakes of interpretation. When an experimenter manipulates some pitch in a stimulus passage, it is not just pitch that changes. Typically, a change of pitch will also cause a change of scale degree, a change of chroma, a change of interval, and a change of scale-degree succession, among other changes. A change of pitch might also lead to a change of melodic contour. Suppose that the experimenter finds that this change is

noticed by a listener. Can we conclude that the listener has heard a change of pitch? No. Unless the experimenter takes care to ensure that no change in contour, interval, scale degree, interval dyad, and so on took place, it is impossible to interpret the listener's discrimination. In research methodology, this problem is referred to as "the third variable problem."[18] If a listener distinguishes stimuli that differ according to property *A* it does not follow that the listener is "sensitive" to or "encoding" property *A*. By changing *A*, an experimenter might inadvertently be changing property *B*—and it may be property *B* that listeners are able to code or distinguish.

Nearly all of the experiments that purport to manipulate some pitch-related property (like interval) also end up transforming other properties as well. This means that taking an experiment to suggest that people are sensitive to, say, pitch contour may be entirely wrong. The high correlation between the various pitch-related representations means that we must be very careful in interpreting experimental results. Music psychologists have largely failed to be sufficiently discerning when manipulating different pitch-related parameters. As a result, much of the existing research related to the mental representation of music needs to be redone. Incidentally, the high correlation between various pitch-related representations does not mean that it is impossible to ever resolve which representations are coded mentally. It is possible to organize experiments so that these confounds are controlled, minimized, or circumvented. However, implementing such experiments requires careful planning.

The Enigma of Melodic Interval

A few years ago I began to doubt whether I hear melodic intervals. Don't get me wrong: I'm as capable of generating and recognizing intervals as the next music teacher. If you ask me to sing an ascending major sixth, I'll readily oblige—but I know that I'm cheating. I'll really be singing the first two pitches of the song "My Bonnie Lies Over the Ocean." I might sing the syllables "la la"—but really, my mind is thinking "so–mi."

I'm also aware that when I hear a major sixth interval, I find it much easier to recognize if the tonal context for the interval is "so–mi" rather than, say, "re–ti" or "fa–re." If the interval occurs in an uncommon context, I sometimes check my intuition by mentally transposing the key so that I can hear the interval as "so–mi." Then I really *know* it's a major sixth.

All of this raises the interesting hypothesis that I have no native mental code for melodic interval. Instead, I may have a scale-degree code, and perhaps a scale-degree dyad code. I tend to hear the rising perfect fourth as "so–do" or sometimes "mi–la"—but almost never as "re–so" or "ti–mi." Rising minor thirds I hear as either "mi–so" or "do–may." Major and minor seconds are an exception: I seem to hear these intervals in a more automatic fashion without being tied to scale relationships.

In many ways, my identification of melodic intervals resembles some of the AP musicians observed by Ken'ichi Miyazaki. I seem to identify the notes and then mentally infer the interval. Since I don't have absolute pitch, I'm not able to identify the pitches per se. Instead, I'm experiencing the tones as scale degrees and then "calculating" the interval. While I can consciously recognize that "so–do" and "la–re" are the same interval, they don't sound at all the same to me. I seem to have no native representation for the concept "perfect fourth" apart from a willfully imposed classification that groups together a bunch of obviously different things.

When I've spoken with other musicians about this, about a third report having similar experiences. If nothing else, it is at least reassuring to learn that I'm not alone. But do some listeners really experience intervals this way or are my introspective descriptions flawed?

With my collaborator, Bret Aarden, I carried out an exploratory experiment to try to determine whether other musicians fail to "hear" pitch intervals. In our experiment we had musicians identify various melodic intervals while we measured their reaction times. Rather than simply playing the intervals to our participants, we preceded each interval by a key-defining context. Specifically, we played a I–IV–V–I harmonic cadence before each trial. Between trials we played a random atonal pitch sequence to try to erase any lingering key that might interfere with the next trial.[19] We predicted that if listeners encode intervals as scale-degree dyads, then certain key-congruent presentations should be more quickly identified than other presentations. Figure 7.3 illustrates two sample trials from our experiment. If a listener tends to experience a major sixth interval as "so–mi" then trial (a) should result in faster interval identification than the "fa–re" presentation of trial (b).

Of course, even with the key-defining context, it is possible that some of our listeners would rapidly change the mental key in order to more quickly identify the presented interval. To control for this possibility, we followed the interval-identification task by a second task. After the participant had identified the interval, we immediately played a ii^6–V–I harmonic cadence and asked our listeners if this cadence was in the same key as the original key-defining cadence that began the trial. This second cadence had a 50 percent chance of being in the same key. We used this latter judgment to gauge whether the listener had maintained the original key or had inadvertently drifted to a different key while making the interval judgment.

Trial (b) in figure 7.3, for example, shows a key-incongruent test cadence. The trial begins in the key of C major and presents a major sixth interval as "fa–re." If a listener tended to hear the interval as "so–mi" then they would have had to reconceive of the pitches in the key of B-flat major. If they did this, it would increase the probability that they would falsely claim that the ensuing test cadence was the same as the initial key-defining cadence.

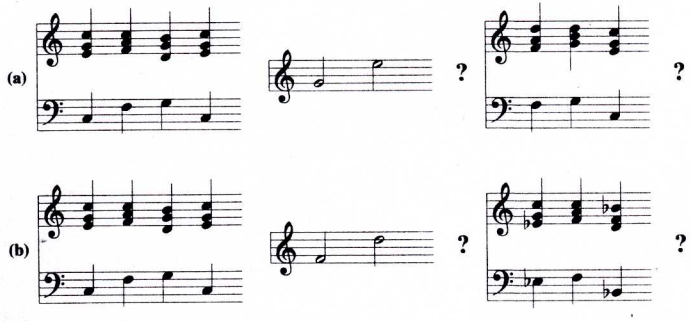

Figure 7.3
Two sample trials used in Huron and Aarden's (unpublished) study of the effect of key context on interval identification. Listeners first heard a key-defining context (I–IV–V–I cadence) followed by an ascending melodic interval. In (a) the interval is an ascending major sixth between scale degrees $\hat{5}$ and $\hat{3}$. In (b) the same interval is played between scale degrees $\hat{4}$ and $\hat{2}$. Musician participants then identified the interval. We predicted that responses to (a) would be faster than to (b). After the participant identified the interval a test cadence is heard (a ii^6–V–I cadence). In (a) this cadence is consistent with the original key-defining context. In (b) the test cadence is consistent with the different key of B-flat major. Participants had to identify whether the key of the test cadence corresponded with the initial cadence. If participants tended to code ascending major sixths as $\hat{5}$ to $\hat{3}$ then we would predict that in stimulus (b), participants would be more likely to incorrectly claim that the test cadence was consistent with the original key.

Our results were somewhat mixed. Some intervals for some musicians appear to be facilitated for certain key contexts. But apparently, not all musicians experience melodic intervals the way I do. With further research, we may be better able to resolve the issue of pitch-interval representation. But there is more work to be done.

Even if we assume that all listeners develop no native code for successive pitch intervals, it does not follow that it is useless to train musicians to consciously identify intervals. Intervals are useful objects in music, and there are lots of circumstances where it is helpful to identify intervals—and to identify them without regard to key context.[20] But we should be careful to distinguish native mental representations from consciously derived labels. At least in my case, it is doubtful that I hear melodies as successions of intervals. My subjective intuition is that I experience melodies primarily as successions of scale degrees. However, this type of introspection is fraught with dangers. Consider, by way of example, the hypothetical mental organization shown in figure 7.4.

Here we see an auditory system that begins with a pitch representation from which an interval representation is generated. In this hypothetical organization, both of

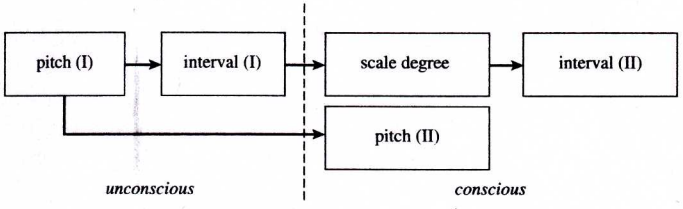

Figure 7.4
A hypothetical mental network for pitch-related representation. In this hypothetical scheme, pitch is a low-level sensory representation from which interval is derived. This interval representation is not experienced consciously, but it is used to derive a mental representation for scale degree. Both pitch and scale degree are then experienced "natively" as a *conscious* perceptual experience. Musical training allows this individual to mentally determine or "calculate" interval based on scale degree.

these mental representations are *unconscious*—codes that are not accessible to consciousness. From the interval representation, a scale-degree representation is created. Also, the pitch representation is duplicated and refined (Pitch II). Both of these latter representations are available to consciousness. Through musical training, this individual has learned to identify intervals (Interval II). This is a conscious task that makes use of the scale-degree code that is accessible to consciousness. Ironically, in this hypothetical scheme, conscious interval recognition would be a reconstruction from a scale-degree representation that itself is constructed from an earlier interval representation that is not accessible to consciousness.

My point is that although my introspective intuition suggests that my preeminent pitch-related representation is scale degree, this does not necessarily mean that interval is perceptually unimportant to my mental representation. In some ways it is difficult to imagine how a scale-degree representation can arise without some sort of intervalic information. Of course, if there is some unconscious interval coding, it will not employ the verbal labels musicians use (major second, minor sixth, etc.).

If such convoluted mental organizations seem far-fetched, there is a precedent for such odd arrangements in the phenomenon of localization. Recall that the auditory system makes use of interaural time and amplitude differences in determining the left–right azimuth for sounds. However, interaural time difference is not accessible to consciousness or introspection. We hear that a sound is "located off to the left," not that the pressure variation in the right ear lags behind the left ear. That is, the representation accessible to introspection differs from the lower-level mental code used to derive it.[21] Biology does not follow the principle of parsimony. Mental representation is not necessarily straightforward.

Other Possible Pitch Representations

Inspired by the twelve-tone music of Arnold Schoenberg, music theorists in the latter part of the twentieth century proposed a number of new music-related representations. The representations suggested by theorists pertained mostly to pitch—or more commonly, pitch-class. Among the more popular representations, theorists would include interval vectors, Z-relations, and Tn and TnI relations.[22] All of these representations pertain to *relative* pitch-class structures, so they are better described as "interval-class" representations. Nevertheless, the designation "pitch-class" (PC) has stuck and remains the common name within music theory circles for these types of representations.

Right from the beginning, these representations spawned skepticism in certain quarters. While many theorists embraced the representations and applied them in their music analyses, other music scholars felt that the constructions were artificial and had little to do with the experience of listening. Much of the disagreement arose from a lack of clarity about the epistemological status of analytic representations in general.[23] A representation does not have to have a perceptual basis in order for it to be useful. For example, a useful representation may pertain solely to the means of musical composition. As I have argued elsewhere, there are plenty of structures in music that have nothing to do with human perception, and that does not make these structures somehow vacuous, wrong-headed, or unworthy of investigation.[24]

The problem arises when an analytic representation is presumed to serve some perceptual or cognitive function. The music analysis literature is full of intimations that a particular set structure has some sort of psychological effect or is important in the aesthetic apprehension of the work by a knowledgeable listener. These presumptions can be tested experimentally, and, over the past half-century, many have been. Pertinent experiments have been carried out by Francès (1958), Thrall (1962), Largent (1972), Lannoy (1972), Millar (1984), Bruner (1984), Gibson (1986, 1988, 1993), Samplaski (2000, 2004), and others. Each of these experiments called into question the ability of listeners to hear the pitch-set relationships posited by some theorists. Don Gibson's 1988 work is especially noteworthy since he studied thirty-two professional music theorists. Gibson showed that even professional theorists perform near chance levels in discriminating pairs of chords according to pitch-class identity.[25]

To be fair, one can't fault theorists for wanting to describe possible new ways of experiencing music. There is little glamour in describing the ordinary—the commonplace musical experiences. The good news from research in music cognition is that, if it is true that mental representations emerge from patterned exposure, then there ought to be many opportunities for artists to shape new and different ways of experiencing sounds. However, the mere proposing of a music representation does not mean that listeners are capable of experiencing music this way. In particular, brains may

have a hard time experiencing already familiar music according to novel mental representations. Moreover, the opportunities to learn new musical representations may be limited to critical periods during childhood.

The theoretical controversy aside, these studies on the perceptibility of pitch-set relations provide important clues concerning mental representation for music. They imply that certain forms of pitch-related representations may be difficult or impossible to form in the adult human auditory system, in the same way that it is difficult or impossible for most adults to acquire AP. Certain music-related representations seem to be preferred. Oddly, these representations may be preferred even though they are just approximations of a theoretical representation that would prove much better at predicting the world. What accounts for our representational preferences?

Representational Preferences

Why do brains appear to favor some types of mental representation over others? Here I'd like to propose four general principles that may influence the preferred mental representations for pitch. All four principles manifest a preference for simplicity over complexity in mental representation. In brief, the preferences are for (1) lower-order relationships, (2) neighboring over distant relationships, (3) lower-derivative states, and (4) event-related binding.

Recall from chapter 3 that a "lower-order relationship" is a relationship between a small number of elements. If the occurrence of state *X* is influenced by the presence of states *Y* and *Z*, it is said to have a higher-order relationship than if *X* were influenced by *Y* alone. The lowest order relationship is a "zeroth-order" relationship; this exists when the presence of some state is independent of any other states. For example, in English text, the most common letter is "e." That is, without considering any context, the highest zeroth-order probability is for the letter "e."

Notice that when we speak of "order relationships," we do not necessarily mean that these relationships are between adjacent or neighboring elements. An event might depend on only a single other event, but that event might be quite remote in time or space. When I push the elevator button in a high-rise building, it might be some time before the elevator appears. But despite the passage of time, the probability that the elevator will appear in response to my pushing the button is almost certain.

Mental representations also appear to favor *neighboring* over *distant* relationships. In the sequence of successive states *A*, *B*, *C*, *D* . . . it is easier to recognize a relationship between neighboring states (e.g., *A* and *B*) than more distant states (e.g., *A* and *D*). One way to think of the preference for neighboring rather than distant relationships is to suppose that it is preferrable for brains to minimize the size of temporal memory. Holding two successive events in short-term memory is easier than holding four events.

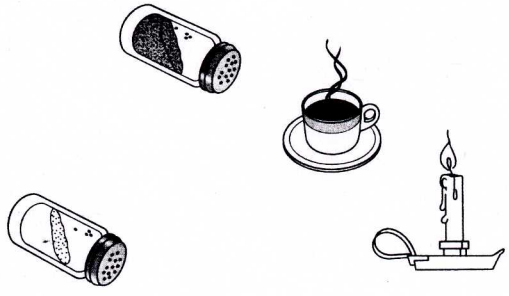

Figure 7.5
Four objects illustrating the failure to code spatial interval. Most observers fail to spontaneously notice that the distance between the salt shaker and the cup is the same as the distance between the pepper shaker and the candle.

Consider now the representation of derivative states. When characterizing relationships between two or more states, we can measure the amount of change, or the change in the amount of change, or the change in the change of the amount of change, and so on. For example, pitch successions might be represented by the amount of change (e.g., interval), or by the amount of change in the interval, and so on. In the language of calculus, we are identifying the derivatives: first derivative, second derivative, third derivative, and so on. Mental representation appears to favor lower derivatives. For example, in vision, the representation for *position* appears to take primacy over *velocity*, while *velocity* takes primacy over representations for *acceleration*, which in turn exhibits primacy over representations for *jerk*.

Finally, mental coding appears to favor representations that bind to auditory events. Consider the nonauditory analogy shown in figure 7.5. When I asked some colleagues to describe this illustration, all of them began by noting that there are four objects: what appear to be a salt and pepper shaker, a cup, and a candle. There is a special relationship between these objects. Can you see it?

The relationship is that the distance between the salt shaker and the cup is the same as the distance between the pepper shaker and the candle. Obvious, right? Of course, this relationship is not at all obvious to viewers. In parsing a visual field, we focus foremost on objects and their positions rather than distance relationships. This is not to say that minds don't apprehend distance relationships, only that our first disposition is to apprehend objects. In the case of sound, we focus foremost on sound events. In fact, observations of electroencephalographic behavior establishes that the brain responds most energetically to events, not to, say, the distance separating events. Brains are especially sensitive to the onsets of events.

In the field of perception, an important mental function is so-called *binding*. When we perceive an object, many different parts of the brain participate in calculating various features of the object. For example, in the case of vision, color is processed separately from shape, while object recognition occurs in yet another part of the brain. But when we see an apple, the phenomenal experience is unified: the shape, color, name, and other properties all adhere together as a single object.

A similar phenomenon occurs in audition. When we hear a sound, the pitch and timbre are not experienced as separate entities. We hear a "cute sound" or a "harsh sound." We don't hear "a sound" accompanied by an independent feeling of "cuteness" or "harshness." The various sound properties—loudness, pitch, location, timbre—all adhere together as a single event-object. Of course, if we are motivated to do so, we can introspect and focus on just the pitch, or just the location, for example. But the brain's disposition is to assemble an integrated "package." The object is experienced as a unified entity rather than a smorgasbord of disparate properties.

When we hear a melodic interval, the experience of interval doesn't somehow replace the two tones. We don't experience the interval as occupying the space between the two tones. Instead, the interval is manifested by the onset of the second tone—the interval feels as though it is a quality or property of the second tone. Although we tend to think of an interval as a certain kind of relationship between two tones, a melodic interval may be better described as a distinctive way of approaching a single tone. Although interval is obviously a distance relationship between two sound events, it is not some disembodied quality. The quality we call "melodic interval" is a property that binds to the ensuing tone.

A demonstration of this disposition is evident at phrase boundaries. Ask a musician friend to indicate whenever they hear a melodic leap while you sing "Happy Birthday." There is a strong likelihood that your friend will miss one or more of the leaps that occur *between* phrases. I think that event-related binding explains why the interval between the last note of one phrase and the first note of an ensuing phrase is often not salient.

My argument here is not that minds don't represent relationships. Rather, my argument is that minds favor event- or object-bound representations. It is easier to process, code, or manipulate representations when they are mentally attached to events or objects.

Preferred Representations

With these four principles of preferred mental representation in mind, let's apply these principles to some concrete examples in mental coding for music. Consider first the phenomenon of post-skip reversal. Why might listeners expect post-skip reversal rather than regression to the mean? In order to infer regression to the mean, a listener

would need to retain a cumulative distribution of all the pitches heard up to a given moment and use this distribution to infer the average pitch. That is, a regression-to-the-mean expectation would require a mental representation based on high-order relationships. All of the preceding pitches would participate in the calculation of an expected pitch. At the same time, a regression-to-the-mean expectation would also require that the representation encode long relational distances. By contrast, with post-skip reversal, a listener need retain only the previous two pitches—inferring the size and direction of pitch motion. Post-skip reversal would require calculating the first derivative—and so would involve a more complicated representation (interval or direction) compared with pitch alone. But the cost of this more complicated representation would be offset by the significant computational savings incurred by not having to maintain a complete past history of pitch heights.

What about the case of absolute pitch? Absolute pitch is the ultimate zeroth-order pitch representation. According to our four principles, it is the simplest possible pitch-related representation. So why doesn't everyone have absolute pitch? In the modern world, stable tuning systems mean that absolute pitch is useful in predicting ensuing tones. If you hear Chopin's Nocturne opus 62, no. 2, a person with AP will know that the initial B3 is followed by G#4. What could be simpler?

As I have already noted, the predictive utility of absolute pitch is limited. Nor is this limited utility merely a product of the variable tuning of voice and instruments for music-making. Consider the research suggesting that there may be a genetic component to AP.[26] If there are genes that facilitate the development of AP, then we are talking about the wrong environment if we focus on modern music-making. The vast majority of human evolution occurred during the long Pleistocene period when hominids slowly took on all of the characteristics of modern *Homo sapiens*.[27] If it is true that there is a genetic disposition that favors absolute pitch, then the sound environment under which it evolved was the Pleistocene savannah, not the salons of eighteenth-century Europe. In the Pleistocene environment, it is hard to imagine any stable absolute pitches. Early humans probably recognized the higher and lower voices of their various friends, but the pitch variability would have been large compared with the better than semitone resolution of modern AP possessors. For millions of years in human evolution, absolute pitch would have had little or no utility.

Moreover, once *Homo sapiens* began to sing, the more pertinent auditory skill would not have been absolute pitch, but a refined form of relative pitch. The triumph of relative pitch in human music-making might even have been the precipitating event for the decline of any existing absolute pitch abilities. If absolute pitch requires one or more enabling genes, then the advent of human music-making may have been the selective agent that has *reduced* the prevalence of these hypothetical genes in the human genome. Ironically, the advent of human music-making may be the reason why so few people have absolute pitch.

The existing research, however, does not favor a genetic account for explaining the variability in AP possession. The capacity for absolute pitch appears to be built into the hearing organ itself with its absolute mapping of frequency to place on the basilar membrane and its faithful phase-lock coding of the temporal fine-structure of sounds. Instead, the existing research suggests that it is the pitch stability of the childhood environment that is the principal determining factor in AP development. Most children fail to encounter an environment where AP significantly enhances accurate expectation, and so the latent capacity either withers or fails to compete with more successful codes such as relative pitch.

Our four principles also provide a plausible reason why pitch interval appears to be problematic. Interval entails calculating the first derivative. At the same time, the interval calculation requires some coding of the two pitches involved. However, if the brain simply encodes pitch-dyad instead of interval, then no derivative need be calculated. Pitch-dyad is a simpler representation than pitch interval.

These four principles of mental representation are admittedly speculative. Additional research may well establish that these suggested principles are incorrect. What they share in common is computational simplicity. Each principle either minimizes the amount of memory needed, or minimizes the number of operations required to code a state.

I think these principles are plausible for several reasons. First, neural networks are easily configured to calculate changes. One neuron can be connected to respond to the difference in potential between two other neurons. This process can be repeated so that a neuron is responding to the difference between two differences and so on. Calculating derivatives is neurologically plausible. Note, however, that each derivative recruits an additional layer of neurons. If, as the research suggests, there is competition for neural real estate, then the brain ought to favor lower-level derivatives over higher-level derivatives.

For similar reasons, the brain ought to favor calculations involving fewer states—hence favoring lower-order relationships. Higher-order relationships would require the recruitment of many more interconnected neurons. Similarly, there are plausible reasons to favor neighboring relationships over distant relationships. Distant relationships can be recognized only when a large number of states can be held in memory, and only when a large number of comparisons between states can be made.

Finally, I suspect that in the evolution of nervous systems, it was inevitable that the "divide-and-conquer" strategy would emerge in analyzing sensations. Brains can apply several different concurrent analyses to a sensation, in some cases using specialized neural tissues for different analytic functions. However, such a strategy incurs a cost: each perceptual property is free-floating and so the experience of the world is fragmented. I suspect that brains have somehow evolved to compensate for this problem

by ensuring that perceptual properties are assembled into a unitary percept. In the case of the auditory system, the profusion of perceptual properties is handled by assembling them into "a sound"—whose phenomenal existence is linked to the onset of an acoustic signal.

A Theory of Auditory Representation

The above discussion, in effect, offers a theory of auditory representation. It is appropriate to take a moment and explicitly summarize the main points of this view.

The theory proposes that the auditory system allows a number of competing neural networks to blossom—each network processing the sensory information in its own way. All of these networks ultimately start with the information provided by the auditory nerve: tonotopic position and firing-rate codings.

Expectations are neural circuits whose activation depends on the pattern of sensory inputs. In practice, these energized circuits facilitate certain motor behaviors and/or cause attention to be directed in particular ways. Motor behaviors may be facilitated directly or may be facilitated indirectly by evoking feeling states that act as motivators. Those neural circuits that predict well the ensuing events evoke a positive prediction response—a positively valenced limbic reward. Different neural representations compete with one another, and it is the prediction response that judges each performance and hands out the prizes. Networks that fail to predict ensuing states atrophy, while those that have some predictive success are retained and strengthened. In evaluating predictive neural networks, the brain follows an ancient biblical adage: the way to identify false prophets is by their false prophecies.

In the development of the auditory system, the initial representations are simple, relying on the codes passed along from the auditory nerve. These representations are low-order and make use of little or no contextual information. However, as auditory development proceeds, networks can connect with and draw information from other networks that have proved successful in predicting ensuing events. As a consequence, subsequent auditory development permits higher-order relationships and more distant contextual information to play a role in the representations that are spontaneously being generated.

In the case of pitch, for example, one might imagine that the brain begins with a simple absolute pitch representation. If a person lives in an environment where absolute pitch provides useful predictive information, then the mental coding for absolute pitch flourishes. Conversely, if the individual lives in an environment where absolute pitches provide little information of value in predicting future events, or where absolute pitch has difficulty competing with the emerging juggernauts of relative pitch prediction, then the capacity for absolute pitch withers.

Notice that the success of a representation is determined largely by the structure of the stimuli encountered by the auditory system. If the auditory environment is organized in a particular way, then mental representations that approximate that external structure will be favored.

Most of the competitive development of auditory representations occurs early in development—probably in infancy and childhood. New representational networks may continue to be spawned later in life, but they have difficulty competing with existing extensive networks that have long proved their worth.

In spawning new representations, simplicity is preferred over complexity. The brain is more likely to form mental representations that code (1) lower-order relationships, (2) neighboring over distant relationships, (3) lower-derivative states, and that (4) can be coordinated or "bound" with sensory events.

Since the auditory system must find its own way in assembling useful representations, there is plenty of scope for individual variation. However, this variation may be masked because different representations are often correlated with one another. Differences may also be masked because brains will normally rely on multiple concurrent representations. Although different listeners may respond to sounds in similar ways, they may be using very different mental codes to perform the same task.

It bears emphasizing that music-related representations exist as real biological patterns in individual brains. They aren't just formal abstractions. With advances in brain imagining, neuroscientists are beginning to show how brain organization reflects the organization of the auditory world. A landmark in music-related neural imaging is the work by Peter Janata, Jeffrey Birk, John Van Horn, Marc Leman, Barbara Tillmann, and Jamshed Bharucha. In Western tonal music, music theorists have long noted a complex web of relationships between major and minor keys. Neighboring keys are related by the "circle of fifths" while major and minor keys are related by a sequence of minor thirds. Music theorists have shown that these relationships can be graphically represented as a three-dimensional torus (donut) shape.[28] By using specially composed music that visited all of the major and minor keys through a series of modulations, Janata and his colleagues were able to observe the activation of torus-like topographies in individual listeners using magnetic resonance brain imaging. The tonal structure of the music was thus evident as a dynamic topography in a part of the rostromedial prefrontal cortex. Note that the torus tonal structure is unique to Western music. Only Western tonal music employs a system of "relative" and "parallel" major and minor keys and dominant/subdominant key-relatedness. The toroidal structures observed by Janata and his colleagues provide direct evidence of neurological adaptations to a particular musical environment.[29]

One final observation is worth highlighting from the work of Peter Janata and his colleagues. The observed toroidal structures were unique to each listener. Each listener's brain exhibited a toroidal pitch-class structure, but the precise organization differed from cortex to cortex. It was like looking at floor plans for houses. Each house might have a kitchen, living room, bathroom, and master bedroom, but the floor plans differed from house to house. These differences imply a unique learning path for each listener—consistent with the theory of neural Darwinism.

In this chapter I have introduced the subject of mental representation—principally by discussing pitch-related representations. Much of my discussion has centered on absolute pitch and interval. From time to time I have also alluded to the role of scale-degree representation, but without much in the way of detail. In chapter 9 we turn our attention exclusively to the phenomenon of scale degree. Scale degree is almost certainly the most musically important mental representation related to pitch for Western-enculturated listeners. But first, let's return to the subject of emotion.