MIT Press books may be purchased at special quantity discounts for business or sales promotional use. For information, please email special_sales@mitpress.mit.edu or write to Special Sales Department, The MIT Press, 55 Hayward Street, Cambridge, MA 02142.

This book was set in Stone Sans and Stone Serif by SNP Best-set Typesetter Ltd., Hong Kong, and was printed and bound in the United States of America.

Library of Congress Cataloging-in-Publication Data

Huron, David.
Sweet anticipation : music and the psychology of expectation / David Huron.
p. cm.
"A Bradford book."
Includes bibliographical references (p.) and index.
ISBN 0-262-08345-0 (hc : alk. paper)
1. Music—Psychological aspects. 2. Expectation (Psychology). I. Title.

ML3838.H87 2006 781′.11—dc22 2005054013

10 9 8 7 6 5 4 3 2 1

Contents

3 Measuring Musical Expectation

Without some way of gathering information about what individuals expect, all theories of expectation would remain purely speculative. If we want to hold our views accountable, we must be able to compare theories with evidence about how real minds anticipate the future. How then, do we go about determining what someone is expecting?

This question raises a host of related questions. What does it mean to have an expectation? How precise are expectations? Do we expect specific events, or do we expect "classes" or types of events? Can a person truly anticipate more than one possibility at a single time—that is, is it possible to have "plural" expectations? How do expectations manifest themselves as psychological or physiological states? How would one go about measuring what a person expects?

One definition of expectation might classify it as a form of mental or corporeal "belief" that some event or class of events is likely to happen in the future.[1] Such "beliefs" are evident in a person's "action-readiness"—that is, changes of posture, metabolism, or conscious thought that prepare the individual for certain possible outcomes but not for others. Such expectations can differ in strength of conviction or certainty.

Over the past four decades, researchers have devised a number of methods for gauging or estimating what people expect. The purpose of this chapter is to describe some of the experimental methods used to characterize listener expectations. Many of the same techniques are used to characterize nonauditory expectations, such as visual expectations. In addition, we will introduce some useful concepts from probability and information theory. These concepts will provide a convenient quantitative method for characterizing the range of expected possibilities, and for expressing the relative strength of conviction or certainty for various expectations.

Experimental Methods in Expectation

At least eight different experimental methods have been used to characterize a listener's expectations. Each method has strengths and weaknesses. Some methods are laborious whereas others are easy; some give fine-grained detail about the relative strengths of various possibilities, where others merely indicate that some outcome is possible for one listener. Some methods require the listener to reflect and introspect; others require no conscious thought at all. Some require the listener to be musically skilled; other methods require no special skills. Some methods are suitable only for adult listeners who can communicate verbally; other methods can be used with preverbal infants and nonhuman animals. Some methods require that the sound experience be periodically halted; other methods can be used without interrupting the listening experience. Each method is able to provide some useful information, but no method is a panacea. Becoming familiar with the different methods will help us better interpret the various experimental results, and also help us understand why different methods sometimes produce diverging—sometimes even conflicting—results.

It is important to understand that none of the following methods measures "expectation" in any direct sense. "Expectation" is a theoretical construct whose meaning and definition is open to debate. In experimental research, theoretical entities are rarely directly observable. Instead, researchers must operationally define some measureable quantity that is assumed to correlate with the theoretical construct. In reading the descriptions of the following experimental methods, it will become obvious that each measurement method is open to the legitimate charge "that's not what expectation really is." But in interpreting the experimental results we need to maintain some perspective. No method will capture the entire essence of expectation, but on occasion, some methods will allow clearer glimpses of how "expectation" operates.

1 Method of Tone Detection

Perhaps the earliest method for measuring auditory expectation was devised by Gordon Greenberg and Willard Larkin in the 1960s.[2] Working at the University of Illinois, Greenberg and Larkin had participants listen to tones in the presence of continuous loud noise. The listeners' task was simply to indicate whether or not they heard the tone.

Greenberg and Larkin discovered that listeners were better able to detect a tone if they expected a tone of a specific pitch to occur at a particular moment. They found that expectation allows listeners to direct their attention in both frequency and time: this directed attention has the effect of lowering the threshold of sensation for the sound. Greenberg and Larkin showed that there was a band of frequencies that was facilitated when listeners were expecting a particular frequency. If a listener expected a tone of 500Hz, she was still able to detect a partially masked 550Hz tone, but not a tone of 800Hz. In effect, Greenberg and Larkin showed that listeners can direct their attention at particular frequency regions and time spans, and they used the method of detection to determine the shape and width of these "attentional bands."

Working at the Catholic University in Washington, Jim Howard, Alice O'Toole, Raja Parasuraman, and Kevin Bennett extended this method so that listeners were asked to detect a tone in some patterned context.[3] Listeners heard a twelve-note sequence presented along with a concurrent sustained noise. Two presentations of the sequence were given. One presentation was complete; the other presentation was missing one of the tones. Listeners were asked to indicate which of the two presentations was complete. They established that the preceding sequence of tones facilitated the detection of some tones but not others. That is, they showed that the *melodic context* influences where listeners direct their attention.

The method of tone detection is rarely used today in experiments related to auditory expectation. But the work of Greenberg and Larkin remains important because it demonstrates two general principles concerning expectation. First, *accurate expectation facilitates perception*. When the events of the world conform to our expectations, we are better able to detect, perceive, and process these events. Over the past half century, this facilitating effect has been observed many times.[4] It is a principle that holds for both visual as well as auditory events. A second lesson from Greenberg and Larkin is that low-level sensory processes (like the hearing threshold for detecting a tone) are influenced by higher-level mental processes (like expectation). It is as though higher mental functions are able to reach down into the sensory apparatus and do some fine-tuning. Sensory systems don't just present information to the higher mental functions; in addition, higher mental functions can reconfigure a sensory system to focus on particular aspects of the sensorial world.

2 Method of Production

At the University of Washington in Seattle, James Carlsen, Pierre Divenyi, and Jack Taylor pioneered the simple technique of having listeners sing a continuation to some interrupted musical phrase.[5] Carlsen and his colleagues simply played a sequence of tones and asked listeners to sing what they thought would be an appropriate continuation. Carlsen used this method to compare the melodic continuations of American, German, and Hungarian listeners.[6] In analyzing the sung continuations, Carlsen found significant differences between the three groups, suggesting that one's cultural background influences listener expectations of what might happen next.

This method has a number of disadvantages. Notably it requires that participants have some singing ability (and be willing to sing while being recorded).[7] The method also relies on the participants' facility and comfort with improvising. Sung continuations can be confounded by vocal constraints. For example, if the antecedent context is low in pitch compared with the singer's vocal range, then there will be a natural tendency for the singer to produce a continuation that rises in pitch. Conversely, if the antecedent context is high in pitch compared with the singer's vocal range, then there will be a tendency for the singer to produce a continuation that falls in pitch.

Thus the melodic contour will reflect the participant's vocal range, rather than general melodic trends. (This problem can be controlled to some extent by determining the singer's vocal range prior to the experiment, and then tailoring the stimuli so they are positioned near the center of the participant's range.) When the antecedent context is short (such as a two-note interval), it may be impossible to infer the key that a participant might be using. In experiments by William Lake, this problem is eliminated by playing a tonic-establishing cadence before the start of the stimulus.[8] Another problem relates to deciphering what a singer sang. When singers are not well trained, the pitch and timing is often quite ambiguous. This introduces onerous technical challenges for the experimenter when transcribing what pitches and durations a singer produced or was intending to produce.

A variant of the method of production has been used by Mark Schmuckler. Instead of having participants sing, Schmuckler asked pianists to perform a continuation on a keyboard. Compared with sung continuations, the use of a keyboard circumvents the problems of pitch transcription, but it introduces a potential problem with skill. In order to reduce the need for keyboard skills, participants are given opportunities to try several different continuations, and rehearse their preferred continuation until they are satisfied with their response. This variation of the method of production has been popular with those researchers, such as Dirk-Jan Povel, who want to reduce the uncertainties introduced by pitch transcription.[9]

Yet another variant of this technique has been developed by Steve Larson at the University of Oregon. Larson contrived a task involving musical notation and expert music theorists.[10] He simply provided a notated antecedent context and asked music theorists to compose a suitable melodic continuation. In some ways, music theorists are ideal participants since they can draw on a lot of experience and know how to respond precisely. But music theorists often pride themselves on being musically clever, so a potential danger with this approach is that participants will be tempted to compose technical flights of creative fancy rather than commonplace or more intuitive melodic continuations. Larson explicitly instructed his participants to compose what they thought would be the most common or obvious continuation. As an inducement to this end, he offered a prize whose winner would be selected from the group of theorists who wrote the same (most frequent) continuation. That is, he gathered all of the responses, identified the most commonly occurring response, and then awarded a prize to one of the theorists (drawn at random) who had composed the most common continuation. One advantage of this approach over Carlsen's method is that it is possible to use complex harmonic or polyphonic stimuli (and responses) that would not be suitable for a sung response. However, there are a number of disadvantages to this method. The principal disadvantage is that notationally literate musicians often hold their own theories about melodic organization, and so the responses hold the potential to be confounded by theoretical preconceptions. Like Schmuckler's keyboard task, the notation task lacks the spontaneity of improvised singing. That is, it encourages conscious, contrived, and reflective responses.

As we have seen, the principal drawback to the method of production is that participants in the experiment must have considerable musical competence—as vocalists, keyboardists, or by having facility with musical notation. This tends to limit the technique to participants who are relatively musical. Further, in requiring participants to "perform," the method of production also assumes that expectation facilitates not perception but *motor production*. Finally, Mark Schmuckler has pointed out that the method of production also requires a certain degree of conscious attention, whereas under normal listening conditions expectations may be largely unconscious and effortless.[11]

Compared with other methods, a unique benefit of the method of production is that it doesn't artificially limit a participant to producing a single tone following the given musical context. That is, whereas many other experimental methods assume that the preeminent expectation will pertain to the immediately succeeding tone, the method of production readily allows a participant to suggest several continuation notes as a coherent group. Later, we will see evidence indicating that a listener's strongest expectation may relate to an event that does not occur until after several intervening tones. The method of production provides better opportunities for an experimenter to study such possible long-term expectations, rather than focusing exclusively on note-to-note relationships.

3 Probe-Tone Method

Without question, the best-known experimental method for testing musical expectations is the probe-tone method pioneered by Roger Shepard and Carol Krumhansl.[12] Krumhansl and her colleagues at Cornell University have carried out numerous experiments using this technique. In simple terms, a musical context is presented—such as several chords or the initial notes of a melody. Following this context, a single tone or chord is played, and the listener is asked to judge this target (or "probe") sound according to some criterion. Often, the listener is asked to judge how well the tone or chord "fits" with the preceding musical context. The original contextual passage is then repeated and a different probe tone or chord is played. Following each presentation, the listener is asked to judge how well the new tone or chord fits with the preceding context.

In probe-tone experiments, a dozen or more repetitions of the same contextual passage may be presented—each presentation followed by a different probe. For example, several dozen possible continuations (probes) might be presented on successive trials. In this way, numerical ratings can be gathered for a large number of possible continuations. Thus, a significant advantage of the probe-tone method is that a detailed picture can be assembled where the listener provides information concerning several possible continuations, rather than only a single continuation. For example, with the probe-tone method a given participant might judge two or three continuations equally good. In the method of production, by contrast, participants must choose just one continuation, so it is problematic for the experimenter to infer that several different continuations would be equally acceptable for the participant. In addition, the probe-tone method can also establish which continuations sound "bad." That is, the method can be used to identify implausible as well as plausible continuations.

An obvious difficulty with the probe-tone method is that it is tedious. Each possible continuation must be tested separately. In practice, the total range of possibilities is reduced by the experimenter. For example, there are 88 tones available on a piano, but most of these tones are unlikely candidates to follow some melodic passage. Most tones will be implausibly high or low. Typically, the experimenter will reduce the candidate pitches to those within a two-octave range (one octave above or below) of the current pitch.

Yet another way to limit the number of possible continuations is to use so-called Shepard tones as probes. Shepard tones are specially constructed complex tones consisting of octave-spaced partials spanning the entire hearing range.[13] This encourages the listener to judge "goodness of fit" according to pitch-class rather than according to a single pitch. In Western music, there are only twelve pitch-classes, so using Shepard tones reduces all possible pitch continuations to just twelve. Using Shepard tones, however, means that the experimenter cannot directly infer the pitch direction (or contour) expected by the listener, given that each pitch class represents several possible pitches.

Apart from the tediousness of the probe-tone method, another difficulty is that it stops the music. When a listener judges "goodness of fit" one might imagine the response to arise from a combination of two sorts of judgments: (1) how well does this tone follow the previous note? and (2) how well does this tone terminate the tone sequence? Theoretically, it is possible that a tone follows well from the previous tone, but it might be judged as a poor fit because it evokes little sense of perceptual closure or completion. Conversely, a tone might follow poorly from the previous tone, yet evoke a strong sense of tonal closure, and so be rated highly by listeners. We will have more to say about these divergent interpretations in chapter 9.

Progressive probe-tone method In some cases, exhaustive experiments have been carried out to trace the changes in the listener's experience as the music progresses. For example, the first three notes of a melody may be played, followed by a probe tone. This procedure is repeated until a large number of continuation tones have been probed. Then the first *four* notes of the melody are played, again followed by one of several probe tones. This procedure continues for the first five notes, six notes, and so on. The progressive probe-tone method has been used to trace in detail such phenomena as how a modulating chord progression begins to evoke a different tonal center.[14]

Continuous probe method An obvious difficulty with the progressive probe-tone method is the tediousness of repeating the stimulus for each probe tone. If twelve probes are used following each note, a simple eight-note sequence will require 96 repetitions of the simulus in order to map the changing expectations over the course of the passage. In 2002, Carol Krumhansl and her colleagues introduced a variation of the probe-tone method in which a single probe tone (or chord) is sustained throughout the passage and the listener provides continuous responses as to the appropriateness of the probe at each moment in time.[15]

A problem with the continuous probe method is that it is hard to regard the responses as relating to expectations. Suppose, for example, that a tonic pitch is sounding continuously throughout a passage. As a cadence approaches, the dominant chord might sound. However, the harmonic clash between the dominant chord and the tonic pitch is not likely to result in a high rating for the tonic. Yet, one might presume that following a dominant chord, a high rating would be expected for the tonic. By comparison, if the penultimate chord is a subdominant chord, the probe-tone tonic is likely to receive a very high rating (because it is consonant with the sounding chord). Yet the dominant chord may well have evoked a greater expectation for an ensuing tonic than is the case for the subdominant chord. Said another way, one would expect the responses to continuous probe tones to be confounded by the resulting harmonic congruence: harmonic congruence is apt to play a much stronger role than expectation for subsequent events in determining a listener's response.

4 Betting Paradigm

Although the various probe tone methods do provide some information about the magnitude of various expectations, it would be useful to gather more precise measures of the subjective probabilities of different outcomes. In the *betting paradigm*, participants are given a "grub stake" of poker chips and asked to place bets on a set of possible continuations. Participants hear an antecedent passage and are invited to bet on what pitch they think will occur next.

I and my collaborators, Paul von Hippel and ethnomusicologist David Harnish, used this approach to compare the expectations for two cultural groups—Balinese musicians and American musicians. The experiment works as follows. Bets are placed on the keys of a mock-up of an instrument (in our experiment, a Balinese *peng ugal*). Bets need not all be placed on a single outcome. Instead, participants are free to distribute the poker chips across several possible continuations—varying the number of chips wagered according to the degree of certainty or uncertainty. Bets placed on the correct pitch are rewarded tenfold. Bets placed on incorrect pitches are lost. Participants are instructed to try to maximize their winnings.

As in the progressive probe-tone method, responses (wagers) can be collected following each note of a melody. The experiment begins with the participant hearing the first note of the melody while the pitch is indicated on a computer monitor. The participant is then invited to bet on what she or he thinks will be the second note. Once bets are placed, the actual second note is revealed, the winnings tabulated, and a sound recording of the melody is played, stopping before the third note. The participant is then invited to bet on what she or he thinks will be the third note. This process is repeated until a complete melody has been revealed.

In our experiment both American and Balinese musicians were tested on a traditional Balinese melody.[16] Throughout the experiment, participants could see the notation up to the current point in the melody, and could try out different continuations using a digital keyboard sampler that emulated the sound of the *peng ugal*. The betting context helped participants consider other possibilities apart from the first one that came to mind.

The principal benefit of the betting paradigm is that it allows the experimenter to calculate the subjective probabilities for different continuations. Assuming that the participant is behaving rationally, bets should be placed in proportion to the subjective likelihood of subsequent events. For example, if a participant thinks that a certain pitch is twice as likely as another pitch, then the participant ought to place twice as large a bet on the more probable pitch. Later in this chapter we will discuss how information theory provides a useful way to quantify such subjective probabilities.

A related advantage of the betting paradigm is that it allows the experimenter to measure confidence independent of relative probabilities. Suppose, for example, that the Balinese and American participants had roughly the same expectations ("A" is more likely than "B" which is more likely than "C"), but differed significantly in their confidence. A lack of confidence would be evident by participants' spreading their bets out more evenly. Once again, this can be calculated using information theory.

There are also several problems with the betting paradigm. For one thing, the procedure is even more tedious than the progressive probe-tone method. On average, we have found that it takes roughly three minutes for participants to complete their wagers for each note in the melody. A thirty-five-note melody thus can take as long as two hours to complete. Fortunately, the majority of our participants report that the task is fun, and that the gambling aspect of the task is highly motivating.

Like the notated version of the method of production, the betting paradigm encourages a conscious-reflective response rather than a spontaneous response. Unlike the method of production, the participant knows that there is a real melody involved, and is motivated to correctly anticipate the next note. Even more so than the method of production, the betting paradigm requires a degree of musical skill. In this regard, the probe-tone method is notably superior, not requiring any musical expertise.

A further problem with the betting paradigm is that the data may be confounded by a learning effect. Since the participants receive constant feedback about the accuracy of their wagers, they are likely to become progressively better at placing their bets as the experiment continues. Typically, experiments last more than an hour, so there is plenty of opportunity to improve and refine one's betting skills. Any apparent decrease in uncertainty as the melody progresses may therefore be an artifact of becoming a more savvy gambler.[17]

5 Head-Turning Paradigm

When we hear an unexpected sound, we will often turn our head in the direction of the sound. This basic reflex is referred to as the *orienting response*, and it is evident in all vertebrates, including young infants and adults. If a stimulus is repeated, after a while an individual will *habituate* to the stimulus and fail to orient to it. Further repetitions are unlikely to provoke a response. If a change is then made to the stimulus, and if the change is sufficiently novel, then a listener might reorient to the sound. This reorienting to a modified stimulus is called *dishabituation*. If an infant reorients to a modified sound, then one might interpret this as evidence that the infant didn't expect the sound.

Experiments employing a dishabituation paradigm typically repeat a stimulus until the participant becomes habituated to it. When habituation is complete then a new stimulus is introduced. If the new stimulus is perceived as the same (or similar) to the preceding stimuli then the participant will typically show no dishabituation. Conversely, if the new stimulus is perceived to differ from the preceding or expected stimuli then the participant is likely to show evidence of dishabituation or orienting to the stimulus.

The dishabituation paradigm is typically used when studying preverbal infants or nonhuman animals. The paradigm is used less commonly among adults since adults can verbally report perceived similarity or difference.

In research with infants, the head-turning paradigm has proved quite popular. There are a number of variants of this experimental method. In some cases, the experimenter merely tabulates whether or not a participant reorients to a modified stimulus. Another variant of the method measures the duration of orienting. This assumes that the greater the discrepancy between the expected and actual stimulus, the longer an infant will look in the direction of the stimulus. Yet another method will compare how quickly an individual habituates to two related events.[18]

Most of the experimental methods described in this chapter are not suitable for use with young children or infants. The principal advantage of the head-turning paradigm is that it can be used with infants and nonhuman animals. Also, unlike the bradycardic method (see below), it requires no special instrumentation apart from a video camera. One problem with the head-turning paradigm is that it requires that the participant first become habituated to the stimulus before some change is made. This makes the procedure extremely time-consuming. Newborn infants have difficulty controlling their head movements, so typically, the head-turning paradigm cannot be used until the infant is at least two or three months old.

An example of an experiment using the head-turning paradigm is one carried out by Michael Weiss, Philip Zelazo, and Irina Swain.[19] Weiss and his colleagues had infants listen to a nonsense word repeated until they habituated to it. The infants then heard either the original sound or one of four variants in which the pitch had been altered. Specifically, the frequency was modified by either 7, 14, 21, or 28 percent. They found that infants were most likely to reorient to the stimulus when the frequency of the nonsense word had been altered by 14 percent or more.

6 Bradycardic Response Method

Another version of the dishabituation paradigm examines changes of heart rate rather than head movements. When a stimulus deviates from an expected stimulus or attracts the attention of an individual, a measureable reduction of heart rate is often observed. Typically, such stimuli will result in a reduction of heart rate of about two to four beats per minute, followed by a recovery back to the normal rate. This response is referred to as *bradycardia*. Bradycardic changes of heart rate are associated with *interest* and *attending* to a stimulus.

Like the head-turning paradigm, the bradycardic response method is useful for studying the expectations of nonhuman animals, and especially for studying preverbal infants. Unfortunately, the method is tedious and the equipment can be cumbersome. Unlike the probe-tone method and the betting paradigm, each trial gives comparatively little information, and building a picture of infant expectation may require hundreds of trials from dozens of participants.

7 Reaction Time Method

Recall that Greenberg and Larkin showed that accurate expectation facilitates perception. When you hear an expected sound, you will typically be able to process it more quickly and respond to it faster (if a motor response is required). A quick reaction time is therefore correlated with high expectation. While reaction-time measures have long been used in experiments related to visual expectation, the method has gained favor only recently in research on auditory or musical expectation. Bret Aarden has shown that the method shows great promise for the study of melodic expectation.[20]

Aarden concocted a task where listeners were required to process and respond to sounds as quickly as possible. While they listened to an ongoing melody, Aarden asked his listeners simply to indicate whether the pitch contour of the melody had ascended, descended, or remained the same. In this task, there are three alternatives, and the listener must press one of three keys as quickly as possible after each note in the sequence. The responses are collected, including the elapsed time between the onset of the heard tone and the key press. The method is based on the assumption that if the pitch contour of a note moves as expected, then this will have a facilitating effect and so produce faster reaction times. Conversely, if a tone moves to an unexpected tone, this will increase the processing time and so result in a slower reaction time.[21]

The reaction time method has two notable advantages over other methods we have seen. First, it can be used in a continuous listening task where data are collected after every tone (except the first). The task is fairly difficult, so stimulus melodies are typically played using a tempo that is about 60 percent of the normal speed.[22] However, apart from the reduced tempo, there are no musical interruptions or pauses. As we will see later, compared with other methods, this method reduces the closure confound—where long pauses encourage listeners to respond to how well the tone provides a good ending point. A related advantage is that data collection is much faster than other methods. Rather than spending an entire experimental session on a single melody, a single session can collect data for dozens of melodies or tone sequences representing many different contexts. Another advantage is that the task happens so quickly that it is difficult for participants to engage in conscious reflection.

The reaction time method also has a number of drawbacks. First, unlike the betting paradigm or the probe-tone method, the reaction time method does not collect data for all of the different possibilities at each moment. Instead, we have a record of the processing time for specific contours within the heard sequence. When a listener makes a slow response, we have no idea of why this occurs—except that the heard tone evoked a more time-consuming mental process. There is no explicit information to tell us which alternative stimulus might have been processed more quickly. Finally, we must note that the reaction time method is fully premised on the idea that accurate expectations facilitate response times. At the moment, not enough is known about this relationship to know what sorts of confounds might be lurking.[23]

Despite these problems and caveats, the reaction time method shows excellent promise. As we will see in chapter 9, the results of the reaction time method and the probe-tone method sometimes diverge dramatically. We will see that these differences illuminate some important aspects of auditory expectation.

8 Evoked Response Potential (ERP)

The activity of neurons results in tiny electrical currents. When large numbers of neurons are active at the same time, the aggregate electrical current can often be detected through the scalp using suitably sensitive electrodes. The complicated hills and valleys of electroencephalographs have been studied for decades. Most of the activity remains an enigma; however, a consensus has slowly emerged about the interpretation of several specific features.

The most pertinent research related to expectation involves those electrical patterns that arise in response to a particular stimulus, like a tone. Since the recorded brain activity is in response to a stimulus, the ensuing electrical behavior is referred to as an *evoked response potential* or ERP.

Researchers still have difficulty interpreting individual ERP recordings. Typically, researchers average together many trials in which the same stimulus condition exists. It is the averaged data set—sometimes averaged across many subjects—that is able to tell a story. After the onset of the stimulus, a characteristic sequence of peaks and troughs can be observed in the ERP data. For convenience, successive peaks are designated P1, P2, P3, and so on, while successive troughs are designated N1, N2, and so on.

Suppose that a repeated sequence of identical sounds is interrupted occasionally by a deviant sound. Typically, this change is reflected in the listeners' response as an increased amplitude of the N2 waveform, which usually peaks around 100 to 250 milliseconds following the occurrence of the deviant sound. Because the electrical potential is negative and because it occurs in response to stimuli that fail to match the expected sound, the event is referred to as a *mismatch negative* or MMN. MMNs can occur in response to changes of pitch, changes of loudness, and changes of timbre, among other things. MMNs also occur if any expected tone is replaced by a silent rest. Interestingly, MMNs occur if the listener *expects* to hear a change in the sound even if the sound remains unchanged.[24]

MMNs have been observed in listeners who are asleep and even in anesthetized rats, so the mismatch negativity can be evoked without conscious awareness of changes in sound. The magnitude of the effect is influenced, however, by attentiveness. For example, distracting tasks will attenuate the MMN.

The location in the brain of the evoked MMN is known to change depending on the type of stimulus change. For example, a change in frequency will evoke the largest MMN response along the sides of the head (temporal cortex). However, suppose an auditory pattern A–B–A–B–A–B is interrupted by the repetition of either A–A or B–B. In this case the site of maximum MMN response shifts toward the top of the head.[25]

ERP methods have a number of advantages and disadvantages. Like the head-turning paradigm and the bradycardic response method, ERP can be useful for studying the expectations of nonhuman animals, and especially for studying preverbal infants. No verbal responses are necessary, and no conscious awareness or thought is required. Unfortunately, many trials must be averaged together in order to infer anything.

Subjective Probability and Uncertainty

Having described the major experimental techniques for characterizing a listener's expectation, let us now turn to a second question: How do we express the strength of conviction or certainty of an expectation?

Suppose you are a participant in a betting paradigm experiment. There might be ten possible notes, and your task is to place your bets according to how likely you think each outcome is. There are lots of possible circumstances. If you are completely clueless, the best strategy would be to simply spread your poker chips evenly across all ten possibilities. Conversely, if you are completely certain of the outcome, the best strategy would be to place all of your chips on the expected note. In the first case, the bets reflect that you are completely uncertain of the outcome, whereas in the second case the bets indicate you are absolutely certain. Of course, there are many intermediate situations. You might be pretty sure that a certain outcome will happen, but you are less than 100 percent certain. In this case, you might place a small bet on all ten outcomes, but place most of your chips on the one outcome you think is most likely. Alternatively, you might be absolutely certain that one of the ten outcomes will *not* happen. In this case, you might spread your bets evenly across the remaining nine possibilities—placing no chips on the outcome you are certain won't occur. A more complicated situation might arise if you are moderately certain that only three outcomes are likely. Here you might split the majority of the poker chips between the three most likely outcomes, while placing small bets on the remaining possibilities.

From a research perspective, it would be convenient if we could distill any complex arrangement of bets into a single number representing the overall degree of certainty or uncertainty. Such a summary measure is provided by information theory. Using the so-called Shannon–Weaver equation, any arrangement of probabilities can be summarized by a single value that represents the aggregate uncertainty—measured in *bits*. When the number of bits is high, it means that the bets (probabilities) represent a high degree of overall uncertainty. Conversely, a low number of bits is indicative of high certainty.[26]

We won't bother to explain the equation here. However, the flavor of using bits to characterize uncertainty can be conveyed by some examples. Suppose we want to characterize the uncertainty of tossing a fair coin. With two possibilities (heads or tails) the amount of uncertainty is precisely 1 bit. Think of a bit as equivalent to one "yes-or-no" question: Did the coin come up heads? A single yes-or-no answer to that question is all we need to know to resolve the uncertainty. Similarly, if there are four equally likely outcomes, the number of bits is 2. If we have to choose from 8 equally likely pitches, then the uncertainty is 3 bits (2 × 2 × 2).

In the case of the von Hippel, Huron, and Harnish experiment described earlier, the *peng ugal* instrument provided ten possible tones. If all ten tones were equally

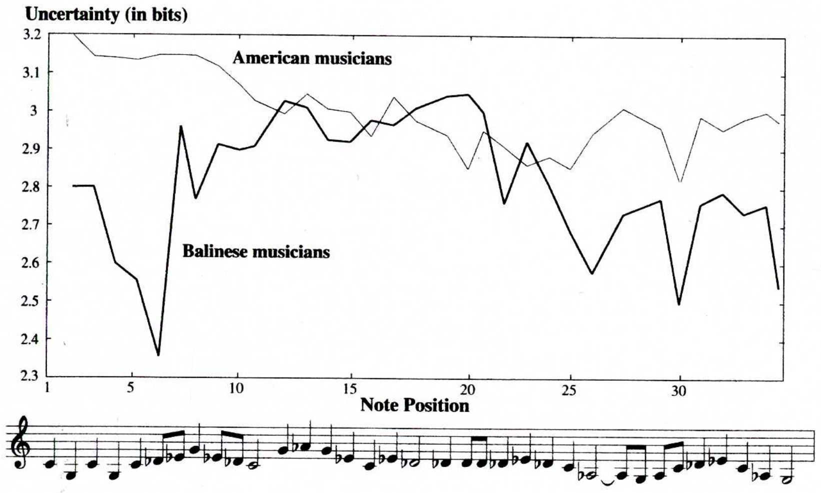

Figure 3.1
Average moment-to-moment uncertainty for Balinese and American musicians listening to an unfamiliar traditional Balinese melody. Uncertainty is plotted as entropy, measured in bits. In general, Balinese listeners show less average uncertainty. Note positions correspond with underlying notational rendering. The pitch levels shown in the notation are only approximate.

probable, then the uncertainty would represent 3.32 bits. A completely clueless listener (acting rationally) would place equal bets on all ten notes, and the Shannon–Weaver equation applied to this arrangement of bets would result in 3.32 bits of uncertainty.

Figure 3.1 plots the average uncertainty (expressed in bits) for the Balinese and American listeners in our experiment. Notice that after the first note, the average uncertainty for the American musicians was 3.2 bits. Since maximum uncertainty for ten outcomes is 3.32 bits, this means that after hearing the first note, the American musicians were almost perfectly clueless about what might happen next. After hearing the first note, the average uncertainty for the Balinese listeners was 2.8 bits, which corresponds almost precisely with seven equally probable states. This advantage is equivalent to being able to eliminate three of the ten notes as possible successors. By the fifth note of the melody, the average uncertainty for the Balinese musicians was roughly 2.35 bits. This is equivalent to being able to exclude five of the ten notes as possible successors.

Notice that by about ten notes into the melody, the American musicians are now comparable in confidence to the Balinese musicians in placing their bets. However, the Balinese musicians continue to exhibit less uncertainty—especially as the end of the melody approaches.

It is important to understand that figure 3.1 portrays average *uncertainty*—not predictive *success*. One can be confidently wrong as well as confidently right. In this case the Balinese musicians were not only less uncertain than the American musicians; they were also more accurate in their bets. A simple summary measure of predictive success is to compare "winnings." We started our participants with a nominal grubstake of $1.50 (not real money—for reasons that will soon become apparent). We rewarded accurate bets tenfold, while inaccurate bets were lost. If a player simply distributed the bets equally across all ten notes on the *peng ugal*, and left them there throughout the melody, then the final winnings would be the same as the initial $1.50 grubstake.

With regard to predictive accuracy, the differences between the American and Balinese musicians were striking. By the end of the melody, the most successful Balinese musician had amassed a fortune of several million dollars. The most successful American musician failed to do as well as the least successful Balinese musician. Moreover, several American musicians went bankrupt during the game and had to be advanced a new grubstake in order to continue.

Not surprisingly, Balinese musicians do better than American musicians in forming accurate expectations related to a Balinese melody. Although the specific melody was unfamiliar to both the Balinese and American listeners, the Balinese were able to take advantage of their cultural familiarity in forming suitable melodic expectations. Familiarity with a musical genre leads to both more accurate expectations and less uncertainty. However, it would be wrong to conclude that the American musicians were utterly clueless when listening to Balinese music. On average, the American listeners performed much better than chance. Either the American musicians were able to adapt quickly to the unfamiliar music, or they were able to successfully apply intuitions formed by their extensive experience with Western music—or both.

Conditional Probabilities—The Role of Context

In casino gambling, there is no link between a previous outcome and a future outcome. Each time we roll a pair of dice, the number that is rolled is independent of numbers previously rolled (this is true even for loaded dice). But in many real-world phenomena, subsequent probabilities do depend on preceding states. The probability of the occurrence of the letter "u" in text increases considerably when the preceding letter is "q." Likewise, in tonal music, the probability of occurrence of the tonic degree increases when preceded by the leading tone (the seventh degree of the scale). When the probability of an event is dependent on some preexisting state, it is referred to as a *conditional probability*.

In describing conditional probabilities, two concerns are the *contextual distance* and *contextual size*. Some states are influenced only by their immediate neighbors (i.e., small contextual distance). Other states are influenced only by states that are far away in space or time (i.e., large contextual distance). At the same time, states might be influenced by just a few other states or by a large number of other states. The size of the context of probabilitistic influence is sometimes also called the *probability order*. When the probability of occurrence for elements is totally independent of preceding elements (as with fairly thrown dice) the probability order is called the *zeroth order*; context sizes that take into account a single preceding element are called *first order*; *second order* denotes the probability order in which two preceding elements are taken into account, and so on.

It is important to note that the contextual size or probability order is independent of the contextual distance. Some event or state might be constrained only by its immediate neighbor (near context, small order). If an event is constrained by many neighbors, it will have a near context and large order. If an event is constrained by the presence of a single distant event, then it will have a distant context and small order.

By way of illustration, consider the following four contrasting examples:

1. Far context, small order A worker who receives a bonus might decide some weeks later to go shopping for a new jacket. Here, the likelihood of a future event (purchasing a jacket) is constrained by a single, somewhat distant earlier event.
2. Near context, small order Hearing her child cry, a mother might pick up the child. Here the future event (picking up the child) is evoked principally by a single immediately preceding state.
3. Near context, large order At a bingo parlor, a winner shouts out "bingo!" This event is provoked only by many preceding events, each of which caused another number on the card to be marked or covered.
4. Far context, large order A talented scientist might carry out a number of experiments leading to a major discovery that many years later results in her receiving a Nobel prize. The prize arose from many activities that were carried out decades earlier.

As we will see in later chapters, music exhibits a complete range of such dependencies. Most of the time, the principal constraints are of low probability order and involve a near context (e.g., one note influences the next note). But music also exhibits distinctive patterns of organization where distant contexts are more influential than near contexts and the probability order is quite large.

Reprise

In this chapter we have covered some basic background that will help us in discussing some of the experimental research pertaining to expectation. In the first instance we have described eight experimental paradigms used to characterize listeners' expectations. None of the methods is without difficulties. Each method makes different assumptions and provides subtly different information.

In addition, we have shown how information theory provides useful tools for measuring the strength or uncertainty of an expectation. We have also provided some conceptual language that will help us describe how the occurrence of a particular event might be shaped by other neighboring or distant events.

As mentioned above, these statistical patterns are absorbed with little or no conscious awareness by listeners. In particular, the experiments with eight-month-old infants indicate that much, most, or all of this learning must occur at a preconscious level. Without the involvement of consciousness, there are not many plausible alternatives to the statistical learning theory. Without the sophisticated mental resources provided by consciousness, one can readily see why brains might rely on simple probability of occurrence: in predicting a future stimulus, our best prediction would be the stimulus that has occurred most frequently in the past. Similarly, given a particular sequence of tones, the most likely next tone is that tone which has most frequently followed after the antecedent context in the past.

At this point I don't expect readers to be convinced that auditory learning is dominated by statistical exposure. (More evidence will be presented in the ensuing chapters.) But with this initial evidence, I hope you will be at least willing to entertain statistical learning as a real possibility. Given this viewpoint, let's see where it takes us.

5 Statistical Properties of Music

As we saw in the previous chapter, listeners appear to be sensitive to the frequencies of occurrence of different auditory events. In Jenny Saffran's experiments, listeners heard tone sequences whose statistical properties were artificially contrived. If we want to understand music-related expectations, then we need to identify the statistical regularities evident in real music.

There are lots of stable probabilitistic relationships that can be observed in music. Some patterns are unique properties of individual musical works. For example, a repeated musical motive or theme may become a statistical feature of a particular composition.[1] Other probabilities appear to reflect properties of particular styles or genres.[2] Yet other probabilities appear to reflect properties of music as a whole. We might begin our musical story by looking for statistical regularities that seem to characterize Western music in general. More narrowly, we might begin by restricting our discussion to melodies.

Mental Representation

Before continuing we might ask what is it that listeners represent when they form mental analogues of probability structures? For example, are tone sequences represented as pitches or as intervals? Are melodies mentally represented as up/down pitch contours? Or coded as pairs of successive scale degrees? Or perhaps music is mentally stored as a combination of two or more such representations? As we will see later, how minds represent music has repercussions for what listeners remember, what listeners judge to be similar, and other musically important functions.

In this chapter, however, we will simply sidestep the issue of mental representation for music—leaving a more thorough discussion for two later chapters. Notice that Saffran's experiments (described in the previous chapter) also sidestepped the issue of how the tone sequences are mentally represented. For example, Saffran's experiments did not directly address whether her listeners learned fixed pitch sequences (such as absolute pitches) or relative pitch sequences (such as intervals or scale-degrees). In principle, these questions can be answered through further experimentation.

As noted, in this chapter we will ignore issues of representation and simply focus on the task of identifying some basic statistical regularities evident in real music. Our sample will be biased primarily toward Western art and folk melodies. But we will draw on music from a number of cultures, including Albanian, American, Bulgarian, Chinese, English, German, Hassidic, Iberian, Irish, Japanese, Macedonian, Norwegian, Ojibway, Pondo, Venda, Xhosa, and Zulu. There are many statistical regularities to be found in music, so we must limit our discussion to a handful. A useful starting point is to identify first those regularities that are the most pervasive or obvious. In this chapter, we identify five such regularities related to melodic organization: *pitch proximity, step declination, step inertia, melodic regression,* and *melodic arch.*

1 Pitch Proximity

One of the best generalizations one can make about melodies is that they typically employ sequences of tones that are close to one another in pitch. This tendency to use small intervals has been observed over the decades by many researchers.[3] Figure 5.1 shows a distribution of interval sizes for samples of music from ten cultures spanning Africa, America, Asia, and Europe.[4] For this sample of cultures, small intervals tend to predominate. There are exceptions to this general trend—such as Swiss yodeling and Scandinavian "yoiks." But overall, there is a marked tendency to construct melodies consisting mostly of small pitch movements.

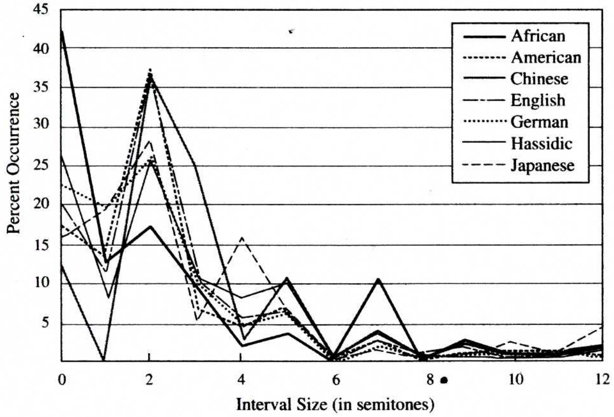

Figure 5.1
Frequency of occurrence of melodic intervals in notated sources for folk and popular melodies from ten cultures. African sample includes Pondo, Venda, Xhosa, and Zulu works. Note that interval sizes only roughly correspond to equally tempered semitones.

If real melodies tend to favor small intervals, what about listeners' expectations? Do listeners *expect* successive melodic intervals to be small? Since accurate expectation promotes event readiness, evidence consistent with an expectation for pitch proximity would show that listeners process small intervals more quickly than large intervals.

In 1978, Diana Deutsch at the University of California, San Diego, showed that listeners are more efficient when processing tones preceded by small intervals than by large intervals. A year later, Paul Boomsliter and Warren Creel at the University of Toronto found that when exposed to extremely brief tones, listeners are faster to form pitch sensations when the stimuli are embedded in sequences where successive pitches are close together.[5] By contrast, listeners take longer to form pitch sensations when the pitch distance separating successive tones is large.

In a musical context, evidence of listeners' expectations for proximate pitch continuations has been assembled from experiments carried out in my Ohio State University laboratory by Bret Aarden.[6] Using a reaction-time method, Aarden asked participants to listen to a number of folksong melodies. As each note in the melody was played, listeners had to indicate whether the pitch was higher, lower, or the same as the previous pitch. Aarden found that listeners respond significantly faster when the successive tones are closer in pitch.

None of the above experiments proves that listeners expect small intervals. It is theoretically possible that these observations have some other explanation. But given the facts that melodies tend to use mostly small intervals and that the auditory system appears to be sensitive to frequently occurring phenomena, it is not unreasonable to suppose that listeners might have learned to expect small intervals. At a minimum, we can conclude that small pitch intervals are a common feature of real music, and that listeners familiar with Western music process small intervals more easily than large intervals.

2 Step Declination

Working at the University of Nijmegen in the Netherlands, Piet Vos and Jim Troost discovered that large melodic intervals are more likely to ascend and that small melodic intervals are more likely to descend.[7] Figure 5.2 shows the frequency of occurrence of ascending intervals for different interval sizes. The dark bars show the results for Western classical music whereas the light bars show the results for mainly Western folk music. Fewer than 50 percent of small intervals ascend. The reverse holds for large intervals.