THE INFLUENCE OF EXPECTANCY ON MELODIC PERCEPTION

Anna M. Unyk
James C. Carlsen
University of Washington

The fulfillment and violation of melodic expectancies influences musicians' ability to perceive, identify, and recall melodic patterns as measured by transcription accuracy. Twenty-seven musicians registered their melodic continuation expectancies by singing. Those expectancies were used to generate six types of brief melodies that varied in their relationships to the individual musician's expectancies: fulfillment of strong expectancies, fulfillment of weak expectancies, interval-size violation of strong or weak expectancies, and contour violation of strong or weak expectancies. The test melodies were presented aurally for transcription. Analysis of variance revealed that violations of strong expectancies led to more errors than expectancy fulfillment. Contour violations did not lead to more errors than mere interval-size violations. Analysis of the pattern of errors suggests that the salience of contour in melody strongly resists the influence of expectancy upon perception, but does not completely overcome it.

The process of expectation or anticipation of future events in listening to music is the focus of several theories of music perception. Meyer (1956, 1967) contends that each music style has characteristic patterns of pitch, rhythm, harmony, etc. that occur more often in that style than in others. Through experience with styles, listeners internalize these probabilities of occurrence which become the basis of their expectancies for future events while listening to unfolding music patterns. Expectancies are also influenced by Gestalt laws. These laws dictate that perceptual processes strive towards completion, good continuation, regularity and symmetry, and expectancies reflect these strivings. If these expectancies are violated in music listening, inaccuracies occur in the memory for these events. According to Meyer, encountering unexpected events contributes to a listener's emotional response to music.

Jones (1981, 1982) suggests that expectancies in music listening are determined by the "ideal prototypes" of a music style. These prototypes reflect the perfect symmetries which underlie, but rarely appear, in a particular style; expectancies, therefore, are often violated in music. While unexpected events create interest in music, they are also thought to be more difficult to recall.

Krumhansl, Bharucha, and Castellano (1982) propose that music patterns are not random events, but conform to patterns that are characteristic of a music culture. Through experience with these patterns, listeners abstract and internalize the underlying regularities in them. These internalizations are called cognitive representations of music, and give rise to expectations as to what is likely to follow in an unfolding music pattern. Krumhansl (1979) suggests that the structure of these internalized regularities also influences the stability of memory traces.

Several empirical studies in the psychological and psychomusical literature lend support to the hypothesized existence of prototypes, cognitive representations,

Psychomusicology, 1987
Volume 7, Number 1

and expectancies. Evidence is also provided for the influence of expectancy on perception, identification, and recall of events.

Rosch (1975a, 1975b) conducted several experiments which suggest that the organization of categorical thinking revolves around prototypes or most representative examples. She asked individuals to rank objects in terms of how representative or prototypical the objects were of a particular category. Rosch found that subjects were able to rank objects belonging to a category differentially along the dimension of prototypicality and that there was a high intersubject agreement on these rankings. She also found asymmetries in similarity judgments of these category members. Less prototypical members were judged to be more similar to more prototypical members than vice versa. Rosch considered this to be evidence that more prototypical examples of a category act as reference points for the category.

Krumhansl (1979) studied cognitive representations of pitch using the method of similarity ratings. She asked subjects to rate the similarity of two tones within the context of a major scale. She found that nondiatonic tones were judged more similar to diatonic tones than vice versa, and diatonic tones were judged more similar to tonic triad tones than vice versa. These results were interpreted to suggest that the organization of cognitive representations of pitch in scale contexts revolve around the central focal point of the tonic triad.

Krumhansl also studied the influence of these cognitive representations upon perception. A standard and comparison tone were presented to subjects with a sequence of interpolated tones. Subjects were to judge if the comparison was the same as the standard. The standard and comparison were either diatonic or nondiatonic within the tonality implied by the interpolated tones. Nondiatonic tones were more confused with diatonic tones than vice versa, reflecting the structure and influence of cognitive representations.

Some evidence for Jones' theory of ideal prototypes was provided by Jones, Maser, and Kidd (1978). Musicians' recall of distance vs. nondistance nested patterns matched for contour were compared. The sequence "c^1 e^1, $g\#^1$ e^1, e^2 $g\#^2$, c^3 $g\#^2$" is an example of a distance-nested pattern. It is defined as such because the distances between the two tones in the paired groupings of adjacent tones are equal to or smaller than the distances between the outer tones of groupings of four adjacent tones or all eight tones. Distance-nested patterns are, therefore, hierarchically organized in terms of distance, and more structurally symmetrical than some other patterns. It was found that the greater the nesting, the better the recall of the pattern.

Additional evidence for the influence of expectancy upon perception was provided by Bruner and Postman (1949). Images of playing cards were flashed on a screen with a tachistoscope. For some of the cards, color and suit were switched so that a card with black hearts was possible. Subjects required longer presentation times to identify these incongruous (unexpected) cards than the normal ones.

In another experiment, Postman, Bruner, and Walk (1951) found that the violation of strong expectancies affected perceptual accuracy to a greater extent than violation of weak expectancies. They presented nonsense and meaningful words to subjects on a tachistoscope. Both of these types of words had a letter reversed. The meaningful words were thought to create a context in which normal, unreversed letters were strongly expected. For the nonsensical words, this expectancy was weaker. *Subjects required a longer presentation time to detect the* reversed letter in the meaningful word than in the nonsensical word. In terms of identification of unreversed letters, however, the letters in the meaningful word were more quickly identified than the letters in the nonsensical word. This suggests that strong fulfilled expectancies facilitate accurate perception.

Given a situation in which unexpected events were not identified correctly, Bruner and Postman (1949) observed three reactions. Subjects either identified the event as being the expected one, identified the event as something in between what was expected and what actually took place, or were unable to label the event at all. Also, an unexpected event in one aspect of a stimulus sometimes caused a type of distortion of perception in the rest of the stimulus that made the event more expected within the context (Postman, et al., 1951).

None of the empirical studies cited thus far examines the nature of expectancies directly. Carlsen (1981) conducted the first such study in music. Twenty-five two-tone melodic beginnings representing all the ascending and descending intervals within the octave and the interval of the unison were presented to subjects. They were instructed to sing what they expected the rest of the melody to be had it not been interrupted. The subject sample was drawn from musicians in three countries, two levels of music training, and four voice ranges.

Carlsen found that different expectancies were sung for the various melodic beginnings. As well, some beginnings produced more of a consensus within the group on a particular continuation than others. Carlsen also found that the sung continuations were different for subjects from the three countries. This result lends support to the theory that music expectancies are dependent on previous experiences with music.

Upon comparing the expectancies of the two training-level groups and four voice ranges, however, no significant differences were found. This suggests that these groups may not have had widely differing music experiences. Another interpretation is that expectancies are developed relatively early in life. Because the perception, identification, and recall of events have been hypothesized to be influenced by expectancies, a cycle of reinforcement of expectancies in perception may be promoted, creating a situation in which expectancies are very difficult to change.

Carlsen (1976) also studied the influence of expectancy fulfillment and violation upon errors made in the transcription of aurally-presented melodies into the form of notation. Based on the theory of Bruner and Postman (1949), he compared errors made in melodies containing *fulfilled and violated expectancies* for weak and strong expectancy generators. The choices of tones for these melodies were based on the expectancy data collected on an earlier sample of subjects. Melodies were presented to subjects which began with either strong or weak expectancy generating melodic beginnings. Strong expectancy generators were defined as those melodic beginnings found to create the most consensus on one expectancy response in the earlier sample. Weak expectancy generators were defined as those beginnings which created less consensus. These beginnings were then followed by intervals which were either produced by the greatest percentage of subjects in the earlier sample (fulfilled expectancy), or by the least percentage of subjects (violated expectancy). The remaining tones of the melodies were chosen to conform to a music theme in the literature. Errors in the notation of dictation melodies containing four types of expectancies were compared: strong fulfilled, weak fulfilled, weak violated, and strong violated. The Bruner and Postman theory would predict inaccuracies in the transcription of these melodies in this same order, with the least errors occurring in the strong fulfilled expectancy melodies and the most errors occurring in the strong violated expectancy melodies. This is the result that Carlsen obtained.

The purpose of this study was to further test the relationship between expectancy and perception within the context of a melodic dictation task. The transcription of aurally-presented melodies into notation is a skill that is emphasized in the training of most musicians. Despite this training, students continue to have problems in notating some melodies, even though they have mastered most other dictation exercises. Theories of music expectancy would suggest that some melodies might be more difficult to perceive, identify, recall and, as a result, notate because they violate expectancies. This hypothesis was tested using an amplification of Carlsen's (1976) experimental design. Data were first gathered on the melodic expectancies of subjects, and then short melodies based on these expectancies were presented to these same subjects in the form of melodic dictation exercises. The melodies either fulfilled or violated strong or weak expectancies. However, several aspects of the experiment were more rigorously controlled in comparison to Carlsen's earlier design. Carlsen defined his strong and weak expectancy generator strengths by the amount of consensus the subject sample displayed on a particular response. In this experiment, the definitions of strong and weak expectancy generators as well as expectancy fulfillment and violation were arrived at through inspection of the expectancy responses of each individual. A fulfillment of expectancy for one person can be a violation of expectancy for another because of differences in past experience with music. These individual differences were taken into consideration in this design.

Carlsen's study (1976) examined the effects of violation of expectancy for interval size only. The difference between the expectancy fulfillment and violation conditions was the size of the expectancy response interval, but not its direction. In this study, the differential effects of violation of interval size as well as contour were compared in relation to transcription errors.

Based on the results obtained by Carlsen (1976) and Postman, et al. (1951), it was hypothesized that more errors would be made in melodies containing violated in comparison to fulfilled expectancies, especially in the case of strong expectancies. It was also hypothesized that a greater degree of expectancy violation would lead to a greater number of transcription errors. Therefore, it was predicted that the greatest amount of errors would be made in melodies with violated expectancies for both interval size and contour, followed by melodies with violated expectancies for interval size but not contour. The least errors were predicted for melodies with fulfilled expectancies for both interval size and contour.

Methodology

Experimental Design

The experiment consisted of two parts. In part one, melodic expectancy data were gathered for each subject. Based on each individual's data, a set of melodic dictation exercises was created for that subject which tested the influence of two levels of the factor of expectancy strength (strong/weak) and three levels of the factor of expectancy fulfillment [fulfilled; unfulfilled (violated) interval size, but fulfilled contour; and unfulfilled (violated) interval size and contour]. In the second part of the experiment, subjects took these dictation exercises, and the amount of errors made in melodies representing the two levels of the expectancy strength factor and the three levels of the expectancy fulfillment factor were compared.

Subjects

Twenty-seven volunteers participated in the study. All were required to have completed at least a one-year college level ear training course. Only persons who could pass a test (Carlsen, 1981) indicating their ability to sing accurately a mentally conceived melody were used in the study. Most of the participants were undergraduate and graduate music majors at the University of Washington. Some faculty and students from Seattle Pacific University and Northwest College also participated. Subjects were paid $3.50 upon completing two experimental sessions.

Expectancy Data

Subjects listened to stimulus tapes of 125 two-tone melodic beginnings and were asked to sing what they expected the continuation of each of these melodies to be had it not been interrupted. The stimulus tapes were a subset of those used by Carlsen (1981). The 125 melodic beginnings consisted of a set of 25 different melodic intervals presented five times. In each presentation, the order of the intervals was different. The intervals were all those possible within the 12-tone equal-tempered octave: 12 ascending intervals, 12 descending intervals, and the unison. The pitches of the 25 intervals were different in at least three out of five presentations of each interval.

Each melodic beginning interval was presented five times to increase the likelihood that the responses obtained accurately reflected the melodic expectancies of subjects. If a melodic beginning inspired several equiprobable expected continuations for a subject, one response would not reflect all of these continuations. A distribution of five responses, on the other hand, would be a much more sensitive measure in this case.

Two different orders of the 125 stimuli were used in the experiment, with the subjects randomly assigned to either order. These two orders of 125 melodic beginnings were presented at four different pitch levels to correspond to the singing ranges of sopranos, altos, tenors, and basses. To insure the ability of the subject to sing expected continuations whether ascending or descending, the second pitch of each melodic beginning in these four pitch levels occurred in the mid-register of the voice range.

The beginnings were produced on a Buchla Dual Square Wave Generator (Model 144) and consisted of a square wave modified by a Buchla Equalizer Line Driver (Model 175) which rolled off the upper register slightly. Envelope was controlled by a Buchla Quad Voltage controlled Envelope Generator (Model 284). Settings on this envelope generator are not finely calibrated, but the envelope had an approximate 30 ms rise time, 500 ms steady state, and 75 ms decay. The tempo of each melodic beginning was one pitch per second. Metronome pulses introduced each melodic beginning. Two pulses sounded before each melodic beginning and continued for nine more pulses in which time the subject sang the continuation. The tempo of these pulses was one per second, and the two-tone melodic beginnings occurred simultaneously with these pulses. A two- to three- second silence occurred after these 11 pulses before another melodic beginning was introduced. This pattern of pulses and stimuli is shown in Figure 1.

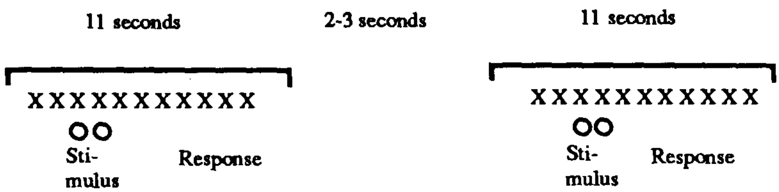


Figure 1. Stimulus and response configuration for melodic expectancy data collection: (x) denates pulse, (o) denotes melodic beginning.

The responses to the 125 melodic beginnings were coded in terms of the size of the melodic interval between the second pitch of the melodic beginning and the first sung pitch. As Carlsen (1981) points out, only this one interval of the response should be analyzed as an expectancy response because subsequent sung pitches may not be a function of an expectancy based on the two-tone melodic beginnings, but rather a function of previously sung pitches. Since each of the 25 melodic beginning intervals occurred five times, a profile resulted for each individual which contained the distribution of the five responses within 25 response categories (all possible ascending and descending intervals within the octave plus the unison) for each melodic beginning.

Construction of Melodic Dictation Exercises

A unique set of 24 melodic dictation exercises was constructed for each subject based on that subject's expectancy response profile from the first part of the experiment. Four melodies represented each of the three levels of expectancy fulfillment within each of the two levels of expectancy strength, creating a total of six experimental conditions.

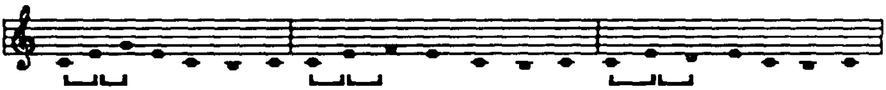
In order to construct these melodies, four melodic beginnings were randomly chosen from each subject's profile of beginnings which proved to be strong expectancy generators, and four which proved to be weak expectancy generators. Strong expectancy generators were defined as those melodic beginnings to which individuals sang the same response interval at least four out of a possible five times within their individual expectancy profiles. Weak expectancy generators were defined as those beginnings to which subjects sang the same response no more than two out of five times. Each of these eight expectancy generators (four strong and four weak) served as the beginning of three melodic dictation exercises representing the three levels of expectancy fulfillment. The three melodies were identical in intervallic structure except for the difference created by an alteration of the third tone of the melody to correspond to the three different levels of expectancy fulfillment.

The three levels of expectancy fulfillment were created as follows. For each of the four strong melodic beginnings, a response interval was chosen from each individual's profile which represented a fulfilled expectancy. The interval chosen was the one that occurred at least four out of five times. Since this interval also had a direction, ascending or descending, this interval represented both a fulfilled interval-size expectancy and a fulfilled contour expectancy for that melodic beginning. Each of these four three-tone patterns (beginning plus response) served as the beginning of a melody consisting of six or seven tones. The melodies represented the fulfilled expectancy condition for strong expectancy generators. The melodies were selected from themes in the literature listed in Barlow and Morgenstern's *A Dictionary of Musical Themes* (1948). The first few tones of these themes are listed in alphabetical order in this source, making it relatively easy to find melodies starting with the three tones based on expectancy fulfillment. Continuations for these melodies for up to six or seven tones were constructed based on these themes. The duration of the tones in the themes were ignored so that all melodies consisted of tones with equal durations. An attempt was made to choose melodies which established a strong tonality as well as a sense of finality within the span of six or seven tones. Where melodies starting with the three expectancy tones could not be found in Barlow and Morgenstern, melodies were composed for the study. In all, 56% of the melodies were from the Barlow and Morgenstern dictionary, and 44% were composed.

Two more melodies were created for each of the four strong expectancy generator melodies already chosen or composed. These melodies represented the two conditions of unfulfilled expectancy. For one of these two melodies, the interval between the second and third pitch was altered to one that never occurred as a response to the particular melodic beginning preceding it in the individual's expectancy data. This unexpected interval, however, preserved the direction of the strongly expected interval. This melody represented the violated interval size but fulfilled contour condition. The third pitch of the melody, forming the unexpected interval, was chosen from the diatonic set of tones within the tonality implied by the melody. Judgments about the tonality of the melody were based on Barlow and Morgenstern's (1948) analysis of the melodies chosen from their dictionary. For the melodies composed for the study, a subjective judgment of the tonality was used as a criterion.

The third melody constructed for each of the four strong expectancy generator melodies represented a condition of violation of expectancy in both interval size and contour for the interval between the second and third tones of the melody. This interval was altered to one that was never produced as an expectancy response by the individual after hearing the two-tone melodic beginning. Its direction (ascending or descending) was in opposition to the direction of the interval representing the fulfilled expectancy condition.

Figure 2 shows an example of three melodies representing each of the three expectancy fulfillment conditions for one subject.



(a) +4 +3 (b)+4 +2 (c) +4 -2

Figure 2. Melodic dictation exercises: 5(a) represents fulfilled expectancy condition; 5(b) represents unfulfilled interval size and fulfilled contour condition; 5(c) represents both unfulfilled interval size and contour condition.

For each of the four melodic beginnings chosen to represent weak melodic expectancies, three melodies were also constructed to represent the three expectancy fulfillment conditions using similar principles. However, some differences occurred in the choice of the expected interval size and contour. Weak expectancy generators were defined as those melodic beginnings for which a subject sang one particular expectancy response interval no more than two times out of five. For the melodies representing the fulfilled expectancy condition, the interval between the second and third pitch of the melody was chosen from those response intervals produced at least once by the subject. However, a further restriction applied. Since sung intervals produced at least once could have occurred in both ascending and descending contours, a problem arose in defining an expected direction for this fulfilled expectancy condition. The expected contour in this case was defined as the one in which the majority (three out of five) of expectancy responses occurred. An interval produced at least once from this set of expected contour intervals was then chosen to represent the fulfilled expectancy conditions. The remaining rules for construction of the other expectancy violation conditions were the same as for the strong expectancy generators.

The construction of melodies was guided by several other rules in addition to expectancy strength and expectancy fulfillment. Approximately half of the melodies within each expectancy strength level began on the tonic note of the implied tonality of the melodies. The others did not. This rule was introduced so that subjects would not develop an overall expectancy that all melodies would begin on the tonic. This might have restricted their range of expectancies in a manner different from the sung expectancy task, thereby diminishing the predictive power of the sung expectancy profiles.

The number of accidentals needed to notate each melody was also systematically varied. For approximately half of the melodies in each expectancy generator strength condition, two or more accidentals were necessary for notation. In the other melodies, one or no accidentals were necessary for accurate notation. This rule was introduced so that a general expectancy of "no accidentals" or "many accidentals" was not developed by subjects during the task which might have changed their expectancies in some major way from those in the sung task.

Each of the melodies presenting the three different levels of expectancy fulfillment for a single melodic beginning began on a different pitch. This rule was used to counteract the possibility of recognition by subjects that the same melody was presented three times in a slightly altered form. Only one exception to this rule occured. For one subject, two melodies began on the same pitch.

The presentation of the 24 melodies was ordered such that melodies in one of the six conditions were not always presented first. This further counteracted the possibility of recognition as a confounding variable in the study.

The 24 dictation melodies were recorded on a cassette tape for each individual. The tones were electronically produced on a New England Digital Able-60 computer. The tones had a rise time of 10 ms, a steady state of 490 ms, and a 50 ms decay time. The time of the tones was one per second. These tones were subjectively matched in timbre and tempo to the melodic beginnings used in gathering the expectancy data. Each melody was preceded by two pulses and a five second silence to warn the subject to attend to the upcoming melody. After each melody, a 30 s silence occurred before the warning pulses for the next melody. The subject was to notate the melody in this span of time. A practice melody which was not scored was included at the beginning of each tape, creating a total of 25 melodic dictation exercises for each individual.

Procedure

Subjects were tested individually in a sound attenuated booth in the Systematic Musicology Laboratory at the University of Washington. The procedure consisted of two experimental sessions, the first lasting 40 minutes, the second 20 minutes.

In session one, subjects sat at a table with a microphone. They were told that they were going to hear a series of melodic beginnings through headphones. They were instructed that each melody would be interrupted after the second tone, and that they were to sing what they expected the continuation of that melody to be if it had not been interrupted. The 125 melodic beginnings and sung responses were recorded on another cassette tape for purposes of data analysis.

In session two of the experiment, subjects sat at a table and listened to the 25 melodic dictation exercises through headphones. They were asked to notate these melodies using accidentals where necessary. Music manuscript paper was provided, on which the first note of each melody to be transcribed was given. A key signature for the melodies was not given. Subjects heard each melody once, and had 30 s to notate it. A complete listing of the melodies presented to each subject as well as the experimental instructions can be found in Unyk (1985).

Scoring

In the first part of the experiment, only the interval between the second melodic-beginning tone and the first sung pitch was scored in order to avoid schema effects growing out of the sung response (Carlsen, 1981). In the second part of the experiment, the number of errors made in the second, third and fourth tone of each of the 24 melodic dictation exercises were analyzed. Even though only the third tone of the melody was based on fulfillment or violation of expectancy, the two tones around it were scored because subjects may have found the third tone to be incongruous, and subsequently varied their perception or memory of the tones around it in order to make the pattern more "expected" (Postman, et al., 1951). In this case, the third tone might be notated correctly, but the tones surrounding it would be inaccurately transcribed. Therefore, all three tones were included in the analysis.

The number of errors in the four melodies in each of the six experimental conditions were added together for the purposes of data analysis. Since the number of errors possible in each melody was three, and there were four melodies in each condition, each subject received an error score out of 12 for each of the six experimental conditions.

Results

A frequency count was conducted of the 25 possible responses to each melodic beginning in the first part of the experiment. Table 1 shows the percentage of responses falling in each interval category of response based on a total response set of subjects and repetitions (27 x 5) for each beginning. The left column denotes the size of each melodic beginning in terms of semitones, with 0 denoting a unison and 12 denoting the octave. Positive numbers signify ascending intervals; negative numbers, descending intervals. The response intervals in semitones between the second pitch of the melodic beginning and the first sung pitch occurring at least as often as chance (one out of 25) are shown in the remaining columns. They are arranged in rank order of percentage of response.

Most of the responses (62.5%) occurring across melodic beginnings were conjunct intervals, i.e., unisons, semitones and whole tones (0, 1, -1, 2, -2). However, each melodic beginning produced a distinct pattern of responses. These patterns showed a considerable similarity to those expectancies obtained by Carlsen (1981) for identical stimuli. For 12 out of the 25 melodic beginnings, the most frequently produced expectancy response was the same in both studies. For an additional seven melodic beginnings, the response most frequently produced in one study was the second most frequently produced in the other study. These results demonstrate that melodic beginnings as short as two tones can produce quite structured expectancies for subsequent tones in a music context.

Some melodic beginnings, such as the -2, -1, 1, 2, 4, 10, and 11 were found to produce a stronger group consensus on one response interval, while other beginnings such as the -12, -10, -9, -8, -3, 7, 8, 9, and 12 produced a weak group consensus. A low consensus on a particular response by the group, however, does not imply that

Table 1

Expectancy response intervals occuring at least as often as chance

Mel Beg[a]	Rank order of response intervals and percentage of responses for each melodic beginning								
	1	2	3	4	5	6	7	8	9
-12	5 (24)	2 (14)	0 (13.3)	7 (11.9)	12[b] (11.1)	4 (6.7)	-3 (4.4)		
-11	11[b] (26)	-1[c] (16.3)	1 (11.1)	0 (8.9)	2 (7.4)	5 (6)	4 (4.4)		
-10	1 (21.5)	10[b] (15)	5 (13)	2 (8.2)	-1 (7.4)	0 (6.7)	-2[c] (5.2)	4 (5.2)	7(4.4)
-9	9[b] (21.5)	5 (17)	-1 (14.8)	2 (13)	1 (8.2)	0 (8.2)	4 (4.4)		
-8	1 (22.2)	8[b] (17.8)	5 (11.1)	2 (10.4)	3 (9.6)	0 (8.2)	-4[c] (6)	-1 (4.4)	
-7	7[b] (27.4)	5 (17.8)	2 (15.6)	0 (7.4)	-1 (7.4)	4 (6)			
-6	1 (26.7)	6[b] (17.8)	-1 (11.1)	2 (9.7)	0 (8.9)	3 (6)			
-5	5[b] (31.6)	2 (23.7)	1 (6.7)	3 (6)	0 (6)	-7[c] (6)	-1 (5.2)	-3 (4.4)	
-4	2 (35.6)	4[b] (24.4)	-1 (11.1)	-3 (8.2)	0 (6.7)				
-3	1 (23)	-2 (21.5)	3[b] (19.3)	-4 (9.6)	2 (8.9)	0 (6.7)	-1 (4.4)		
-2	-2 (47.8)	2[b] (19.4)	-1 (17.9)						
-1	-2 (39.6)	1[b] (26.1)	-1 (20.9)						
0	0 (35.6)	2 (30.4)	-2 (7.4)	4 (6.7)					
1	2 (42.1)	1 (31.6)	-1[b] (12)						
2	2 (61.5)	1 (14)	0 (6.7)	-2[b] (5.2)	3 (4.4)				
3	4 (27.4)	2 (24.4)	-3[b] (12.6)	-1 (11.1)	1 (8.1)	0 (5.9)	5 (5.2)		
4	3 (49.6)	1 (20)	0 (6.7)	-4[b] (6.7)	-2 (11.1)	-1 (4.4)			
5	2 (32)	4 (20.9)	0 (13.4)	-5[b] (12.7)	-1 (9)	7[c] (6)			
6	1 (34.1)	0 (11.1)	-2 (10.4)	2 (8.9)	-6[b] (8.1)	-1 (7.4)	5 (5.9)	3 (5.2)	
7	2 (17.8)	-7[b] (16.3)	5[c] (9.6)	1 (9.6)	0 (9.6)	-3 (9.6)	-1 (8.1)	-2 (8.1)	-4(4.4)
8	-1 (21.5)	2 (14.8)	0 (14.1)	1 (12.6)	-8[b] (11.9)	-3 (6.7)	-2 (5.0)		
9	-2 (21.5)	1 (20)	-4 (14.8)	0 (11.1)	-9[b] (11.1)	3[c] (7.4)	-1 (6.7)		
10	-1 (42.5)	2[c] (14.2)	-2 (13.4)	-10[b] (11.1)	0 (7.4)	-4 (7.4)	-11 (6)		
11	1[c] (41.5)	-2 (11.9)	-1 (8.9)	0 (7.4)	-4 (7.4)	-11[b] (6)			
12	-1 (21.5)	0 (21.5)	-12[b] (17)	-2 (13.3)	2 (7.4)	-3 (6)	1 (4.4)		

[a]The number of the expectancy response interval refers to the numbers of semitones in the interval. Negative numbers refer to descending intervals.

[b]Indicates a return to the first pitch of the melodic beginning.

[c]Indicates a completion of the octave.

individuals, in their own profiles of five repetitions for each melodic beginning interval, produced a low consensus on any one response. A frequency count was conducted to determine the degree of intrasubject consensus on responses for each melodic beginning. Based on this analysis, the 25 melodic beginnings in the individual response profiles were divided into three groups: strong, medium, and weak expectancy generators. Strong beginnings were defined as those for which four or five out of the five repetitions produced the same response interval in a subject. Weak beginnings were those which produced two or less of the same response intervals. The remaining melodic beginnings were denoted as medium expectancy generators. Table 2 shows the percentage of subjects for which each melodic beginning was scored as strong, medium or weak.

From this analysis, it can be seen that no melodic beginnings produced exclusively strong or weak within-subject consensus on a particular response interval in the five repetitions of each melodic beginning. Even in beginnings which produced a high group consensus on a particular response, such as the ascending major second (2), some subjects (11.1%) produced a weak consensus (2 or less out of 5) on a particular continuation in their individual profiles. Conversely, a melodic beginning such as the ascending perfect fifth (7), which produced a relatively low consensus on a particular response in the group, produced a strong consensus on one response for 40.7% of subjects. These results indicate that a beginning might elicit a rather unambiguous expectancy for a particular continuation in a particular subject, but this expectation may be different from one elicited in another subject. This poses an interesting problem for measuring the relationship between expectancy and perceptual accuracy using any predetermined "expected" patterns based on group consensus (Carlsen, 1976) or other structural principles (Jones, 1981, 1982). If expectancy is defined as the anticipation of future events in an unfolding music pattern, and the content of these expectancies are found to differ across subjects, then a measure of the effects of violation of expectancy on perception must also reflect these individual differences. For this reason, the definitions of strong and weak expectancy generators, as well as definitions of fulfillment and violation of expectancies used in this experiment were based on individual profiles.

This research method produced a unique set of melodic dictation exercises for each subject. Table 3 shows the frequency of various patterns of melodic beginnings followed by expectancy intervals used to represent two experimental conditions combined, that of fulfilled expectancy for strong and weak expectancy generating melodic beginnings

The frequencies in this table are to be interpreted in comparison to a total sample frequency of 27 (subjects) x 8 (melodies) = 216. The distribution shows a substantial amount of variety in the melodic patterns used to represent fulfilled expectancies. The most frequently occurring pattern of melodies was the melodic beginning of the -1 followed by -2, occurring in seven out of 216 melodic dictation exercises. The same combination of melodic beginning and response, however, did not necessarily result in identical melodies across subjects because different

Table 2

Percentage of subjects for which melodic beginning was scored as strong, medium, and weak.

Melodic Beginning	% Subjects Strong	% Subjects Medium	% Subjects Weak
-12	14.8	33.3	51.9
-11	18.5	25.9	55.6
-10	22.2	18.5	59.3
-9	11.1	40.7	48.2
-8	29.6	33.3	37.0
-7	22.2	29.6	48.2
-6	37.0	33.3	29.6
-5	40.7	37.0	22.2
-4	51.9	29.6	18.5
-3	37.0	33.3	29.6
-2	44.4	22.2	33.3
-1	63.0	18.5	18.5
0	48.2	25.9	25.9
1	40.7	40.7	18.5
2	51.9	37.0	11.1
3	51.9	25.9	22.2
4	66.7	14.8	18.5
5	44.4	33.3	22.2
6	18.5	29.6	51.9
7	40.7	44.4	14.8
8	33.3	33.3	33.3
9	22.2	44.4	33.3
10	29.6	37.0	33.3
11	40.7	25.9	33.3
12	25.9	29.6	44.4

continuations for these tones were used to produce the rest of the melody up to a total of six or seven tones. Only eight pairs of individuals and two groups of three individuals had one identical melody in their entire set of dictation exercises. This ensured that the attributes of particular melodies did not have a confounding influence on the measured relationship between expectation and errors.

Analysis of Errors in Dictation Exercises

Each subject notated four melodies within each of six experimental conditions. These six conditions formed a 3 x 2 factorial design with three levels of expectancy fulfillment crossed with two levels of expectancy generator strength. Three errors were possible for each melody, and these errors were added together for the four melodies in each condition, creating a total of 12 possible errors in each of the six conditions per subject. These totals were submitted to a 3 x 2 analysis of variance for repeated measures, which tested for main effects across the factors of expectancy generator strength and expectancy fulfillment as well as for their interaction.

Table 3
Frequency of responses chosen to represent fulfilled expectancies for each melodic beginning.

	Rank order of interval representing fulfilled expectancy and its frequency (in parentheses)						
Melodic Beginning	1	2	3	4	5	6	7
-12	12 (2)	-3 (1)	5 (1)				
-11	2 (2)	-1 (2)	8 (1)	-5 (1)	1 (1)	7 (1)	11 (1)
-10	5 (4)	1 (3)	-1 (2)	3 (1)	10 (1)	4 (1)	
-9	5 (3)	-2 (2)	-1 (2)	2 (1)			
-8	1 (3)	8 (2)	5 (2)	2 (1)	3 (1)	7 (1)	-4 (1)
-7	5 (2)	2 (2)	7 (1)	-1 (1)	4 (1)	1 (1)	
-6	1 (4)	-1 (1)	3 (10	2 (1)			
-5	5 (2)	2 (2)	1 (1)	-2 (1)			
-4	4 (3)	2 (4)	5 (1)	-1 (1)	-2 (1)		
-3	1 (3)	-4 (2)	-1 (1)	3 (1)	-2 (1)	5 (1)	
-2	-2 (4)	-1 (3)	2 (2)	-3 (2)			
-1	-2 (7)	-1 (4)	1 (2)	-4 (1)			
0	2 (2)	4 (1)	-2 (1)	5 (1)			
1	2 (4)	1 (3)	4 (1)	3 (1)	-1 (1)		
2	2 (5)	1 (1)					
3	2 (5)	4 (3)	5 (1)	1 (1)			
4	3 (6)	1 (3)	-2 (2)				
5	2 (5)	4 (2)	-1 (2)	-2 (2)	7 (1)	-5 (1)	
6	1 (4)	2 (1)	-2 (1)	3 (1)			
7	2 (3)	5 (2)	-7 (1)	-2 (1)	-3 (1)		
8	-1 (4)	1 (1)	2 (1)	-2 (1)			
9	-2 (3)	-4 (2)	1 (2)	-1 (2)	3 (1)		
10	-1 (5)	-2 (1)	-10 (1)	2 (1)			
11	1 (3)	5 (1)	-1 (1)	-12 (1)	-2 (1)	-4 (1)	
12	-2 (2)	2 (1)	-7 (1)				

Both main effects were found to be significant. (expectancy strength, F=15.2 (1,26), $p<.001$, MSe=46.72; fulfillment, F=15.6 (2,52), $p<.001$, MSe=48.2). The number of errors varied significantly with the level of expectancy fulfillment that was created.

Also, the expectancy strength conditions differed significantly in the amount of errors produced. The interaction between levels of expectancy strength and levels of expectancy fulfillment was not significant.

In order to identify sources of critical difference, a multiple comparison of the mean errors in the three expectancy fulfillment conditions was carried out within

Table 4
Results of the Newman-Keuls comparison procedure for mean errors in six experimental conditions.

	Expectancy Fulfillment		
Expectancy Strength	Fulfilled	Unfulfilled Interval	Unfulfilled Interval and Contour
Strong	3.93	*6.00*	*6.22*
Weak	*5.78*	*6.53*	*7.07*

Note: Italics signify subsets of means which do not differ significantly from each other at the .05 level. Comparisons conducted separately within each expectancy strength condition.

each level of the expectancy strength factor using the Newmann-Keuls procedure. Table 4 shows the results of this analysis.

For the melodies beginning with strong expectancy generators, the mean number of errors in the violated interval size but fulfilled contour condition was not significantly different from the condition of violated interval size and violated contour. However, both of these conditions produced significantly more errors at the .05 level than the fulfilled expectancy condition.

For melodies beginning with weak expectancy generators, the mean number of errors in the three expectancy fulfillment conditions did not differ significantly from each other at the .05 level.

A comparison of the mean errors was also carried out for the two expectancy generator strength conditions within each of the three levels of the expectancy fulfillment factor using a *t*-test for means of correlated samples. Table 5 shows the results of this analysis.

In the fulfilled expectancy conditions, significantly more errors were made in melodies with weak expectancy generating melodic beginnings. For the conditions of violated interval size and fulfilled contour, and violated interval size and contour, no differences were found for strong and weak expectancy generating melodic beginnings.

Direction of Errors

A frequency count of the types of errors made by subjects was carried out in order to ascertain if errors in the violated expectancy conditions occurred in the direction of expected tones. The analysis was carried out for two conditions only: strong expectancies which violated interval size but fulfilled contour, and strong expectancies for which both interval size and contour were violated. An analysis of weak expectancy generators was not conducted, since no significant differences were found in the number of errors made in fulfilled as compared to violated expectancies.

The procedure involved analysis of the third tone of the melodies. If an error was made in notating this tone, it was included in the analysis. The first step in the analysis involved a frequency count of the number of errors in which subjects notated the tone they expected rather than the one actually presented. For the violated interval size but fulfilled contour condition, 24% of the errors involved notating the expected tone instead of the one presented. For the violated interval size and contour condition, six percent of errors involved notating the expected tone.

The second step of the analysis involved a comparison of the direction of the notated interval between the second and third tone of the melody to the expected direction. For the violated interval size and fulfilled contour condition, all of the errors in the third tone maintained the expected direction of the interval between the second and third pitch which also matched the actual presented direction. For the violated interval size and contour condition, 13% of the errors conformed to the expected direction between the second and third tones of the melody. These errors contradicted the actual presented direction of the intervals.

Internal Validity

Three structural features of the melodies could have possibly occurred more often in the violated expectancy as compared to fulfilled expectancy conditions, thus being the cause of the difference in errors between these conditions. These features were the number of accidentals that were necessary in notating the three scored tones in the melody, the size in semitones of the two intervals between the second and fourth tone of the melody added together, and the repetition of the third tone of the melody elsewhere in the melody. In order to ascertain the relationship of these three

Table 5

Results of comparisons of mean errors within the expectancy fulfillment factor using the t-test for means of correlated samples.

	Expectancy Fulfillment	
Expectancy Fulfillment	Strong	Weak
Fulfilled	3.93	*5.78*
Unfulfilled Interval	*6.00*	*6.52*
Unfulfilled Interval and Contour	*6.22*	*7.07*

Note: Italics signify means which do not differ significantly from each other at the .05 level (one-tailed). Comparisons conducted separately within each expectancy fulfillment condition.

variables to the amount of errors made in notating the melodies, three Pearson Product-Moment correlations were calculated between the amount of errors made in the second, third, and fourth tones of each of the 24 melodies presented to each subject and the three structural features. The correlation between errors and accidentals was calculated as .19, the correlation between interval sizes and errors as .08, and between repeats (1 designating repeated, 0 designating not) and errors as .06. None of these suggests any important relationship.

Interrater reliabilities were calculated in the scoring of both expectancy response intervals and the number of errors in the melodic dictation excercises. On a sample of 5.5% of total expectancy responses, interrater agreement on interval sizes was found to be 92.8%. On a sample of 12.5% of dictation melodies, the interrater agreement on the number of errors was 97.5%.

Discussion

Meyer (1956, 1967) and Jones (1981, 1982) propose that listening to music involves entertaining expectancies about future events in the unfolding music pattern. The contents of these expectancies reflect ideal types or prototypes of music patterns in each music style. These expectancies are therefore often violated in the music. A review of the literature in this area suggests that more errors are made in the perception, identification, and recall of events which are unexpected as compared to those which are expected.

This experiment tested one aspect of this theory, the relationship of fulfillment and violation of expectancy to errors in transcription of aurally presented melodies into the form of music notation. The performance of this task necessarily involved accurate perception, identification, and recall of music. It was predicted that more errors would be made in transcribing melodies with violated as opposed to fulfilled expectancies, especially in the case of strong expectancy generators. This in fact was the result obtained. For strong expectancy generators, significantly more errors were made in melodies containing violated interval size but fulfilled contour and violated interval size and contour in comparison to melodies with fulfilled expectancies. With weak expectancy generators, however, violation of expectancy did not produce significantly more errors.

These results lend support to the theory that listeners entertain certain expectancies about future events in music while music events are unfolding. If these expectancies are strong, violations of expectancy lead to a diminished ability to identify and recall these music events. These results were obtained in spite of the fact that the melodies used in the experiment were very much simpler and more tonally structured than the vast majority of music compositions. The deviations from expectancy were all diatonic within the tonality of the melody and therefore did not require a reorientation to another diatonic system in order to comprehend their relationship to the other tones. Yet, musicians who had received or were receiving advanced professional training did poorly at identifying these simple deviations from the expected. A generalization suggested by this observation is that in the more complex examples of music, deviations from expectation may be even more elaborate, thus influencing even greater errors in identification.

Violations of weak expectancies did not result in more errors than fulfillment of weak expectancies. This result was predicted since the effect of fulfillment or violation of an event which is not exclusively or strongly expected should not be that different. However, this does not mean that these weakly expected events were more accurately transcribed. In fact, more errors were made in averaging across all three expectancy fulfillment conditions for weak as compared to strong expectancies. This result is in contrast to the one obtained by Carlsen (1976) and Postman, et al. (1951), who found that although weakly expected events which were fulfilled were more difficult to identify than strongly expected and fulfilled events (as in this experiment), just the opposite result occurred when these expectancies were violated. More errors occurred in strong violated expectancies than weak violated expectancies. In this experiment, violations of strong expectancies did not create any more errors than violations of weak expectancies.

This contradiction in results may be explained by differences in the stimuli and/or methodologies used in the three experiments. The findings of this experiment converge more closely with hypotheses that derive from Meyer's (1956) theory of music expectancy. He suggests that both weak and strong expectancies emerge out of the same cognitive framework. Strong expectancies ensue from music listening contexts that resemble the redundant features of a music style. Weak expectancies ensue from music contexts which differ greatly from characteristic patterns or whose elements are undifferentiated. These types of contexts are hypothesized to be more difficult to accurately recall. An implication of this theory is that a melodic beginning which most resembles more typical beginnings in a style should generate strong expectancies. A melodic beginning which deviates from the typical or whose structure is undifferentiated should create weak expectancies. The violation of such weak expectancies may simply continue to differentiate the pattern from more structured ones which produce strong expectancies, thus maintaining the difficulty in recall. Therefore, violated weak expectancies would create the same or more errors than violated strong expectancies. The results of this experiment support this prediction.

Another hypothesis in this experiment was that more errors in transcription would be made in the condition of violated interval size and contour than in the condition of violated interval size and fulfilled contour. This hypothesis was not supported by the results. The violation of interval size and contour did not produce significantly more errors than violation of interval size only in either strong or weak expectancy generators. A conclusion might be drawn that expectation for melodic direction is not important in identification and recall of music. However, an inspection of the direction of errors made in transcribing the melodies suggests another interpretation. In the strong violated interval size but fulfilled contour condition, no errors occurred in notating the direction of the second and third interval of the melody. In the violated interval size and contour condition, only 13% of errors occurred in notating the direction of the second and third interval of the melody. Very few errors in transcription, therefore, involved inaccurately notating contour. These results can be interpreted to suggest that contour is so salient in the perception of music that expectations do not affect accurate transcription of contour.

Bruner and Postman (1949) suggest that violations of expectancy may create perceptual error in the direction of the expected event. An analysis of the type of errors made in the third tone of the melodies revealed that 24% of the errors in the strong violated interval size and fulfilled contour condition were made to the expected tone, and even six percent of the errors in the violated interval size and contour condition were to the expected tone. The lower percentage in the violated contour condition can be explained with the hypothesis of salience of contour in perception. Perceptual error in the direction of the expected tone in this condition would have involved an inaccurate perception of the salient features of contour. In this interpretation, a six percent error towards the expected tone, although low, attests to the important role played by expectancy in perception.

Even in the violated interval size condition, however, error to the expected tone occurred in only 24% of the cases. This percentage is not completely surprising in light of the results found by Bruner and Postman (1949). They observed that perceptual error in the direction of the expected event was only one of three possible perceptual reactions under conditions of violation of expectancy. The other two were a compromise between the expected and the real, and complete disorganization of perception.

Further Research

The study of the relationship of expectancy to error in melodic transcription in this experiment was limited to the intervals between the first three pitches of a melody. Music involves many more variables than this. In future research in this area, emphasis must be placed on expanding the number of music elements studied in order to gradually approach the complexity of music. This can be achieved by examining expectancies for two-tone melodic beginnings with differing rhythms and timbres as well as studying expectancies in relation to music beginnings which are longer than two tones, and which are harmonic in nature.

The manner in which expectancy responses are gathered can also be expanded in order to more closely approach the complexity of music. Sung responses can only contain information on interval and rhythm. However, harmonic and timbral expectancies can also be in operation in music. These could be studied by presenting a series of possible expected continuations to subjects after which they could rate how well the continuation approximated their expectation. For study of harmonic expectancy, expected continuations could be played by subjects on a keyboard.

The dependent measure of transcription of aurally presented music patterns into notation may be easily expanded to accommodate fulfillment and violation of rhythmic expectancy. The accuracy of identification of elements in multi-part music, however, might be better examined by asking a subject to play back an excerpt on a keyboard, or choosing the most accurate pattern from a set of patterns. Even in the case of melodies, a measure of accuracy in identification other than transcription into notation can be used, such as repeating a melody vocally after its aural presentation. An experimental comparison of these various methods, however, should be carried out in order to ascertain the similarity in results yielded by them. An in-depth discussion of strengths and weaknesses of some of these approaches has already been presented in Carlsen, Divenyi & Taylor (1970).

The study of the effects of various types of expectancy violations can also be expanded. In the case of melodies, violations of interval size and contour have been tested. More comparisons can be carried out along the lines of comparison of errors for diatonic vs. nondiatonic violations and conjunct vs. disjunct interval violations. Beyond this level, the effect of various rhythmic, harmonic and timbral violations deserve to be examined.

The theory of expectancy outlined in this experiment includes reference to the role of stylistic prototypes as determinants of expectancies. This aspect of the theory was not directly addressed in this experiment. In order to study this hypothesis, a methodology must be developed to ascertain the structure of prototypes and their relationship to expectancy patterns. Similarity ratings between tones such as those used by Krumhansl (1979) can reveal some aspects of cognitive representations of music such as pitch organization. However, the internalized probability systems described by Meyer (1956) involve more complex, style-specific organizations with an interaction of elements such as melody, rhythm and timbre. At this level, perhaps the methodology developed by Rosch (1975b) to study categorical prototypes is more suitable. Subjective rankings of the degree of prototypicality of excerpts from a music style category could be compared to the frequency of expected continuations produced by subjects for patterns within these excerpts. The most prototypical excerpt from a music style category should be reflected in the most frequently produced continuation.

References

Barlow, H., & Morgenstern, S. (1948). *A Dictionary of Musical Themes.* New York: Crown Publishers.

Bruner, J., & Postman, L. (1949). On the perception of incongruity: A paradigm. *Journal of Personality, 18,* 206-223.

Carlsen, J. (1976). Cross-cultural influences on expectancy in music. A plenary address to the World Congress of the International Society for Research in Music Education. In Egon Kraus (Ed.), *International Music Education Yearbook* III. Mainz: Schott, 61-62.

Carlsen, J. (1981). Some factors which influence melodic expectancy. *Psychomusicology, 2* (1), 12-29.

Carlsen, J., Divenyi, P., & Taylor, J. (1970). A preliminary study of perceptual expectancy in melodic configurations. *Council for Research in Music Education Bulletin, 22,* 4-12.

Jones, M. (1981). Music as a stimulus for psychological motion: Part II. Some determinants of expectancies. *Psychomusicology, 1*(2), 14-31.

Jones, M. (1982). Music as a stimulus for psychological motion: Part II . An expectancy model. *Psychomusicology* , *2*(1), 1-13.

Jones, M., Maser, D., & Kidd, G. (1978). Rate and structure in memory for auditory patterns. *Memory and Cognition, 6*(3), 246-258.

Krumhansl, C. (1979). The psychological representation of musical pitch in a tonal context. *Cognitive Psychology, 11,* 346-374.

Krumhansl, C., Bharucha, J., & Castellano, M. (1982). Key distance effects on perceived harmonic structure in music. *Perception and Psychophysics, 32*(2), 96-108.

Meyer, L. (1956). *Emotion and Meaning in Music.* Chicago: University of Chicago Press.

Meyer, L. (1967). *Music, the Arts, and Ideas.* Chicago: University of Chicago Press.

Postman, L., Bruner, J., & Walk, R. (1951). The perception of error. *British Journal of Psychology, 42*(1&2), 1-10.

Rosch, E. (1975a). Cognitive reference points. *Cognitive Psychology, 7,* 532-547.

Rosch, E. (1975b). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General, 104*(3), 195-233.

Unyk, A. (1985). *The influence of expectancy on melodic perception.* Unpublished doctoral dissertation, University of Washington.

Author Notes

This article was based upon the doctoral research of the first author. We thank W. Jay Dowling and Jack Taylor for helpful suggestions and comments in the preparations of this manuscript.

Requests for reprints should be sent to James Carlsen, School of Music, DN-10, University of Washinghton, Seattle, WA 98195.