Editorial and Business Offices:	**Psychomusicology** **Illinois State University** **College of Fine Arts** **CVA 116** **Normal, IL 61761**
Published by:	**The Office of Research in** **Arts Technology** **Illinois State University**

Psychomusicology **abstracts appear in *The Music Index*, *Psychological Abstracts*, *Music Psychology Index*, *CAIRSS™ Project*, and the *International Repertory of Music Literature (RILM)*.**

Psychomusicology **is produced using an in-house system of: Apple Macintosh microcomputers and LaserWriter IINT printer networked on an AppleShare LAN with software including Microsoft *Works*, Aldus *PageMaker*, and Silicon Beach *SuperPaint*.**

Psychomusicology **logo by James C. Carlsen.**

EXPECTANCY PROFILES GENERATED BY MAJOR SCALES: GROUP DIFFERENCES IN RATINGS AND REACTION TIME

Bradley Frankland
Annabel J. Cohen
Dalhousie University

Sixty-one subjects with a wide range of formal music training were presented with ascending and descending C major scales, each followed by a single probe tone drawn from the 13-note chromatic scale. They were asked to decide whether the fit of the probe to the preceding context was good or poor. For each probe, the proportion of good-fit replies was considered as a measure of the degree to which the context established an expectancy (i.e., primed) for that probe. For each subject, the ratings of all 13 x 2 probes were considered as a profile of the internal representation of expectancies generated by the major scale for that subject. A cluster analysis based upon the individual profiles was used to group and classify subjects. This resulted in groups which could be described in terms of the relative emphasis on tonic triad, diatonic, and proximity relations as previously identified by Krumhansl and Shepard (1979). Correlations of the average rating with the average reaction time for each of the 13 x 2 probe tones for each group suggested a relation between the consistency of the ratings (either always good fit or always poor fit) and the reaction time: reaction times were faster when ratings were consistent and slower when ratings were inconsistent. It is concluded that a dichotomous rating procedure reveals expectancy profiles for the major scale context that are consistent with those described by Krumhansl and Shepard (1979) and that reaction time measures complement these ratings.

In many theories of music perception, listeners are considered to have internalized probabilistic rules (e.g., Krumhansl, 1979; Schmuckler, 1989) or prototypes (e.g., Jones, 1982) that define the occurrence of particular notes or stylistic regularities within a given music idiom. When listening to music, these rules guide the generation of expectations for future events (Unyk & Carlsen, 1987). Subsequent events may conform to or deviate from these expectations and it is likely that the nature of this deviation affects the processing of those events (e.g., accuracy and speed).

Expectancies in music may be based on any perceived regularity or structure. The present article discusses the expectancies associated with tonality. Tonality refers to the establishment of one of the 12 chromatic tones as a reference or tonic. The remaining 11 chromatic notes have a relative importance with respect to this reference tone. A distinction is made between the seven diatonic notes (i.e., the scale tones) and the five nondiatonic notes. The diatonic level is further subdivided into three tonic-triad notes and four nontriad notes. Generally, the triad notes are considered to be the most related to the tonic, while the nondiatonic notes are the least related. Furthermore, within each level, the relative importance of the individual notes may differ.

If such a hierarchy is internalized when listeners establish tonality, as shown by Krumhansl and Kessler (1982), one would predict the generation of expectancies on three levels: tonic-triad tones (conforming), diatonic, nontriad tones (intermediate) and nondiatonic tones (anomalous). As well, one would predict that the accuracy and the rate of processing subsequent events might depend upon the level of expectancy produced by the prior tonal context. That is, triad tones would be processed accurately and quickly while nondiatonic tones would be processed inaccurately and slowly. Events between those two extremes would be processed with intermediate accuracy and speed.

Such a concept was tested by Janata and Reisberg (1988). They presented subjects with an ascending major scale or major chord context followed by a probe tone. Subjects were asked to decide whether or not the probe belonged in the key of the triad or scale. As predicted, they found that probe tones which were tonic-triad tones were responded to more quickly and accurately than the remaining diatonic tones. The data indicated that, in the chord context, response to the nondiatonic tones was as quick as that of their nearest diatonic neighbor, while in the scale context, this response was faster than the diatonic, nontriad tones and almost as fast as the triad tones. The results for scales suggest an inverted U-shaped function when subjects are required to label events as either context-congruent or context-incongruent. In other words, both conforming (triad tones) and anomalous (nondiatonic tones) events may be labelled quickly and accurately because the level of expectancy or activation is either very high or very low. Inaccurate and slower processing of intermediate events (diatonic, nontriad tones) may occur because there is greater uncertainty concerning the item with respect to the context.

To seriously consider this hypothetical U-shaped function based on an interpretation of the scale data of Janata and Reisberg (1988) is perhaps premature due to limitations of their data and to other procedural problems. First, the data of 9 of the 17 subjects were lost due to equipment failure and thus nondiatonic data from 8 subjects were compared against diatonic data for 17 subjects. Secondly, Janata and Reisberg (1988) considered a yes response as correct for diatonic tones and incorrect for nondiatonic, and a no response as correct for nondiatonic and incorrect for diatonic. Within this scheme, subjects performed the labelling task poorly with error rates approaching 60%, even though they had not been instructed to respond quickly. This is not the "near perfect performance" typically required for a reaction time analysis (i.e., Pachella, 1974, pp. 48-80). Rather, such error rates imply genuine uncertainty or lack of knowledge. As such, the error profiles provided might rather be regarded as measures of the stability of the fit of each note to the prior context. Thus, the reaction times for both correct and incorrect responses could have been analyzed. However, in the absence of the data for incorrect responses, it is unclear what the given reaction time profile actually means. This point is further emphasized by the lack of agreement in the error profiles for the chord and scale contexts. Krumhansl (1979) found that the ratings from both types of context correlated at .83, while the diatonic data of Janata and Reisberg (1988) seem to hover near .35 (computations based upon Figure 2; .32 based upon Figures 2 & 5).

Further indirect evidence supporting an inverted *U*-shaped relation between expectancy and reaction time is also provided by Bharucha and Stoeckig (1986). Subjects heard two chords in succession and were asked to label the second chord as either major or minor. It was assumed that the first chord (the prime) would set up an expectation for related chords. This would aid processing if the second chord (the target) was related and would hinder processing if the second chord was unrelated. However, it was found that when the prime was a major chord, the unrelated major chords were processed more slowly and more inaccurately than either the related major chords or related minor chords or unrelated minor chords. That is, what might be defined as moderate incongruence led to higher reaction times than did either very high or very low congruence. A similar pattern of results occurred when the prime was a minor chord. Although the authors explained their results in terms of response bias and consonance/dissonance, it is possible to invoke the inverted *U*-shaped function as an alternative way of regarding the data.

The following experiment was designed to assess the expectancies established by a major scale context for the 12 chromatic tones and to evaluate associated reaction time data. The subject's task was to rate the fit of a probe tone to a preceding major scale context in a two-level, forced choice (good-fit or poor-fit) paradigm, using any criteria that they wished. Both the proportion of good-fit responses and the reaction time for each of 13 probe tones were obtained. Following Krumhansl and Kessler (1982), it was postulated that hearing the tonal context would lead to the activation of the internal hierarchy of notes. If the probe tone was perfectly congruent with the expectancies, the decision concerning its fit would be relatively easy; the response would be consistently good-fit and fast. If the probe tone was completely incongruent with those expectancies, again the decision would be easy, the response consistently poor-fit and fast. Only if the probe tone was associated with moderate expectancies would the decision be difficult, and the response inconsistent and slow. As a rule, when the proportion of good-fit responses approached either 0.0 or 1.0, reaction times would be short and when the proportion approached 0.5, reaction times would be long. This inverted *U*-shape function will be referred to as the rating/reaction time rule.

The present procedure resembles that of Krumhansl and her colleagues, particularly Krumhansl and Shepard (1979). In their work, response profiles when averaged over highly trained subjects, typically showed the predicted three-level tonal hierarchy, with only minor variations. Similar findings with highly trained subjects have also been demonstrated in variations of the task (i.e., Krumhansl, 1979; Krumhansl & Kessler, 1982). Krumhansl and Shepard also examined subjects with less training and found two other basic response profiles, in addition to the three-level hierarchy associated with high training. For some subjects, the profile exhibited two levels in which diatonic notes (including the triad notes) were delineated from the nondiatonic notes. For other subjects, judgments seemed to be based on proximity. Notes nearest, in absolute pitch, to the last note of the context were judged to be the most highly related. The three different profiles seemed related to the average level of training, 7.4, 6.1, and 0.7 years respectively. Further evidence for the importance of training in the development of the hierarchical response profile was provided by Krumhansl and Keil (1982), using a variation of the Krumhansl and Shepard (1979) task that involves two rather than one probe tone. They found a developmental differentiation of the 12 chromatic notes that proceeded through the two-level hierarchy (grades 1 and 2, averaging 0.9 years of instruction) to the three-level hierarchy (achieved by adults, averaging 6.8 years of instruction).

However, Cuddy and Badertscher (1987) found no such developmental sequence using the completed ascending scale (as well as the major and diminished triads) as context and a single probe tone. Even the youngest group (grades 1 and 2) evidenced a three-level hierarchy. The same was true for three groups of undergraduate university students averaging 0.5, 7.5, and 14.5 years of training: all groups evidenced a three-level hierarchical profile, although those of the less trained subjects were not as sharply defined. Janata and Reisberg (1988) following the previous results of Cuddy and Badertscher (1987) also collapsed over all subjects. However, their profiles for the scale context demonstrated a strong effect of proximity (which they discussed in terms of recency).

Given the ambiguity in the empirical evidence relating training to the internal representation of tonality, it would seem that the continued examination of individual profiles is in order. It was therefore decided to collect probe tone, response profiles from a large number of subjects having a wide variety of training. These subjects would then be classified using a cluster analysis in a manner similar to that of Krumhansl and Shepard (1979). The response profiles for groups of classified subjects would be assessed for the effects of tonal expectancy as measured by reaction time. In order to circumvent the possibility of judgments based upon contour continuation, completed scales and probe tones within the same octave as the scale were chosen. The use of a simple binary decision was intended to facilitate the acquisition of reaction-time data.

If distinct response profiles emerged from the clustering, it was expected that the associated patterns for the reaction times would also be distinct, but in keeping with the U-shaped response rating/reaction-time rule mentioned previously. For profiles having three levels, reaction times would be short for tonic triad and nondiatonic tones and longer for diatonic, nontriad notes. For profiles having only two levels, two patterns of results were considered possible. If, on one hand, the response for all diatonic notes was consistently good-fit and for all nondiatonic notes consistently poor fit, then no differences in reaction time should be noted. If, on the other hand, the response for either the diatonic or nondiatonic notes was inconsistent (a rating near 0.5), then longer reaction times would be associated with those notes. Finally, for a proximity profile, an inverted *U*-shaped function might be expected because tones closest to the final note of the scale would be highly primed and those most remote would be most weakly primed. However, if the proximity effect dropped off abruptly, no differentiation of reaction times should be observed.

In summary, subjects heard complete ascending or descending scales, followed by a single probe tone. Subjects were asked to rate the fit of the probe to the preceding context in a two-level, forced choice decision. Averaged responses to the

13 probe tones were expected to yield either a three-level, two-level, or proximity profile. Reaction time profiles were expected to differ with response profile, and were expected to follow the inverted *U*-shaped response rating/reaction-time rule.

Method

Subjects

Sixty- one subjects (38 females and 23 males) having a mean age of 21.97 ±.59 years (range: 16 - 38) and a wide range of formal music training (range: 0 - 29 instrument-years, mean: 5.64 ±.85 years), were recruited from the university community. Because the basis for the assignment of subjects to groups would be a cluster analysis, no rigid selection criteria were imposed. However, in order to obtain a complete range of training levels, 19 subjects were selected with no training, 24 with between 1 and 7 years (inclusive), and 18 with more than 8 years. Three subjects reported having absolute pitch. Three subjects were volunteers, 26 participated for class credit, and 23 were paid at a rate of $6.00 per hour.

Stimuli

For each trial, the context sequences consisted of an 8-note ascending or descending scale in the key of C. It was followed by a single probe tone drawn randomly from 13 chromatic notes in the octave bounded by the context. The order of presentation for each context-probe combination was randomized for each subject such that within each block of 26 trials each probe-combination was presented only once.

Thirteen sine tones, representing the equal-tempered chromatic scale within the octave C^4 to C^5 (523.25 to 1046.50 Hz) were digitally synthesized on a Commodore Amiga 500 computer, with 8-bit resolution and a sampling frequency of 21 kHz (cf., Cohen & Mieszkowski, 1989). All tones had durations of 175 ms and rise/decay times of 25 ms. The intertone interval was also 75 ms. Tones were presented binaurally through a pair a Realistic LV10 headphones connected directly to the audio output of an Amiga 500 or 1000, in identical Industrial Acoustics single-walled, sound-attenuating rooms.

Procedure

Subjects were seated in front of the computer. Instruction was provided by the computer at a pace determined by the subject. The experimenter remained with the subject throughout the presentation of the instructions and during the practice sessions in order to provide any necessary clarifications. Abbreviated instructions remained on the screen for the duration of the experiment.

Subjects were instructed to compare the last note of the sequence (the ninth note) with the preceding eight. They then indicated whether the last note was a good or poor fit to the preceding eight notes. Subjects were instructed to use whatever criteria they wished with the restriction that the response was to be as natural or "instinctive" as possible. The presentation of each trial was initiated, under control of the subject, by the the keyboard space bar, and the response was indicated by one of two arrow keys. The use of the left hand for the space bar and two fingers of the right hand for the arrow keys was demonstrated, but not enforced. The computer visually indicated immediately when each response had been recorded. Subjects were instructed to rest when desired by not initiating the subsequent trial. Subjects completed a practice phase with 2 blocks of 12 trials each. The subsequent test phase consisted of 4 blocks of 26 trials each. The practice phase was the same as the test phase except that the context sequences consisted of only three notes, and there were only 12 context/probe combinations. The entire experiment lasted approximately 25 minutes.

Results

Performance for each subject was scored as the proportion of good-fit replies out of four for each of the 13 probes for ascending and descending scales. The proportion of good- fit replies are, in future, referred to as ratings. The performance of each subject thus consisted of a profile of 26 proportions ranging from 0.0 (poor-fit) to 1.0 (good-fit). In addition, the median reaction time for each of the 26 context-probe conditions (using both the good- and poor-fit responses) was also calculated, yielding a corresponding reaction-time profile.

As in Krumhansl and Shepard (1979), the correlation matrix of the ratings profiles of all subjects was used as a measure of intersubject distance in a cluster analysis (Johnson, 1969) using both the compactedness and connectedness options. For a correlation coefficient based on 26 pairs, an r value greater than 0.49 is significantly different from 0.0 ($p<.01$), so groups were considered distinct when delineated at the $r = 0.50$ level. This implied that individuals within each group accounted for at least 25% of the variance in each other. Reducing the criterion to $r = 0.38$ ($p<.05$) would have reduced the number of groups by two (see Figure 1). Unfortunately, this criterion would have implied that individuals within a group accounted for only 15% of the variance in each other. (Examination of the data of Krumhansl & Shepard indicated that they delineated groups using an $r = 0.70$, or 50% of the variance.)

The clustering for the compactedness option (Figure 1) resulted in 12 groups with 3 unclassified subjects who are not further discussed. The results using the connectedness option, while qualitatively similar, were not as well delineated. A summary of group characteristics is presented in Table 1.

Ratings Analysis

The average ratings were analyzed in two parallel manners. In the first analysis, subjects were represented by their ratings profiles (the mean proportions of good-fit replies for each of the 26 tones). The data from all subjects within a group were averaged to create a profile of the rating for that group. Each group was independently analyzed with a repeated measures one-way ANOVA having 26 cells, using a set of 25 planned orthogonal contrasts designed to aid in the identification of the ratings pattern or profile, as triadic (i.e., having three levels), diatonic (i.e., having two levels), or proximity. In the second analysis, mean proportions for each individual subject were transformed to z-scores and were then averaged to create an average z-score rating profile for each group. Each group was

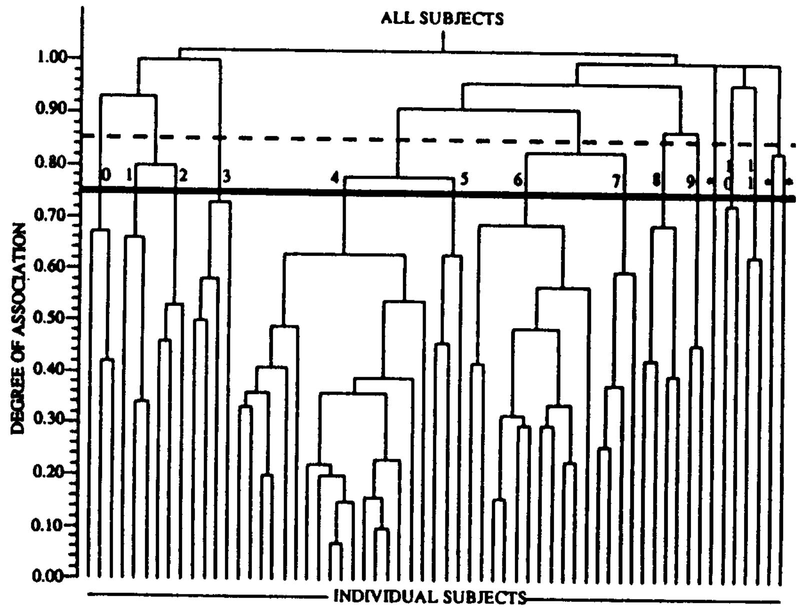

Figure 1. The cluster analysis using the algorithm of Johnson (1969). Each individual subject is represented by a line along the abscissa. The ordinate represents the degree of association between elements (either individuals in a group or between groups of individuals). The degree of association is measured as the percent of variance accounted for (essentially r^2). The criterion for delineating groups ($r = 0.50$, $r^2 = 0.25$) is represented by the dark line. Numbers adjacent to the dark line indicate the group label. Relaxing the criterion to $r = 0.38$ ($r^2 = 0.15$) would have resulted in the cut line indicated by the dash line.

Table 1
Group characteristics and labels.

Group	#	Age	Total	Max.	Instr.	Label
0	3	21.00 ±2.65	0.33 ±0.58	0.33 ±0.58	0.33 ±0.58	Proximity - Diatonic
1	3	20.07 ±2.52	3.00 ±1.00	2.00 ±1.73	1.33 ±0.58	Proximity
2	3	20.67 ±1.15	2.67 ±4.62	1.00 ±1.73	1.00 ±1.73	Proximity + extras
3	4	19.25 ±1.89	2.38 ±2.29	1.96 ±2.19	1.25 ±0.96	Proximity & Diatonic
4	17	23.15 ±5.54	8.23 ±7.89	5.53 ±4.65	1.65 ±1.32	Diatonic
5	3	24.00 ±4.58	10.00 ±9.17	5.67 ±5.13	2.33 ±2.08	Diatonic
6	11	21.82 ±3.43	7.27 ±4.78	4.64 ±3.29	2.36 ±0.50	Triadic
7	4	20.25 ±3.30	2.13 ±3.92	1.13 ±1.93	0.75 ±0.96	Triadic - Diatonic
8	4	21.25 ±2.63	5.75 ±5.44	4.50 ±4.20	1.50 ±1.00	Diatonic & Proximity
9	2	26.00 ±7.07	2.50 ±3.54	2.50 ±3.54	0.50 ±0.71	Diatonic & Proximity
10	2	20.50 ±2.12	0.40 ±0.57	0.40 ±0.57	0.50 ±0.71	Diatonic + extras
11	2	28.00 ±14.14	0.00 ±0.00	0.00 ±0.00	0.00 ±0.00	Proximity + extras
12	3	18.67 ±0.58	9.67 ±13.43	6.67 ±8.83	1.33 ±1.53	Not Classified

Note: Total training (Total) is the number of years formal training on all instruments studied. Maximum Training (Max) is the maximum number of years of formal training on any one instrument. Number of instruments (Instr) is the number of instruments formally studied. All errors are standard deviations. A '-' in the label indicates that the group used one profile for the ascending and the other for the descending context. A '+' indicates that emphasis was given to specific notes. An '&' indicates that the group used a mixture of the 2 profiles.

independently analyzed using a simple one-way ANOVA having 26 cells, and the same set of 25 planned orthogonal contrasts. Two analyses were used because it was thought that the first would give undue weight to those subjects having widely varying responses (i.e., large standard deviations), while the second would give undue weight to those subjects who did not differentiate between conditions. Because both analyses yielded the same pattern of results, only the first will be discussed in detail. The relevant means, standard deviations, and omnibus F tests for both analyses are presented in Table 2. For simplicity, only the omnibus significance levels have been cited (a complete listing of the 25 contrasts for each of the 12 groups for two different analyses is available from the authors). Also, means and standard deviations for the z-score analysis are not presented (they are uniformly 0.0 and 1.0, respectively). Included in Table 2 is the correlation between the profile for the ascending context and the profile for the descending context. A high positive correlation indicated that the group performed similarly in both contexts: a high negative correlation indicated that the group did the opposite in the two types of context. Ideally, a positive correlation was expected from the triadic or diatonic groups, while a negative correlation was expected for the proximity groups.

In the following, the groups of subjects are referred to by numbers which arbitrarily fell out of the cluster analysis. Only Group 6 (Figure 2), accounting for

Table 2
Mean ratings and omnibus tests for all groups.

Group	*n*	mean	*sd*	Analysis 1	Analysis 2	Correlation
0	3	.481	±.092	F(25,50) = 06.32**	F(25,52) = 06.57**	0.37
1	3	.551	±.106	F(25,50) = 08.30**	F(25,52) = 08.87**	-0.78**
2	3	.401	±.039	F(25,50) = 07.94**	F(25,52) = 08.52**	-0.65**
3	4	.255	±.122	F(25,75) =0 7.68**	F(25,78) = 08.15**	-0.32
4	17	.553	±.053	F(25,400) = 69.09**	F(25,416) = 72.91**	0.88**
5	3	.667	±.111	F(25,50) = 07.92**	F(25,52) = 08.23**	0.67**
6	11	.403	±.060	F(25,250) = 34.58**	F(25,260) = 35.97**	0.85**
7	4	.478	±.119	F(25,75) = 12.00**	F(25,78) = 13.25**	0.44
8	4	.464	±.120	F(25,75) = 08.54**	F(25,78) = 09.02**	-0.01
9	2	.414	±.136	F(25,25) = 0 6.19**	F(25,26) = 06.78**	0.47
10	2	.534	±.061	F(25,25) = 02.80**	F(25,26) = 03.33**	0.54
11	2	.577	±.095	F(25,25) = 04.12**	F(25,26) = 04.28**	0.71**

Note: The means were calculated by first determining the mean proportion of good-fit responses for each subject (all 26 context-probe conditions) and then the mean and standard deviation for each group. Each group was analyzed in 2 ANOVAs. The first was a one-way, within subjects design on the raw (untransformed) scores. The second was a simple one- way design using the z-scores (transformed). Hence the difference in the *df* in the denominators. The *n* is the number per group. Correlation is the correlation between the ascending and descending context within each group. The '*' and '**' mean significant at the .05 and .01 respectively.

11 subjects, was identified as triadic. It exhibited a significant triadic profile for both the ascending and descending contexts with slight exceptions, likely due to proximity: Note B was higher than expected in the ascending context, as was note D in the descending context. Subjects in this group performed similarly in the ascending and descending contexts: the correlation between the two types of contexts was $r = 0.85$. The average of the ascending and descending contexts correlated at $r = 0.97$ with the average profile of Krumhansl and Kessler (1982). This group of subjects corresponded to Group 1 of Krumhansl and Shepard (1979).

Groups 4 and 5 (Figure 2), accounting for a total of 20 subjects, were identified as diatonic. Group 4 produced a classic diatonic profile for the ascending context: all diatonic tones were rated near 1.0 and all nondiatonic tones were rated near 0.0. In the descending context, the note B was rated at the level of the nondiatonic notes. Group 5 was almost identical to Group 4, with the exception that nondiatonic notes, particularly in the ascending context, were rated somewhat higher (i.e., not rated at 0.0 or poor-fit). Both of these groups performed similarly in the ascending and descending contexts: the correlations between the two types of context were $r = 0.88$ and $r = 0.67$ for Groups 4 and 5 respectively. These two groups corresponded to Group 2 of Krumhansl and Shepard (1979). It is interesting to note that these two groups would have merged if the criterion for group delineation were to be relaxed to $r = 0.38$ ($p<.05$).

Three groups, totalling 10 subjects, evidenced proximity profiles (Figure 2). Of these three groups, only Group 1 exhibited a prototypical proximity pattern. The most favored notes were those lying in close pitch proximity to the final note of the sequence. Note that this profile does not represent a recency pattern because the last tone of the context (either C^4 or C^5) was not given the highest rating. Group 2 made use of proximity, with extra emphasis on note G in the ascending context and on note F in the descending context. The profile for Group 3 could be a variation of proximity or possibly, a true recency effect since the highest ratings went to the last tones of the context. These three groups corresponded to Group 3 of Krumhansl and Shepard (1979). Groups 1 and 2 would have merged if the criterion for group delineation were to be relaxed to $r = 0.38$ ($p<.05$).

The remaining groups, accounting for only 17 of the 61 subjects, could be described as a mixture of the above profiles (Figure 2). Group 7 represented a mixture of the triadic and diatonic groups. The profile for the ascending context seemed to be triadic while the profile for the descending context was almost identical to the descending contexts of Groups 4 and 5. Interestingly, Group 7 would have merged with Group 6 (triadic) if the criterion for group delineation had been relaxed to $r = 0.38$ ($p<.05$). Group 0 was diatonic, almost triadic, in the ascending context, and was a mixture of triadic, diatonic, and proximity in the descending context. Groups 8 and 9 appeared to mix diatonic and proximity profiles. In Group 10, the diatonic pattern, although significant, was difficult to discern in the ascending context. Finally, Group 11 showed evidence of proximity in the ascending context and an unidentifiable pattern in the descending context.

The three subjects reporting perfect pitch did not cluster together; one remained unclassified while the other two appeared in Group 4.

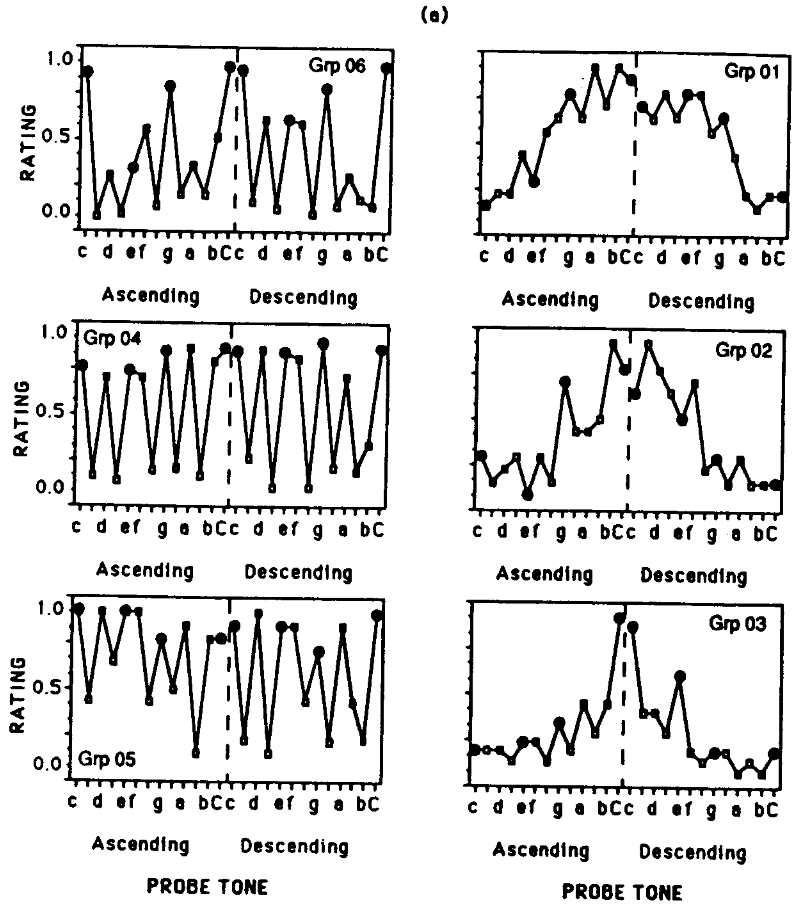

Figure 2a. The response ratings profiles for Groups 6 (triadic), 4 and 5 (diatonic), and 1, 2, and 3 (proximity). These profiles are composed of the average rating of each subject in each context-probe condition. The dark circles correspond to the tonic triad tones, the dark squares to the diatonic, nontriad tones, and the open squares to the nondiatonic tones.

(b)

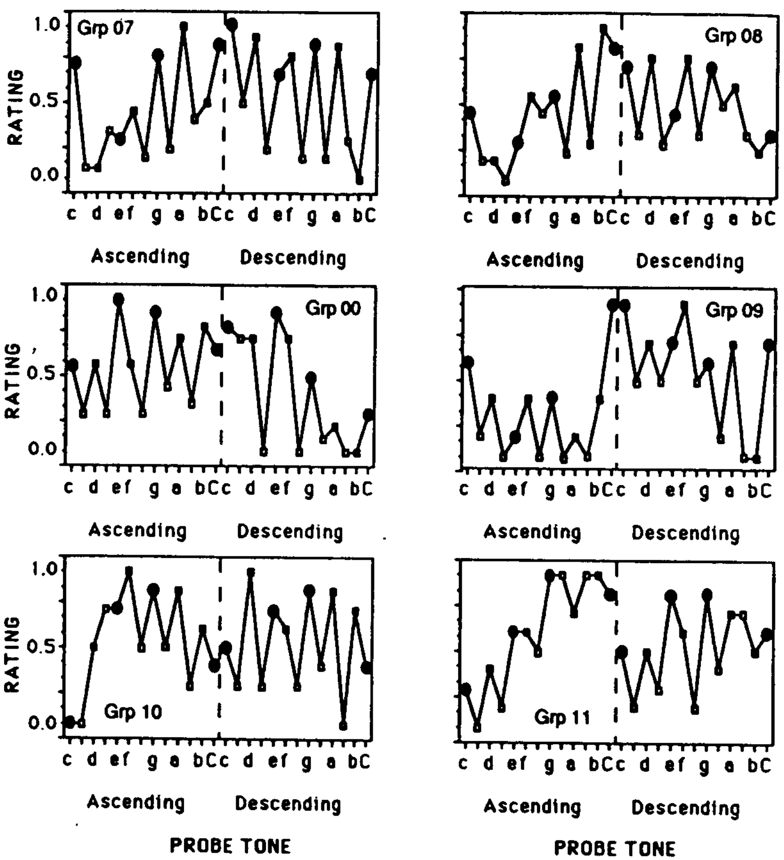

Figure 2b. The response ratings profiles for Groups 7 (triadic and diatonic), 0, 8, 9, and 10 (diatonic and proximity), and 11 (undefined). These profiles are composed of the average rating of each subject in each context-probe condition. The dark circles correspond to the tonic triad tones, the dark squares to the diatonic, nontriad tones, and the open squares to the nondiatonic tones.

Relation Between Ratings Profile and Training

The total number of years of formal training, the maximum number of years of formal training on a single instrument, and the number of instruments on which formal training was received were analyzed in individual one-way ANOVAs (unequal n) over the 12 groups. No measure was significant, although there did seem to be a trend (see Table 1). The groups having diatonic profiles had the highest level of training (Groups 4 & 5: averaging 9.12 years). The triadic group followed closely behind (Group 6: 7.27 years). The group defined as a mixture of triadic and diatonic (Group 7: 2.13), however, did not exhibit a level of training between or near those groups. The group showing a proximity profile (Group 1: 3.00 years) and those groups described as proximity plus embellishments were lower in training (Groups 2 & 3: averaging 2.50 years). The remaining groups were slightly lower (Groups 0, 9, 10, & 11: averaging 1.07 years), with the exceptions of one group defined as diatonic and proximity (Group 8: 5.76 years) and the "group" containing the unclassified subjects one of whom had 25 instrument-years of training and two of whom had almost none. A set of post hoc contrasts, in the method of Rodger (1974; see also Howell, 1982), revealed no significant effects between groups with different profiles.

Reaction-Time Analysis

The reaction-time data were analyzed in two parallel ways in a manner analogous to the rating data, for the reasons previously given. For the first analysis, the median of four response times for each of the 26 tones were used to create a reaction-time profile for that subject. Within each group, individual profiles were averaged to create an average reaction-time profile for that group. The data for each group were submitted to a repeated measures one-way ANOVA having 26 cells, using a set of 25 planned orthogonal contrasts, that were designed to complement those used with the ratings data and to demonstrate the aforementioned response rating/reaction-time rule. In the second analysis, respective *z*-scores for each subject were averaged to create an average *z*-score reaction-time profile for the group. Again, both analyses yielded generally the same pattern of results. The *z*-score analysis will be discussed in detail. The relevant means, standard deviations, and omnibus *F* tests for both analyses are presented in Table 3.

As shown in Figure 3, for Group 6, previously defined as triadic, subjects were generally slower for the diatonic, nontriad tones (although B in the descending context was actually the fastest). Groups 4 and 5 were defined as diatonic (Figure 3), and the reaction-time profiles for both groups were fairly flat, although for Group 5, notes D and B in ascending context and B in the descending context were slower (for the raw data, only B in the descending context was significantly slower). It is noteworthy that for Group 5, the nondiatonic tones resulted in generally longer reaction times (see Figure 2): This was not the case for Group 4. This generally matched the prediction that reaction times would be longer when the rating approached 0.5: in both Groups 4 and 5, the diatonic tones were rated near 1.0, but in Group 5, the nondiatonic tones were rated near 0.5, while in Group 4, the nondiatonic tones were rated near 0.0.

Of the groups defined as proximity (Figure 3), in Group 1, the profiles for the ascending and descending scales indicated that subjects were slower when the ratings were near 0.5. Group 2 was essentially undifferentiated in response, while Group 3 exhibited a pattern that was reminiscent of the triadic group: notes D and A in the ascending context were slow.

Group 7 (Figure 3) did not exhibit any obvious pattern in the reaction time profile. Group 0 had long reaction times on notes C# D A and C^5 in the descending context. Groups 8, 9, and 10 (Figure 3) all demonstrated undifferentiated profiles. However, for Group 8, the ratings in the descending context were, on average, closer to 0.5 than in the ascending context (see Figure 2) and reaction times were significantly longer for the descending context. Group 11 (Figure 3) was difficult to classify in any manner. Reaction times were slow for notes D and F# in the ascending context and G and A# in the descending context.

Relation between Ratings and Reaction Time

The final analysis explored the relation between the rating profile and the reaction-time profile within each group, using a correlational analysis. A rating of 0.0 or 1.0 (proportion of good-fit's) indicated certainty. It was expected that such ratings would be paired with fast reaction times. A rating of 0.5 indicated uncertainty and, as such, was expected to be paired with a slow reaction time. To

Table 3

Mean reaction times and omnibus tests for all groups.

Group	*n*	mean	*sd*	Analysis 1	Analysis 2
0	3	944	±473	F(25,50) = 1.76*	F(25,52) = 2.14*
1	3	703	±196	F(25,50) = 1.41	F(25,52) = 1.64
2	3	845	±192	F(25,50) = 1.37	F(25,52) = 1.39
3	4	730	±168	F(25,75) = 1.06	F(25,78) = 1.16
4	17	1035	±827	F(25,400)= 1.42	F(25,416)= 1.53*
5	3	1013	±605	F(25,50) = 0.92	F(25,52) = 0.94
6	11	895	±314	F(25,250)= 2.52**	F(25,260)= 3.21**
7	4	1092	±646	F(25,75) = 1.23	F(25,78) = 0.99
8	4	1009	±154	F(25,75) = 1.04	F(25,78) = 1.02
9	2	1238	±975	F(25,25) = 0.85	F(25,26) = 0.58
10	2	984	±22	F(25,25) = 1.09	F(25,26) = 1.13
11	2	773	±16	F(25,25) = 3.08**	F(25,26) = 3.39**

Note: The reaction times (in ms) were calculated by first determining the mean reaction time for each subject (all 26 context-probe conditions) and then the mean and standard deviation for the group. Each group was analyzed in 2 ANOVAs. The first was a one-way within subjects design on the raw (untransformed) scores. The second was a simple one-way design on the *z*-scores (transformed). Hence the difference in the *df* in the denominators. The *n* is the number per group. The '*' and '**' mean significant at .05 and .01 respectively.

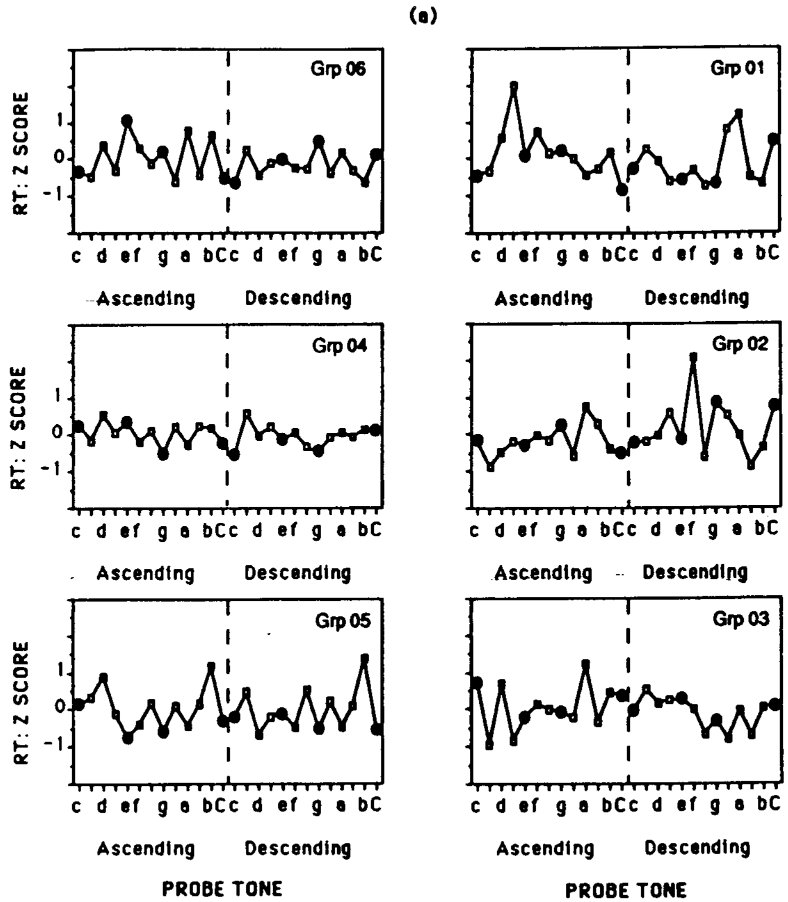

Figure 3a. The reaction-time profiles for groups 6 (triadic), 4 and 5 (diatonic), and 1, 2, and 3 (proximity). These profiles are composed of the average rating of each subject in each context-probe condition. The dark circles correspond to the tonic triad tones, the dark squares to the diatonic, nontriad tones, and the open squares to the nondiatonic tones.

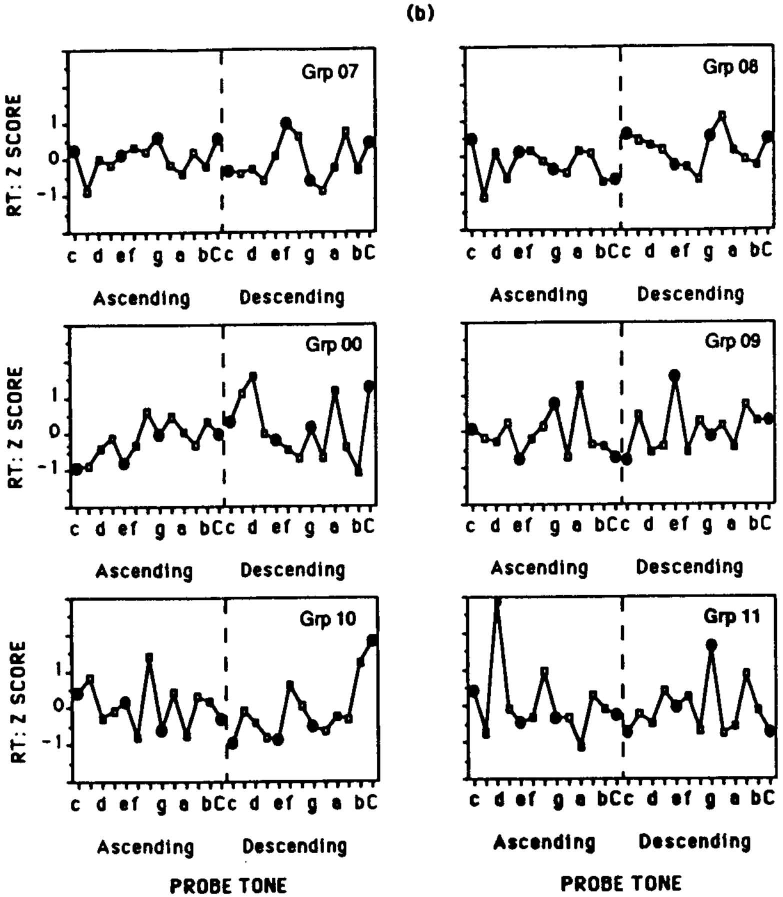

Figure 3b. The reaction-time profiles for groups 7 (triadic and diatonic), 0, 8, 9, and 10 (diatonic and proximity), and 11 (undefined). These profiles are composed of the average rating of each subject in each context-probe condition. The dark circles correspond to the tonic triad tones, the dark squares to the diatonic, nontriad tones, and the open squares to the nondiatonic tones.

make it possible to compare the responses with the reaction time in a Pearson product moment correlation, responses were converted according to the formula:

new response = 0.5 - absolute value(old response - 0.5)

Thus, highly certain ratings (good or poor) had a value of 0.0 and the level of certainty decreased from 0.01 to 0.50.

For Group 6, which had exhibited a three-level hierarchical pattern, there was a significant correlation of $r = 0.68$ ($p<.01$) between the reaction time and adjusted ratings profiles, supporting the rating/reaction-time rule.

Although for both diatonic Groups 4 and 5, reaction-time profiles were undifferentiated, Group 4 produced a significant correlation, $r = 0.52$ ($p<.01$). Hence, although the ratings in Group 4 indicated that the subjects were always certain of their response, subjects were still slower when less certain. In Group 5, the correlation was not significant ($r = 0.28$; $p>.10$), despite the fact that these subjects were generally slower when less certain (i.e., slower for the nondiatonic tones): the lack of a significant correlation is likely due to the notes B and D, which produced the longest reaction times.

For Group 1, the correlation between the reaction-times and adjusted ratings approached significance ($r = 0.34$; $p<.09$). Group 2, which was essentially undifferentiated in reaction-time response, demonstrated no correlation between the reaction-time and adjusted ratings profiles ($r = 0.26$; $p>.10$). Group 3, which exhibited a reaction-time pattern reminiscent of the triadic group, had a significant correlation of $r = 0.47$ ($p<.05$).

The absence of any pattern in the reaction time profile of Group 7 was reflected in a nonsignificant correlation ($r = 0.23$; $p>.10$) between the reaction-time and adjusted ratings profiles. Although Group 0 exhibited some pattern in its reaction-time profile, the slowest reaction times were not always associated with the notes on which the ratings approached 0.5, as shown by the nonsignificant correlation ($r = 0.23$; $p>.10$). Groups 8, 9, and 10 all demonstrated undifferentiated reaction-time profiles and correspondingly, Groups 9 and 10 had nonsignificant correlations ($r = 0.15$; $p>.10$, and $r = .05$; $p>.10$, respectively). However, the correlation for Group 8 was significant at $r = 0.50$ ($p<.01$) which can be explained by the differences between the ascending and descending contexts. Group 11, which was difficult to classify had a nonsignificant correlation ($r = 0.07$; $p>.10$).

Discussion

With respect to the role of tonality in the generation of expectations, it was predicted that hearing a tonal context would lead to the activation of an internal hierarchy of notes and a consequent set of expectations. Decisions concerning the fit of a probe tone would be consistently good-fit and fast for tones congruent with expectations, and consistently poor-fit and fast for tones incongruent with expectations. Only for a probe tone that was an imperfect or partial match to the context, would the decision be difficult and the response inconsistent and slow. The following relationship between the proportion of trials in which the tone was rated as a good-fit and the speed of the response was predicted: when the proportion approached either 0.0 or 1.0, reaction time would be low and when the proportion approached 0.5, reaction time would be high. Subjects were grouped and classified according to their ratings (proportion of good- fit replies) for all 13 probe tones–the ratings profile. For each group, the average reaction-time profile was compared with the average rating profile.

There were a number of groups for which the predicted relationship between reaction time and rating held true. The group classified as triadic (Group 6 ; $n = 11$) demonstrated such a pattern. Of the Groups that were classified as diatonic (Group 4; $n = 17$ and Group 5; $n = 3$), Group 4 exhibited the predicted pattern in its reaction time profile. Group 5 demonstrated generally longer reaction times for those tones having a rating near 0.5 (i.e., the nondiatonic tones), but the slowest reaction times belonged to the notes D and B, although these notes had received ratings near 1.0 or 0.0. Of the groups classified as proximity, subjects in Group 3 ($n = 4$) also indicated that when they were certain of the rating, they were fast, and when they were uncertain of the rating, they were slow. Group 1 ($n = 3$), much like Group 3, also produced the predicted pattern, although the correlation did not quite reach significance. Of the remaining groups, Group 8 ($n = 4$), defined as diatonic and proximity, demonstrated the predicted pattern.

In total, groups comprising 39 of the 61 subjects (including Group 3, but not Group 5) demonstrated the predicted pattern of results. Of the remaining groups, most did not demonstrate any effects at all, which was not surprising given the small number of subjects within each group (Group 2; $n = 3$; Group 7; $n = 4$; Group 9; $n = 2$, and Group 10; $n = 2$). Only Groups 0 ($n = 3$) and 11 ($n = 2$) exhibited patterns truly contrary to expectations.

These findings, especially given the results for the larger groups, add to the evidence indicating that individuals have an internal representation of a tonal hierarchy that affects the processing of subsequent tones. These results were also consistent with the scale data of Janata and Reisberg (1988). Furthermore, since the results from the reaction-time analysis generally complemented the ratings analysis, this experiment demonstrates some promise for the utility of reaction time in the study of the internal representation of tonal hierarchies. It is also possible that the reaction-time measure may be more sensitive than the rating measure. For example, for Group 5 (diatonic), the notes B and D were rated near their respective neighbors (i.e., C^5 & A# or D# & C#) but these notes resulted in the longest reaction times, indicating some uncertainty in the fit of those tones.

The evidence for individual differences in profiles in this study supports the findings of Krumhansl and Shepard (1979) and Krumhansl and Keil (1982) concerning individual differences for the internalized tonal representation. In this study, not all subjects provided evidence of an internalized three-level hierarchy. Some demonstrated a two-level hierarchy and some responded on the basis of proximity. In their classification, Krumhansl and Shepard obtained only three groups of subjects with one unclassified subject. In the present study, there were 12 groups of subjects with three unclassified subjects. Nevertheless, because the present work seemed to result in three major categories of subjects, it may be reconciled with the

work of Krumhansl and Shepard. A total of 11 subjects (Group 6) were classified as triadic. A total of 20 subjects were classified as diatonic (Groups 4 & 5) and a total of 10 subjects were classified as proximity (Groups 1, 2, & 3). Hence, 41 of 61 subjects could be grouped analogously to that of Krumhansl and Shepard. The remaining subjects (20 of 61), fell into some intermediate classification that combined elements of the above classification and/or used different response styles for the ascending and descending contexts. Given that Krumhansl and Shepard based their classification on 24 subjects while this study had 61, the observation of a larger number of groups, some representing intermediate concepts of tonality is not so surprising. It should also be noted that most of these intermediate groups represented a low number of subjects. Moreover, a total of 31 of the 61 subjects responded strictly on the basis of a two- or three-level hierarchy, and furthermore, 48 of the 61 subjects (51 when the unclassified subjects were included) responded to some degree on the basis of diatonism.

It is possible that the number of groups could have been reduced by combining similar groups together. However, this would inevitably lower the degree to which individuals within a group would be associated. In our opinion, this is to be avoided since the process of averaging unrelated subjects could result in undifferentiated profiles, or perhaps, the existence of a tonal hierarchy that is not representative of all the individuals within that group (i.e., if an undifferentiated profile were to be averaged with a triadic profile, the result would resemble a triadic profile). The large number of different profiles may have resulted, in part, from the use of a dichotomous choice method for obtaining the probe-tone profiles rather than the seven-point rating method used by Krumhansl and coworkers. Perhaps increasing the number of trials per probe tone would have led to more consistent responding and consequently, a fewer number of groups. This is an empirical question. Given the high correlation between the data of the triadic group (Group 6) and those of Krumhansl and Kessler (1982), we believe that for many of the subjects, ratings profiles had stabilized within the duration of the experiment.

As with Krumhansl and Shepard (1979), within any particular group, individual members demonstrated a wide variation in the number of years of formal training. However, in both studies, there was a tendency for groups evidencing tonal hierarchies (i.e., triadic or diatonic) to have a higher average level of training. However, the link between type of profile and training was not strong.

In summary, this work has demonstrated possession of an internal tonal hierarchy among a number of listeners who were considerably heterogenous with respect to music training. To some extent, this hierarchy influenced reaction-time of the goodness of fit judgment, according to an inverted U-shaped function: those events that conform to strong expectancies or are anomalous with respect to these expectancies are also labelled quickly. Events that are intermediate in the level of expectancy tend to be labelled slowly.

However, groups of individuals evidenced different types of response profiles, which were mirrored in their reaction-time profiles. Therefore, if a researcher wishes to manipulate expectancy, care must be taken to ascertain which of the possible tonal hierarchies is actually used by a particular subject. Training can be a guide in this assessment, but it is only a guide. This conclusion is consistent with Unyk and Carlsen (1987) who felt it necessary to assess the expectancy pattern for individual subjects in order to subsequently investigate effects of violating expectancy on melodic processing.

References

Bharucha, J.J., & Stoeckig, K. (1986). Reaction time and musical expectancy: Priming of chords. *Journal of Experimental Psychology: Human Perception and Performance, 12*, 403-410.

Cohen, A.J., & Mieszkowski, M. (1989). Frequency synthesis with the Commodore Amiga for research on perception and memory of pitch. *Behavior Research Methods, Instruments & Computers, 21*, 623-626.

Cuddy, L.L., & Badertscher, B. (1987). Recovery of the tonal hierarchy: Some comparisons across age and levels of musical experience. *Perception and Psychophysics, 41*, 609-620.

Howell, D.C. (1982). Statistical methods for psychology. Boston: Duxbury Press.

Janata, P., & Reisberg, D. (1988). Response-time measures as a means of exploring tonal hierarchies. *Music Perception, 6*, 161-172.

Johnson, S.C. (1969). Hierarchical clustering schemes. *Psychometrika, 32*, 241-254.

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

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

Krumhansl, C.L., & Keil, F.C. (1982). Acquisition of the hierarchy of tonal functions in music. *Memory and Cognition, 10*, 243-251.

Krumhansl, C.L., & Kessler, E.J. (1982). Tracing the dynamic changes in the perceived tonal organization in a spatial representation of musical keys. *Psychological Review, 89*, 334-368.

Krumhansl, C.L., & Shepard, R.N. (1979). Quantification of the hierarchy of tonal functions within a diatonic context. *Journal of Experimental Psychology: Human Perception and Performance, 5*, 579-594.

Pachella, L. (1974). The interpretation of reaction time in information processing research. In B. Kantowitz (Ed.), *Human information processing: Tutorials in performance and cognition* (pp. 48-80). New York: John Wiley and Sons.

Rodger, R.S. (1974). Multiple contrasts, factors, error rates, and power. *British Journal of Statistical Psychology, 27*, 179-198.

Schmuckler, M. (1989). Expectation in music: Investigation of melodic and harmonic processes. *Music Perception, 7(2)*, 109-150.

Unyk, A.M., & Carlsen, J.A. (1987). The influence of expectancy on melodic perception. *Psychomusicology, 4*, 3-23.

Author Notes

The research was based on part of the Masters of Science Thesis by Bradley Frankland and was supported by a grant from the Natural Science and Engineering Council of Canada to Annabel J. Cohen. The assistance of L.A. Wisker in production of figures is gratefully acknowledged. Requests for reprints should be sent to either author at the Department of Psychology, Dalhousie University, Halifax, Nova Scotia, B3H 4J1.