2003

Online Detection of Tonal Pop-Out in Modulating Contexts

Petr Janata
Dartmouth College

Jeffrey L. Birk
Dartmouth College

Barbara Tillmann
Dartmouth College

Jamshed J. Bharucha
Dartmouth College

Dartmouth Digital Commons Citation

Janata, Petr; Birk, Jeffrey L.; Tillmann, Barbara; and Bharucha, Jamshed J., "Online Detection of Tonal Pop-Out in Modulating Contexts" (2003). *Open Dartmouth: Published works by Dartmouth faculty*. 2454.
https://digitalcommons.dartmouth.edu/facoa/2454

Music Perception
Spring 2003, Vol. 20, No. 3, 283–305

Online Detection of Tonal Pop-Out in Modulating Contexts

PETR JANATA, JEFFREY L. BIRK,
BARBARA TILLMANN, & JAMSHED J. BHARUCHA
Dartmouth College

We investigated the spontaneous detection of “wrong notes” in a melody that modulated continuously through all 24 major and minor keys. Three variations of the melody were composed, each of which had distributed within it 96 test tones of the same pitch, for example, A_2. Thus, the test tones would blend into some keys and pop out in others. Participants were not asked to detect or judge specific test tones; rather, they were asked to make a response whenever they heard a note that they thought sounded wrong or out of place. This task enabled us to obtain subjective measures of key membership in a listening situation that approximated a natural musical context. The frequency of observed “wrong-note” responses across keys matched previous tonal hierarchy results obtained using judgments about discrete probes following short contexts. When the test tones were nondiatonic notes in the present context they elicited a response, whereas when the test tones occupied a prominent position in the tonal hierarchy they were not detected. Our findings could also be explained by the relative salience of the test pitch chroma in short-term memory, such that when the test tone belonged to a locally improbable pitch chroma it was more likely to elicit a response. Regardless of whether the local musical context is shaped primarily by “bottom-up” or “top-down” influences, our findings establish a method for estimating the relative salience of individual test events in a continuous melody.

Received June 12, 2002, accepted December 15, 2002

When listening to a performance of Western tonal music, it is generally easy to detect mistakenly played notes or chords. Notes that do not fit into the ongoing tonal context tend to pop out perceptually and have a need to be anchored by the ensuing context in order for coherence in the music to be maintained (Bharucha, 1984, 1996). The pop-out phenomenon could be explained by surface features of the music, for example,

Address correspondence to Petr Janata, Dept. of Psychological and Brain Sciences, 6207 Moore Hall, Dartmouth College, Hanover, NH 03755. (e-mail: petr.janata@dartmouth.edu)

ISSN: 0730-7829. Send requests for permission to reprint to Rights and Permissions, University of California Press, 2000 Center St., Ste. 303, Berkeley, CA 94704-1223.

sudden deviations from locally expected pitch distributions, and/or deviations from expectations generated by stored psychological representations of musical structures. Tonal hierarchies are psychological representations of musical keys that indicate how well each element of the chromatic scale fits into the key (Krumhansl, 1990). The basic features of the tonal hierarchy are that members of the tonic triad (e.g., C, E, and G in C major) fit best, followed by the four remaining diatonic tones. The five nondiatonic tones fit most poorly.

Evidence for tonal hierarchies has been obtained across several experiments using a variety of musical contexts such as ascending and descending scales (Krumhansl & Shepard, 1979), harmonic contexts varying in length from a single chord to several chords (Krumhansl & Kessler, 1982), and short melodies (Hebert, Peretz, & Gagnon, 1995). In such experiments, a tonal context is terminated or followed by a single tone that is rated on a scale of 1 to 7 as to how well it fits with the preceding context. Tones that serve prominent functions in the key of the context, such as the tonic or dominant, are judged to fit well with the context, whereas other diatonic tones and nondiatonic tones are judged to fit more poorly. Probe-tone profiles can also be constructed using the time it takes to determine whether a probe tone fits into the preceding context (Janata & Reisberg, 1988). Tones such as the tonic and dominant are recognized more quickly and accurately than other diatonic tones, whereas nondiatonic tones are recognized with intermediate speed. Tonal hierarchies have been measured using the probe-tone rating technique across participants varying in age and degree of musical experience (Cuddy & Badertscher, 1987; Halpern, Kwak, Bartlett, & Dowling, 1996; Hebert et al., 1995). Although the exact shapes of the profiles vary across studies, the basic features just mentioned are generally preserved. The relative differences among the tones are generally more pronounced in musicians than in nonmusicians and in adults than in children. The existence of tonal hierarchies in nonexpert groups suggests that much of our knowledge of tonal hierarchies is acquired automatically through repeated exposure to musical regularities (Francès, 1958; Krumhansl, 1990; Tillmann, Bharucha, & Bigand, 2000).

As noted above, the pop-out effect we experience on hearing a “wrong” note could arise also by virtue of the note representing an infrequent event given the current distribution of pitch events. Probe-tone judgments are also influenced by the relative frequency of occurrence of pitches in melodic sequences, that is, the strength of their representation in sensory memory (Huron & Parncutt, 1993; Leman, 2000; Parncutt & Bregman, 2000). Using melodies composed in such a way that they violated pitch distributions typical of Western tonal music, Oram and Cuddy (1995) found that tones that matched the more frequently occurring pitches of the melodic contexts were considered to fit better with the context than those pitches that occurred less frequently. Thus, multiple factors appear to influence the subjective match between an individual tone and the local context.

The purpose of the experiment described in this article was to obtain measures of the tonal congruence of individual test tones without interruptions in a relatively natural musical context. To this end, we developed a task that relied on continuous monitoring of a modulating melody for the presence of notes that were incongruous with the locally established tonality. A different continuous monitoring approach has been proposed recently in which listeners hear a continuously repeated probe tone in one ear and must continually evaluate using a slider how well the tone fits into a piece of music presented to the other ear (Krumhansl & Toiviainen, 2000). In contrast to this dichotic approach, listeners in our task are required to monitor only a single stream, thus approximating more closely a natural listening situation. We expected that our results would complement the evidence for tonal hierarchies that was obtained in the traditional probe-tone experiments cited earlier. In contrast to experiments that employ discrete trials consisting of short contexts followed by single probe notes, our participants were unaware of the concept of specific probe notes about which they were to make judgments. Rather, they were asked to evaluate continually how well notes fit the melody and to respond whenever they heard a note that seemed not to fit.

Materials and Methods

CONSTRUCTION OF THE BASIC MUSICAL MATERIAL

A continuously modulating melody (sequence) was constructed using Performer 6.03 (Mark of the Unicorn). The sequence included all 12 major keys and all 12 minor keys. It was composed so that each key context lasted for eight measures, and each of these eight bar groupings lasted approximately 19 s. The total duration of the melody was approximately 7.7 min.

The sequence contained internal modulations in order to provide a shifting harmonic context over which the task was to be performed. The modulations themselves were not the primary object of study. Rather they existed in order to provide a smooth and unobtrusive bridge from each key area to the subsequent one. The following compositional constraints were determined on the basis of this consideration. First, a progression of closely related keys should be used in order to avoid surprising harmonic shifts. Second, major and minor keys should alternate as much as possible in order to maintain consistency and symmetry throughout the sequence. Third, the modulations should not occur too quickly. In other words, the preparation, the "pivot chord," and the arrival at the new key should last at least 2 seconds in every instance. On the other hand, the modulation process should not be so slow as to obscure the harmonic context. Therefore a fourth constraint was added that at least 6 of the 8 measures belonging to a given context must unambiguously establish the key center. The remaining two measures may then be used for the modulation.

The first constraint of closely related key centers was met by referring to the circle of fifths. Our definition of a close relationship between keys requires that they occupy either identical or adjacent spots on the circle of fifths. That is, they must either share the same key signature or have a key signature differing by one flat or sharp. In order to meet the second constraint of alternating major and minor keys as well, the following two rules were devised. Major keys should be followed by minor keys having a root of a minor third below (relative minor relationship), and minor keys should be followed by major keys having a root of a perfect fifth above (dominant relationship). There were only three places in the sequence where the pattern had to be broken in order to continue the modulation through all 24 keys. The key relationships in these places, however, still qualified as close harmonic relationships by our definition. The full tonal scheme of the sequence was as follows. Note that each key letter represents 8 measures of music. Major keys are represented by upper case letters, minor keys by lower case letters, and the double bar (//) indicates the necessary breaks in the pattern:

C - a - E - c♯ - A♭ - f - // - c - G - e - B - g♯ - E♭ - // - B♭ - g - D - b - F♯ - e♭ - // - b♭ - F - d -
A - f♯ - D♭.

Two basic chord progressions were devised for the sequence. These progressions were made to be as similar as possible to one another for the sake of consistency. Each one applied to two groups containing six keys each. One progression governed the chords for the group of six keys starting with major keys (C major and B♭ major). The second progression governed the chords for the group of six keys starting with minor keys (c minor and b♭ minor). In the presentation of the progressions below, note that the letter followed by a colon (e.g., C:) indicates the tonic key at that point in the progression, and the vertical bars (|) indicate the beginnings of measures. The slash (/) indicates a pivot chord. The roman numeral symbol before this slash is the chord function in the old key, the letter and colon following the slash indicate the new tonic key, and the roman numeral symbol immediately following is the chord function in the new key (e.g., I / a: III where I is the tonic in C major). The first progression is as follows:

C: I IV | V vi | I iii | IV ii | I vi | I$^{6}_{4}$ V^{7} | I / a: III iv | ii$^{\circ 6}$ V^{7} | i iv | V VI | i III | iv
ii$^{\circ 6}$ | i VI | i$^{6}_{4}$ V^{7} | i V / E: I | ii V^{7} |

This pattern then repeats for E major – c♯ minor and then again for A♭ major – f minor. The pattern is also used for B♭ major – g minor, where it then repeats for D major – b minor and then again for F♯ major – e♭ minor. The second progression is much like the first in terms of the roman numerals used and the inversions of its chords:

cm : i iv | V VI | i III | iv ii$^{\circ 6}$ | i VI | i$^{6}_{4}$ V^{7} | i V / G: I | ii V^{7} | I IV | V vi | I iii | IV
ii | I vi | I$^{6}_{4}$ V^{7} | I / e: III iv | ii$^{\circ 6}$ V^{7} |

This pattern then repeats for e minor – B major and then again for g♯ minor – E♭ major. The pattern is also used for b♭ minor – F major where it then repeats for d minor – A major and then again for f♯ minor – D♭ major.

Because some places required a break in the pattern of key relationships, the progression had to be modified slightly for the transitions E♭-B♭ and D♭-C. The E♭ – B♭ transition required the following change to flow logically into the new key. The last two measures in E♭ were changed from E♭: I / c: III iv | ii$^{\circ 6}$ V^{7} | to E♭: I / B♭: IV I | ii V^{7}|. In order to create a continuous loop out of the sequence, the last portion of the progression in D♭ needed to be modified so that a modulation to C major occurred. This modification was accomplished as follows. The last three measures in D♭ were changed from D♭: I$^{6}_{4}$ V^{7} | I / b♭: III iv | ii$^{\circ 6}$ V^{7} | to D♭: iii IV | I ♭II^{6} / C: V^{6}/V | vii$^{\circ 7}$/V V^{7} |. The use of the Neapolitan sixth chord (♭II^{6}), the secondary dominant chord (V^{6}/V), and the secondary seventh chord (vii$^{\circ 7}$/V) were considered permissible despite their absence in the rest of the sequence since they facilitated a smooth modulation.

It was decided that the contour of the sequence should be regular so that harmonic changes within each context could be clearly perceived and were not given differential emphasis. The melody was composed with a 6/8 meter. An up-down-up contour was used for each arpeggiated chord. Each chord contained six notes: the first note, always the bass of the chord, was followed by three ascending notes, and the final two notes were a repetition of the second and third notes of the pattern (e.g., C_2, G_2, E_3, C_4, G_2, E_3). Thus each arpeggiated chord consisted of four distinct pitches.

In the aim of creating a smooth context, an effort was made to conform to the voice-leading rules of 18th-century harmony and counterpoint when connecting each of these chords to the next one in the sequence. For this purpose, the arpeggiated chords were considered to function as if all the notes sounded together. This assumption is valid because of the occurrence of subtle streaming effects that create voice-leading expectancies in the listener. There are five principal rules governing the composition of the sequences at this level. First, whenever an authentic cadence occurs (V-I or V^7-I), the leading tone of the dominant chord should ascend to the root of the tonic chord. Second, whenever a dominant seventh chord precedes a tonic chord (V^7-I), the seventh of the first chord should descend to the third of the second one. Third, whenever chords built on adjacent tones of the scale occur side by side (e.g., IV-V or iii-IV), the upper voices should move in contrary motion to the bass voice. Fourth, whenever possible, parallel fifths, octaves, and unisons should be avoided. Fifth, direct fifths and octaves should be avoided as well unless they occur in the inner voices. (Direct fifths and octaves are defined as perfect fifths and octaves that are approached in the same direction by two voices that previously formed an interval other than a perfect fifth or octave.)

CONSTRUCTION OF THE SPECIFIC STIMULUS SEQUENCES

Note that the modulating sequence just described will be referred to from this point on as the original sequence, the initial 24 measures of which are illustrated in Figure 1. The actual experiment required a slight modification. The task in the experiment focuses on the spontaneous perception of "wrong notes" in the continuously modulating context. For this experiment, three test pitches were selected: C_3, A_2, and $E\flat_3$. These pitches were chosen for two principal reasons. First, they fall within the center of the distribution of pitches for the entire sequence, and second, they correspond to the tonic elements of tonalities that are distantly related to one another (i.e., three or more positions away on the circle of fifths). For each of these probes, a separate sequence was constructed from the original looping sequence that lacked any specific instances of the test pitch (e.g., C_3). In order to accomplish this requirement, the voice-leading needed to be altered in several places. The solution for avoiding the test pitches was to replace the note with another chord member in every instance. Special care was taken not to disrupt the contour excessively as these changes were made. The test tones were then reinserted into their respective sequences according to the following two constraints. First, test tones should occur four times per tonal context (i.e., approximately once every 4.8 seconds) in order to keep listeners on task and fully attending to the sequence. Second, test tones should occur only as the third, fifth, ninth, or 11th notes of the 12 notes in each measure This requirement prevented the test tones from ever occurring as the lowest or highest note in an arpeggiated chord and therefore guarded against disruptions of the progression's bass line as well as an overly salient "pop out" effect in the upper register. It also guarded against an overly salient effect due to rhythmic prominence in the metrical framework. On average, five notes preceded the first test tone in each key. An example of how the original melody was modified to create a melody with test tones is provided by the open notes in Figure 1.

For each test pitch, two or three sequences were created that began at different positions in the original sequence. Like the original sequence, these new sequences looped through all 24 keys and lasted 7.7 minutes. The starting keys were varied in order to avoid creating a bias toward any one key (or any region of adjacent keys) on the circle of fifths. Starting keys were selected by finding those with the maximum distance to the test tone's corresponding tonality on the circle of fifths. Such starting tonalities would bias the test tones toward popping out, thereby causing listeners to respond relatively frequently at the beginning of

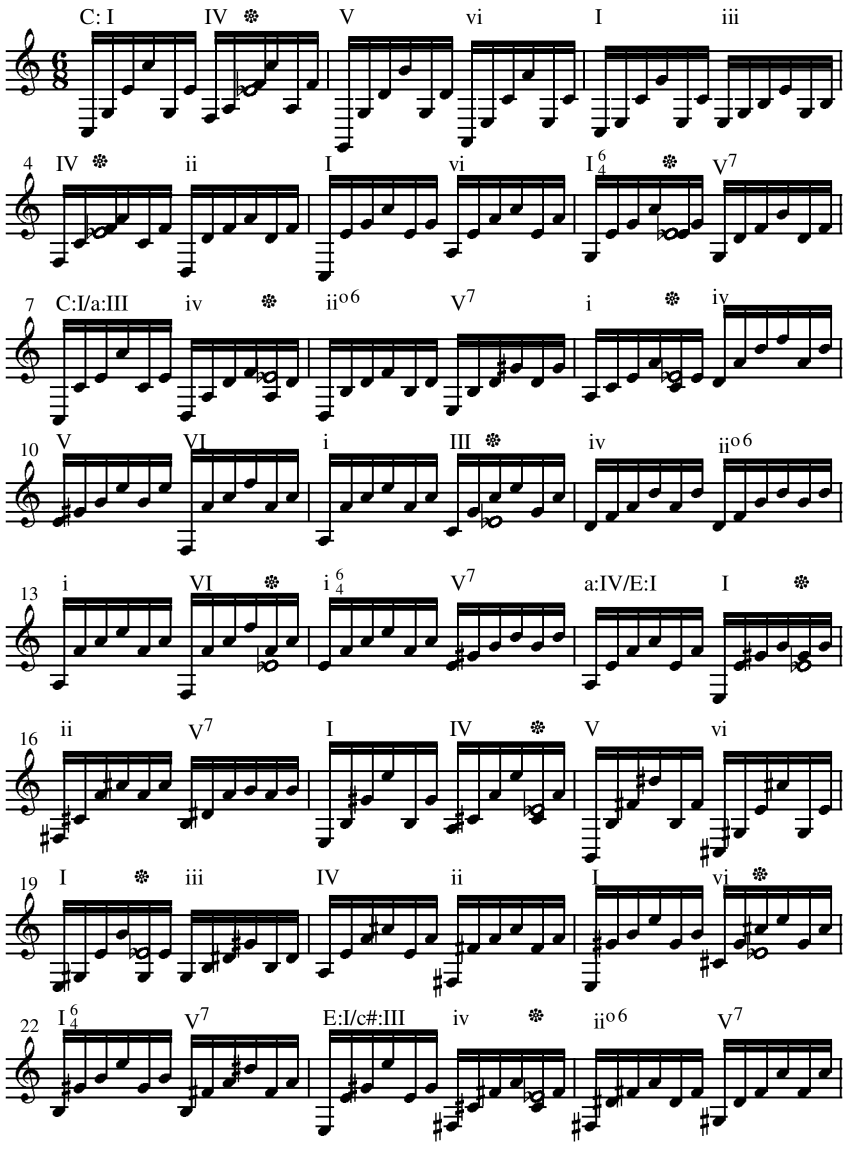

Fig. 1. Example of the original melody in the first 3 keys. The beginning of each key is denoted by the letter name of the key followed by a colon. The harmonic function of each arpeggiated chord is denoted by a Roman numeral. Capital letters and Roman numerals refer to minor chords and keys, respectively. Pivot chords that mark the transition from one key to the next are labeled with the chord function in the old key followed by a slash and the name and chord function of the new key. Multiple versions of this basic melody were used in the experiment with individual notes replaced by test tones. Asterisks mark the test tone sites for one of the melodies in which the original note was replaced with an E♭ test tone(unfilled notes). Note, that the 2nd half of measure 19 had to be rewritten to eliminate the naturally occurring instances of E♭. This modification is not shown for the sake of clarity.

the run. We felt that this bias was acceptable because salient test tones at the beginning of each run were likely to reassure listeners that they knew what type of notes they were listening for. Sequences that included the anomalous D♭-C modulation within the first 32 measures were not used, even if their starting keys were of maximum harmonic distance with respect to the probe. This resulted in a total of seven starting keys. For the C_3 test tone, the keys of G♭ major and B major were used as starting positions. For the A_2 test tone, B♭ major, E♭ major, and A♭ major were used. For the $E♭_3$ test tone, E major and D major were used.

ANALYSIS OF THE LOCAL CONTEXT OF THE TEST TONES

As mentioned earlier, the placement of test tones in the stimulus sequences was guided primarily by metrical and contour-preservation considerations and an attempt to distribute test tones somewhat evenly but unpredictably across the time window that the melody occupied in each key. The local harmonic context into which the test tone fell was not taken into account when the sequences were composed. We recognized, however, that while a probe test tone might not fit well with the global harmonic context of a particular key, it might fit reasonably well with the local harmonic context. Thus, the local context might induce a note to blend in, even though it would be expected to pop out given the key. In order to better understand the effect of local context on the perception of wrong notes, an extensive catalog of information about the test tones was created. Among the variables examined were the relationship of the test tone to the current key (diatonic or nondiatonic), relationship to the current chord (in or out of chord), and formation of a plausible seventh chord (yes or no).

Before discussing how this information was used, we clarify the codification of these variables. For major keys, any test tone that coincided with a note in the major scale of the current key was considered to be diatonic. For minor keys, any test tone that coincided with a note in the harmonic minor scale was considered diatonic, and test tones that were a minor seventh above the current tonic also were considered diatonic. All other instances of test tones were labeled as nondiatonic. Plausible seventh chords included major-minor seventh chords (i.e., dominant seventh chords) in root position, minor-minor seventh chords in root position, major-major seventh chords in root position, and fully diminished seventh chords. All other chords that included a seventh (i.e., minor-major and half-diminished seventh), were not considered to be plausible seventh chords because of their infrequency in tonal music and/or their total absence from the original sequence.

The relationships of the test tones to the local contexts is shown in Figure 2. Figure 2A illustrates, arranged by test tone type, the proportion of test tones that formed a part of the chord into which they were inserted. Two aspects of these distributions are important. First, they differ slightly for the different test tones. Second, the distributions are different for major and minor keys. The latter differences reflect the simple fact that for major keys, the test tone, (e.g., C) is a member of the diatonic scale for the five keys counterclockwise to it (F, B♭, E♭, A♭, D♭) along the circle of fifths and for the one key (G) that is clockwise to it. For minor keys (as defined above), the test tone (e.g., C), is a member of the diatonic scales of the two adjoining counterclockwise (f, b♭) and four adjoining clockwise keys (f, d, a, e), and the key situated five steps counterclockwise (c♯). Figure 2B illustrates the distributions of the proportion of test tones that formed a plausible seventh chord (as defined above) at each step along the circle of fifths. This analysis revealed an idiosyncrasy of our heuristic for distributing test tones: a test tone did not necessarily belong to the current chord more often when the test tone was the tonic of the current key. We made no attempt to modify the distributions shown in Figure 2.

PARTICIPANTS

Forty Dartmouth College undergraduate students (30 females) from the introductory psychology course served as participants in return for partial course credit. Their mean age was 19.5 years (SD, 1.2 years). Thirty-six participants reported that they were right handed,

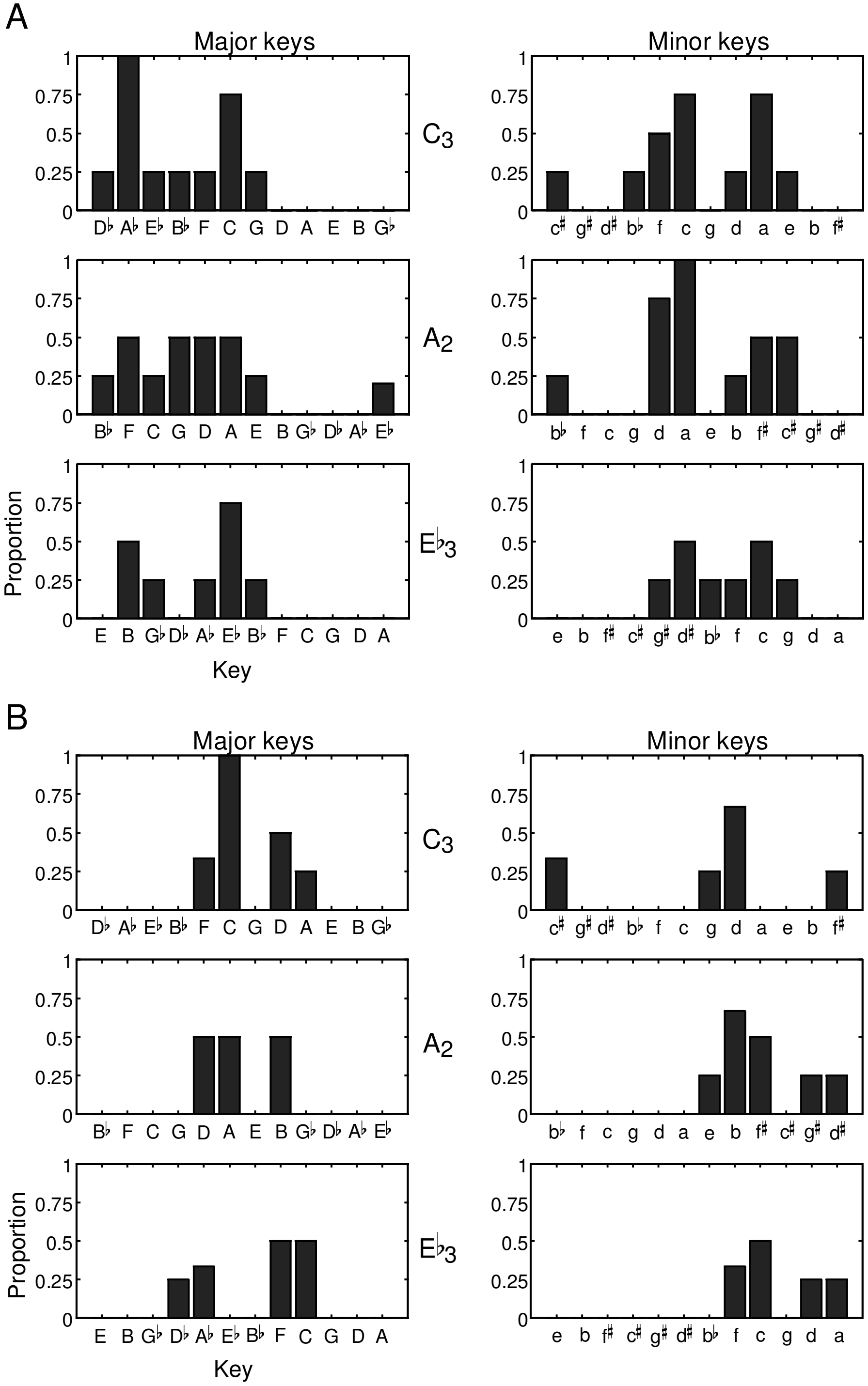

Fig. 2. Test tone statistics. (A) Proportion of test tones in each key that formed a part of the current chord. When C test tones occurred in A♭ major, they always formed a part of the current chord which would lead toward their blending into the melody. (B) Proportion of test tones in each key that could have formed a plausible seventh chord in the local context.

and one reported being ambidextrous. One participant reported possessing absolute pitch, though the veracity of this claim was not tested. Thirty-three participants reported having at least 1 year of formal musical training for voice or instrument. Among these participants, the mean amount of formal musical training was 8.3 years (SD, 4.1 years).

PROCEDURES

Performer 6.03 running on an iMac (Apple Computer) was used to play the melodies and record participants' responses. The melodies, which were stored as MIDI sequences, were rendered via the QuickTime interface using the Grand Piano sound patch at a tempo of 150 eighth notes per minute. This tempo corresponds to a note duration of 200 ms. Listeners were instructed to listen attentively to each modulating sequence and press a foot pedal every time they detected a single note that seemed "wrong." They were told to do so as quickly as possibly while still retaining a relatively high degree of confidence that it was indeed a wrong note. Wrong notes were described to the listener as sounding "out of place," "incorrect," or "possibly out of tune." The listeners heard two examples of familiar melodies ("Happy Birthday" and "Twinkle, Twinkle Little Star") with one wrong nondiatonic note in each of them. Listeners confirmed detection of the wrong notes by pressing the foot pedal and they did so with ease. A brief 30-s portion of a modulating sequence in which test tones would tend to pop out was then played as an example, and listeners were instructed to listen attentively and respond to notes they perceived to be wrong. If they failed to make any presses or expressed any confusion about the task, the example was repeated until they felt confident that they understood the task.

Each listener was then presented with a total of three modulating sequences, each one containing a different test tone (C_3, A_2, or $E\flat_3$). The order of the three test tones and the seven starting key locations was randomized across listeners. Listeners were instructed to press the pedal when they perceived that a wrong note had occurred. For the true experimental trials, they were told that there was no upper or lower limit to the number of responses they made and that it was important only that they remain concentrated and stay on task. They were told to respond if they were unsure whether a note was wrong but were leaning in the direction of thinking that it was. Listeners were instructed, however, that they should refrain from responding if more than five or six notes (about 1 s) had passed since the note in question. The listener's performance was monitored by the experimenter throughout each trial. The instructions were clarified between the first and second trials for listeners with evidently low hit rates, long self-reported delays between the notes and responses, or low self-reported confidence ratings regarding the task.

The pedal press data were recorded via a Korg digital piano (model SG-1D) and USB MIDI interface onto MIDI tracks using Performer 6.03 software. Recording the pedal press data onto tracks adjacent to the stimulus tracks facilitated online monitoring of participants' performance on the task.

DATA ANALYSIS

The pedal press MIDI data were converted to text files with time-stamp information corresponding to each pedal press and release. The time-stamp information was analyzed by using custom scripts written in Matlab (Mathworks). Several analyses were performed separately for each of the three sequences presented to a listener. First, the total number of responses occurring in each key segment (epochs of 19.2 s) was tallied. These responses included both responses to test tones ("hits") and false alarms. Second, tallies were constructed taking into consideration only the first response following a test tone. Thus, if a participant made two responses between two test tones, only the first was added to this second tally. Finally, we determined the response latency between each test tone and the first subsequent pedal press. These were then combined into a cumulative distribution of response latencies.

Because different listeners heard sequences starting in different keys, it was necessary to realign the data to a common starting position before comparing the number of responses in each key across listeners. The starting position was defined as C major. Similar data realignments were performed in order to view the response frequencies in terms of the circle of fifths or semitone distance of the test tone from the tonic of each key.

Results

REACTION TIME DISTRIBUTIONS

The cumulative distribution of reaction times across all participants following all test events for each of the three test tones (A_2, C_3, and $E\flat_3$) is shown in Figure 3. The mean reaction time was 963 ms (SD, 843 ms). The median reaction time was 703 ms. Two aspects of the distribution are important. First, 88% of the responses occurred between 300 and 1500 ms. Normally, the observation that the bulk of the responses fell into a latency

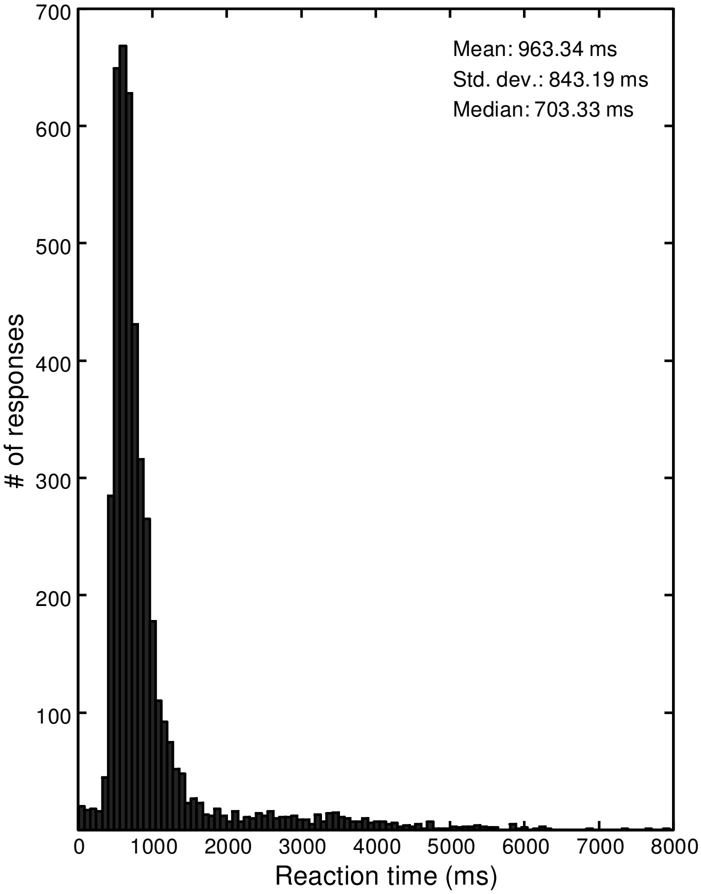

Fig. 3. Distribution of response times following all test tone events.

range that is typical for most cognitive tasks would be neither startling nor interesting. However, in our experiment, the test tones were embedded into the ongoing melody and were not presented or judged in isolation. This feature raised the concern that what we consider to be a response to a test event may not necessarily indicate a response to the test tone, but rather a response to any of several notes that occurred in a time window preceding the response. Two pieces of evidence suggest that participants were responding to the test tones.

The rather sharp peak in the distribution of reaction times provides one piece of evidence that participants were responding to a specific event rather than indiscriminately to any of several events in the same time window. Another important aspect of the distribution is the presence of a small number (<20) of responses at short latencies (<300 ms) and at longer latencies (>1500 ms). The responses at the very short and long latencies probably reflect false alarms during times around test tones that blended into the key. In other words, because all responses shown in the distribution were expressed relative to the closest preceding test event, false alarms that were made to an event following a known test tone that itself was missed show up as a long-latency response in the distribution.[1]

The second and more critical piece of evidence that participants were responding to embedded test tone events rather than other unintended features was a comparison of the number of responses that were made between 300 and 1500 ms following a test tone with the number of responses that were made during the exact same section of the melody when no test tone occurred. We were able to perform this comparison because test tone locations varied somewhat between trials that used the different test tones. For example, during the detection of C test tones, a test tone would occur in measure 57. However, during the detection of A test tones, there would be no test tone event at the corresponding location in measure 57 or during the ensuing 1500 ms. Thus the melody in measure 57 was identical between the two test tone conditions with the exception of the single C test-tone event. Consequently, if a large number of responses was made following the C test tone, but none or very few were made when the same melodic segment was heard during detection of A test tones, one could argue that responses were being made selectively when the test tone was present because the only difference was the presence of the C. This type of comparison was made for 43, 47, and 41 time windows for the C, A, and E♭ test tone sequences, respectively. We found that in these time windows, 94.5%,

---

[1] An additional reaction time distribution was computed for a hypothetical situation: responses in each run for each subject were randomly distributed across the run, and the same set of analyses was performed. In this case, only 29% of the first responses following an event fell into the window from 300 to 1500 ms following a test tone. The hypothetical distribution did not have the peaked shape of the observed distribution.

95.2% and 92.2% of the responses were made following C, A, and E♭ test tones, respectively. In other words, on average only 6% of the responses in these time windows were made when no test tone was present and could therefore be considered false alarms. This result, together with the reaction time distribution makes us confident that we observed responses to the intended test tones, even though they were embedded in an ongoing musical context.

RESPONSE LIKELIHOOD AS A FUNCTION OF TONAL CONTEXT

Although the cumulative distribution in Figure 3 indicates that participants were detecting a large number of test tone events, we were interested specifically in the number of responses observed in each key for each test tone sequence. We hypothesized that large numbers of responses would be obtained when the inserted test tone did not fit into the current key (e.g., the note C in the key of F♯ major), but that few, if any, responses would be obtained when the test tone fit well with the current key (e.g., C in F major). Figure 4 shows the average number of responses that were made during each key in the modulating melody as a function of the test tone. Low values indicate that test tones blended in and were not detected, whereas large values indicate keys in which the test tones popped out. The keys are arranged in the order that they occurred in the melody. The up and down pattern of the profiles shown in Figure 4 for each of the test tones across the 24 keys highlights that the modulating sequence was composed in a way that would result in alternating salience of any given test tone, rather than a single long period of blending in followed by a long period of salience (popping out) as might be expected if the melody followed a circle of fifths modulation.

A repeated-measures analysis of variance (ANOVA) that was performed on the profiles shown in Figure 4 showed no significant main effect of Test Tone, $F(2,78) = 2.46$, n.s., indicating that the rates of responding were not different for the different test tone trials. However, there was a significant main effect of Key, $F(23,897) = 21.18$; $p < .0001$, and most importantly, a significant Test Tone x Key interaction, $F(46,1794) = 46.27$, $p < .0001$, which indicated that the number of responses for any given key depended on the test tone that was being presented. Unless otherwise noted, all subsequent analyses were performed on all of the responses that were made, rather than only the first response following a test tone. A mean of 5.1% (SD, 1.1%) of all responses were not the first response following a test tone and could be regarded as true false alarms. These false alarms were included in the analyses to show that the profiles we obtained conformed to expectations based on theoretical considerations and prior evidence for tonal hierarchies and were therefore robust against false alarms. Profiles

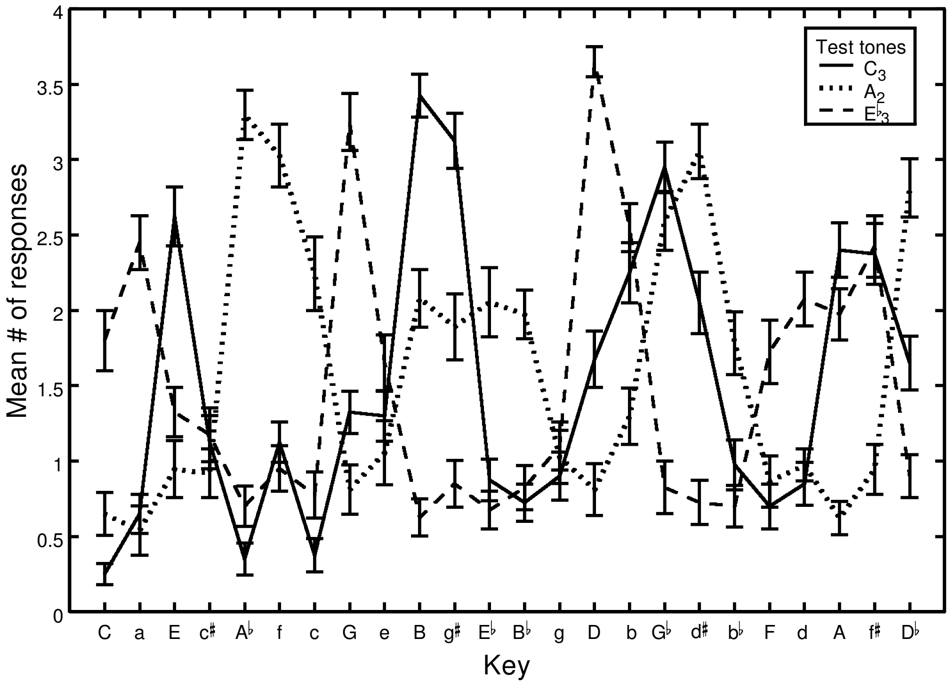

Fig. 4. Overlay of average number of responses following test tones in each key for the entire progression through the 24 major and minor keys. The keys are arranged by order of occurrence in the original melody starting with C major even though there were two possible starting keys for each of the test tones. Only the first response following a test tone was counted as a valid response. The profiles did not differ substantially from those constructed from all responses. The legend indicates which profile corresponds to each test tone. As expected, C test tones elicited few responses when the melody dwelled in C major, but elicited many responses while the melody was in B major. The error bars indicate the standard error of the mean (N = 40).

calculated with only initial responses following test tones did not differ appreciably.

TEST TONE RESPONSE PROFILES AS A FUNCTION OF POSITION IN KEY

As expected, test tones elicited fewer responses when they served important tonal functions in the current key than when they represented nondiatonic notes. Figure 5 shows the data displayed in terms of the semitone distance between the test tone and the tonic of each key. The test tones are referred to in terms of their intervallic names relative to the tonic. For example, the top panel of Figure 5A shows that few C test tones were detected while C notes served as a major third (A♭ major), but many C test tones were detected while C notes served as a minor sixth (E major). Note that the relative numbers of responses to minor and major thirds and mi-

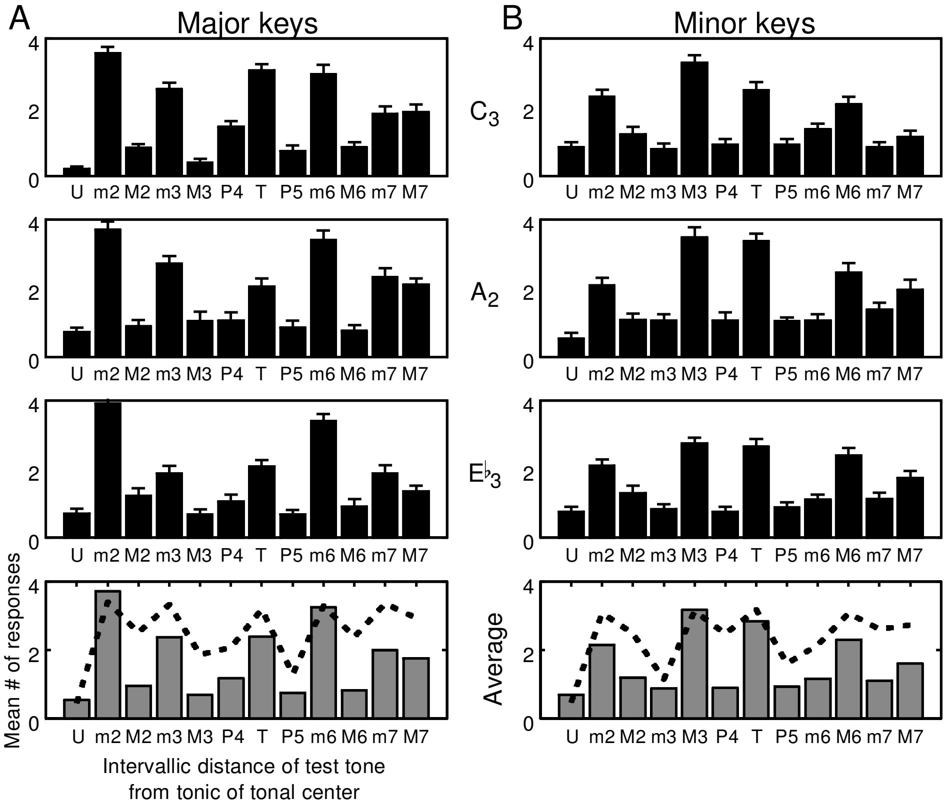

Fig. 5. Average number of responses in each key arranged by intervallic distance of the test tone from tonic of the current key. For example, P4 corresponds to G major for C test tones. The identity of test tones is indicated between sections A and B. (A) Response profiles for major keys. (B) Response profiles for minor keys. The bottom panel shows the profiles averaged across the individual test tone profiles. The dashed line represents the key profiles reported by Krumhansl and Kessler (1982), which have been inverted and scaled to facilitate direct comparison with our data. Interval labels corresponding to semitone distances of 1 to 12: U, unison; m2, minor 2nd; M2, major 2nd; m3, minor 3rd; M3, major 3rd; P4, perfect 4th; T, tritone; P5, perfect 5th; m6, minor 6th; M6, major 6th; m7, minor 7th; M7, major 7th.

nor and major sixths switched when the test tones were in major and minor keys (Figure 5A vs Figure 5B). The major-key profiles obtained using the different test tones were highly correlated (mean $r = 0.92$) as were the minor-key profiles (mean $r = 0.93$). However, repeated-measures ANOVA indicated that there were significant Test Tone × Position interactions for both major, $F(22,858) = 4.32$; $p < .0001$, and minor, $F(22,858) = 3.18$; $p < .0002$, keys. These interactions may be due to the slightly different "relatedness" profiles associated with each of the sequences (Figure 2). In other words, at any given position along the circle of fifths, the likelihood of blending into the local tonal center may have differed across the test tones and may have therefore influenced the observed profiles.

The averaged key profiles were compared with the major and minor key profiles reported by Krumhansl and Kessler (1982, as reported in Krumhansl, 1990). Because closely related tones elicit high ratings using the probe-tone method, whereas closely related tones elicit few responses using our method, we adapted the Krumhansl and Kessler profiles by subtracting them from 7 (the maximum possible value) and multiplying by 5/7 to account for the difference in the possible number of values using each method (0–4 in ours, 1–7 in the probe-tone method). The averaged major profile was significantly correlated with the Krumhansl and Kessler major key profile ($r = 0.80$; $p < .001$) as were the minor key profiles ($r = 0.76$; $p < .002$). As the superimposed Krumhansl and Kessler profiles in the bottom panels of Figure 5 show, the overall contours of the profiles matched well, which gave rise to the strong correlation.

INFLUENCE OF MUSICAL EXPERTISE

Previous research has examined the influence of experiential factors on the form of tonal hierarchies obtained using the probe tone technique. Figure 6 shows profiles, averaged across the three test tone types, that were obtained from the data of 8 participants with 1 year or less of musical training, and 8 participants with more than 10 years of musical training. Overall, the profiles of experts and novices were similar and were significantly correlated for each of test tone types: $r = 0.924, 0.833, 0.808$ for C, A, and E♭ test tones respectively ($p < .0001$ in all cases). This result replicates the findings of Cuddy and Badertscher (1987). Intriguingly, they reported a similar dissociation between listeners with high and low levels of musical training in their responses to the probe when it serves as the M3 and P4, as well as a tendency for novices to accept a m3 as fitting better into the context.

REACTION TIME PROFILES

The median response times (for the entire group of participants) to test tones as a function of the semitone distance of the test tone from the tonic of the local context are shown in Figure 7. As in the previous response time analysis (Figure 3), only the first response following a test tone was considered. The numbers under each bar correspond to the percentage of participants that responded to at least one of the four test tones that occurred for each semitone distance. In general, median response times increased and the number of participants responding decreased for test tones that were members of the diatonic scale. The lengthening of response times for dia-

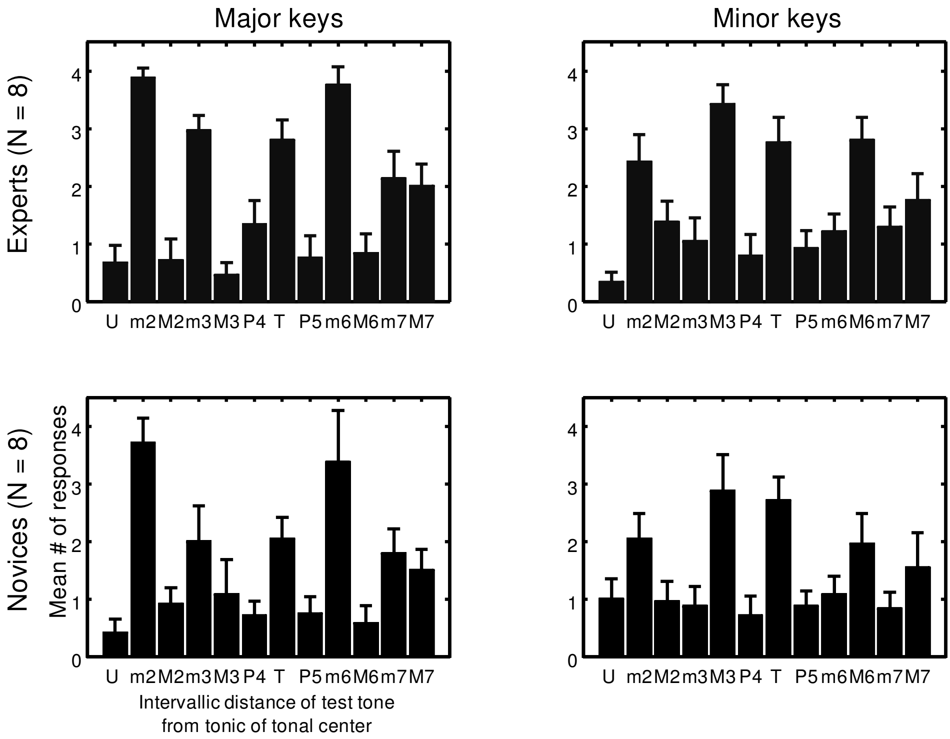

Fig. 6. Number of responses in each key arranged by intervallic distance of the test tone from tonic of the current key, and averaged separately for a subgroup (N = 8) of "experts" with more than 10 years of musical training and a subgroup (N = 8) of "novices" with 0–1 years of musical training. See Figure 5 for an explanation of the interval labels.

tonic tones was even more pronounced when mean response times were used in the analysis (data not shown). The increased response times are evidence that it took longer to decide that a test tone did not fit into the ongoing context when the test tone was more closely related to the ongoing context, than it did to reach the same decision when the test tone was unrelated to the context. One must remember, however, that some of the responses following diatonic test tones may reflect false alarms to other notes in the melody rather than responses to the test tones per se.

CIRCLE OF FIFTHS

Another way of displaying the data is in terms of the distance around the circle of fifths by which each key in the modulating sequence is separated from the key that the test tone is the tonic of (Figure 8). In this representation, it becomes evident that test tones blend in better with keys that are situated counterclockwise (negative axis labels; see Figure 2 for key labels) from the test tone's key when major keys are considered. For minor keys, there is a

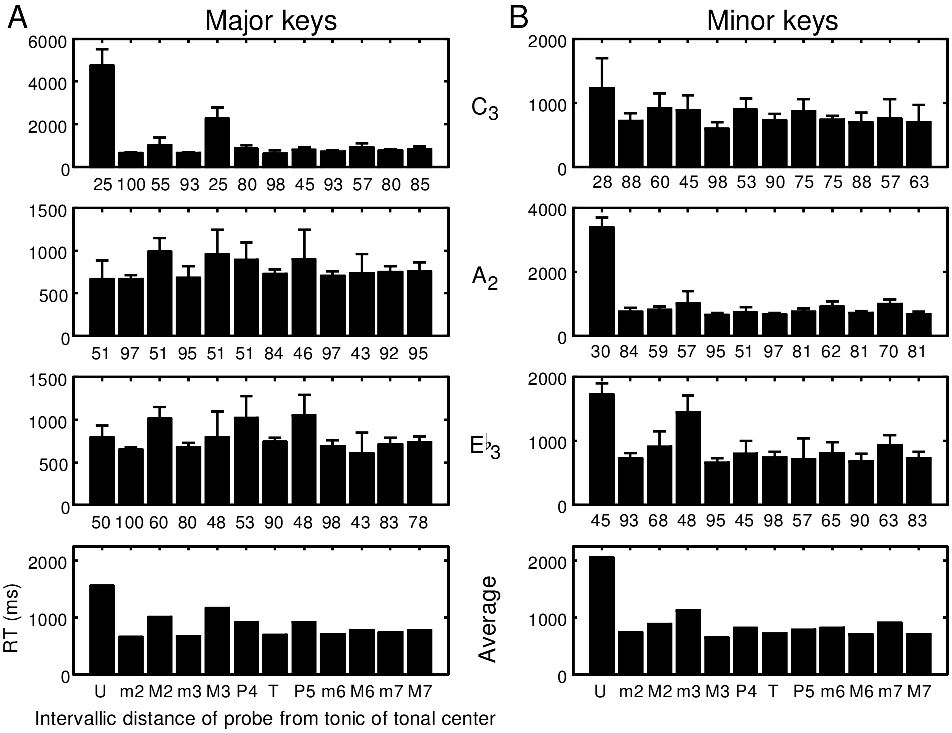

Fig. 7. Average median reaction times to test tones in each key arranged by intervallic distance of the test tone from the tonic of the current key. Test tones are indicated between sections A and B. (A) Response profiles for major keys. (B) Response profiles for minor keys. The numbers below each bar indicate the percentage of participants who responded at least once while the melody was in each key. Error bars indicate the standard error of the median reaction time. The number of participants used in the standard error calculation was the number of participants for which a median response time was available. The bottom panels show the average response time profile. Overall, when the test tone served as a nondiatonic tone, it was responded to more quickly and by a larger number of participants than when it served as a diatonic tone. The long average median response times to diatonic tones may reflect false alarms rather than responses to actual test tones.

greater bias in the clockwise direction for the test tone to blend into the current key. These effects are readily explained if one considers in which keys the test tone is a member of the diatonic scale. For example, an A_2 test tone is a diatonic member of five major keys (B♭, F, C, G, D) that are situated counterclockwise to A and one major key (E) that is situated clockwise.

Discussion

In this study, we investigated spontaneous responses following deviations from tonal contexts in a continuous melody that modulated through

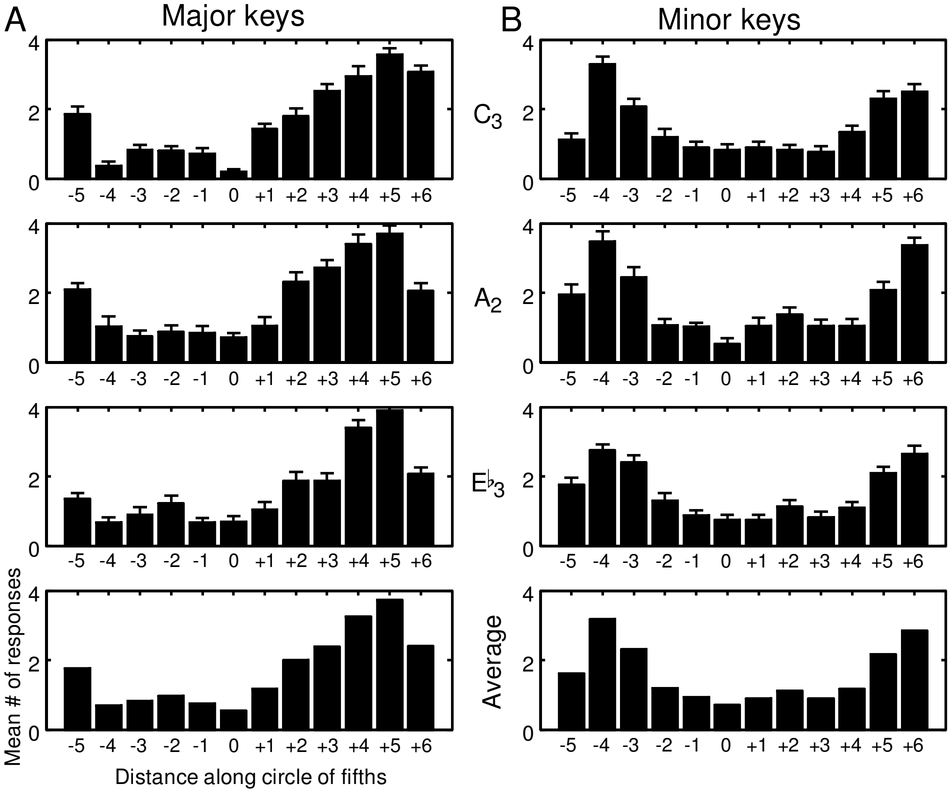

Fig. 8. Number of responses per key arranged by distance around the circle of fifths. The distance is the number of steps of the current key from the key of which the test tone would be considered the tonic. For example, a distance of -2 corresponds to B♭ major for C_3 test tones, and G major for A_2 test tones. (A) Profile for major keys. (B) Profile for minor keys. Error bars reflect the standard error of the mean (N = 40).

all 24 major and minor keys. The basic melody was modified slightly to accommodate each of three different test tones. A test tone was presented four times during the time the melody dwelled in each key. Thus, the test tones were equiprobable in all keys. Participants were not instructed to detect the test tones; rather they were asked to respond whenever they heard a note that sounded "wrong" or "out-of-place." Our results indicate that the contours of profiles generated from spontaneous responses following test tones embedded in an ongoing melody closely match the contours of profiles obtained using the traditional probe-note judgment method (Figure 5). In other words, participants responded when test tones were contextually distant from the current key, but were less likely to respond when test tones fit well with the current key. As noted in the introduction, judgments of how well any given note or chord fit into the preceding context could be influenced by long-term memory for tonal structures (Krumhansl, 1990) or short-term memory for the pitch chroma distribution in some time window preceding the event of interest (Huron & Parncutt, 1993; Leman, 2000; Oram & Cuddy, 1995; Parncutt & Bregman, 2000). We now discuss in turn the relationship of our measured profiles to profiles representing the canonical tonal hierarchy and the pitch distributions of our stimulus.

Important features of the canonical Krumhansl and Kessler (1982) tonal hierarchy were preserved: fewer responses were made following test tones that functioned as a member of the tonic triad, more were elicited following other diatonic tones, and the most were elicited following nondiatonic tones. However, the relative magnitudes of the components of the profiles obtained using the two methods were somewhat different. Overall, test tones in our study tended not to elicit a response even though they may receive fairly low relatedness ratings using the traditional probe-tone method (e.g., when the test tone is in a major-second relationship to the tonic). This discrepancy may be due in part to the number of response options afforded by each method. In the traditional probe-tone method, participants are able to respond on a seven-point scale whereas using the wrong-note detection method, participants have only two choices. When more response options are available, it is possible to assign ambiguous events an intermediate relatedness value. However, when a binary decision must be made, whether an ambiguous event is tagged as a wrong note depends in part on the response criterion maintained by the listener. It is possible that this response criterion is shaped by musical factors, such as how quickly a deviant note is melodically anchored, that operate in a real musical context but not in a probe-note judgment setting in which the probe note is the final musical event.

Our profiles could also depend on the short-term pitch distribution statistics of the melody we used. Figure 9 shows that the average pitch height of notes used while the melody was in each key varied across keys. We calculated the correlation of the vector of average pitch heights with a vector specifying distance along the circle of fifths and found them to be uncorrelated ($r = 0.008$), indicating that average pitch height did not predict location on the circle of fifths. More salient is the question of the influence of representations of the test pitch in sensory memory. For example, if middle C is the test tone, one might expect a C to pop out if it has not been heard recently, as would be the case while the melody dwells in B major. Although we eliminated all instances of the test pitch from the melody so that the distribution of test tones upon reinsertion would be uniform, we did not eliminate all instances of octave equivalents of the test pitches. To the extent that participants are implicitly tracking distributions of pitch chroma instead of or in addition to pitch height, octave equivalents of the

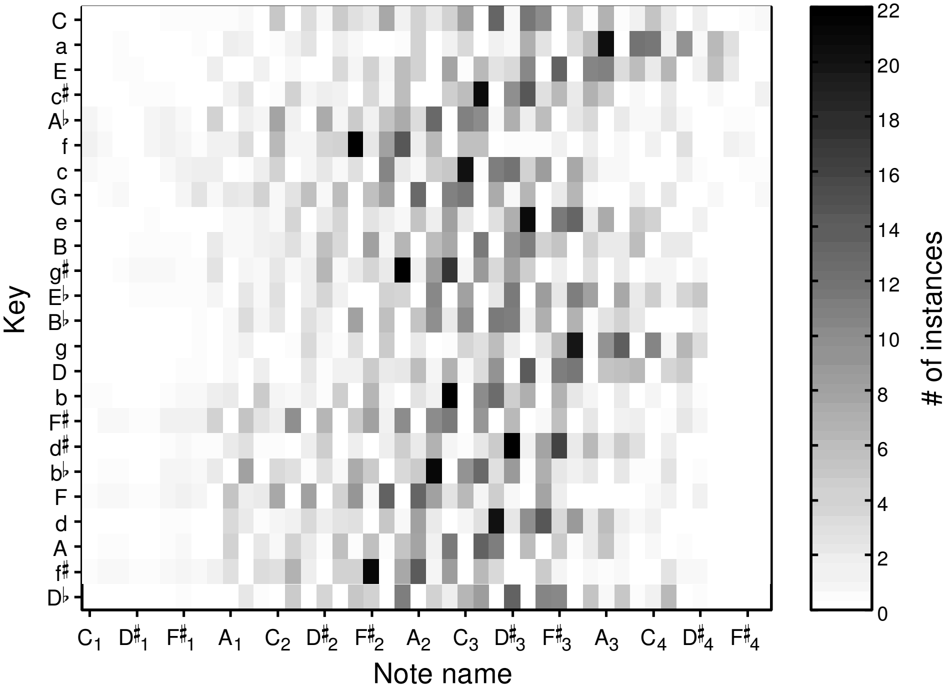

Fig. 9. Distributions of pitches used in the original melody while it dwelled in each key.

test pitch could be expected to influence the results. To assess the degree to which the pitch distribution characteristics might be responsible for the response patterns we observed, we tallied the number of times that octave equivalents of the test pitch occurred in each key. The resulting profile was then correlated, separately for major and minor keys, with the Krumhansl and Kessler profiles (in their original orientation) and our observed average response profile. The correlations are given in Table 1.

TABLE 1
Correlations Among Key Profiles and the Pitch Distribution of the Stimulus Melody

	Response Profile	Krumhansl and Kessler Profile	Pitch Distribution
Response profile	1.000	-0.805	-0.877
Krumhansl and Kessler profile	-0.765	1.000	0.930
Pitch distribution	-0.772	0.932	1.000

NOTE—Major key profile correlations are given in the upper right triangle. Minor key profile correlations are given in the lower left triangle.

The pitch distribution profiles were correlated positively and strongly with the Krumhansl and Kessler major and minor profiles. The correlations of the pitch distribution profiles with the response profiles were weaker. As expected, the correlation was negative. In other words, the more likely the occurrence of an octave equivalent of the test tone, the smaller the likelihood that the test tone would elicit a response. In the case of the minor key profiles, the Krumhansl and Kessler and pitch distribution profiles were correlated equally strongly with the response profile. For the major key profiles, the correlation of the response profile with the pitch distribution was greater than the correlation with the Krumhansl and Kessler profile. An interesting aspect of the correlation matrix was that the pitch distributions were more predictive of the Krumhansl and Kessler hierarchies than of our profiles derived from the subjective impressions of notes popping out of the local context. The very strong correlation of the "sensory" and "cognitive" profiles highlights the difficulty in teasing apart sensory and cognitive contributions to judgments of tonal membership in real musical contexts (cf. Cuddy & Thompson, 1992). The weaker correlations of the response profiles with either of the reference profiles also indicate that factors other than the statistics of the pitch distributions in each key were influencing listeners' responses. For example, as mentioned earlier, the deviation of the observed response profiles from the other profiles might be due to the degree to which the test tones could be bound into the local chord or anchored by the notes of subsequent chords.

Although our experiment did not enable us to determine the extent to which short-term and long-term memory influence the moment-to-moment evaluation of notes in the melody, it establishes an alternative method for confirming that participants' percepts of individual notes are sensitive to the momentary tonal center even as the tonal center changes from key to key over the course of a modulating melody. Our findings support previous research showing that musically trained and untrained listeners readily perceive modulations in both harmonic (Cuddy & Thompson, 1992; Krumhansl & Kessler, 1982) and melodic (Thompson & Cuddy, 1989, 1992) sequences.

One important aspect of our task is that it assesses listeners' percepts without interrupting the melody in order to have them make a judgment about a probe note. This property has enabled us to use this task and stimulus set to identify regions of the brain that maintain a representation of the distance relationships among the major and minor keys (Janata, Birk, Van Horn, Leman, Tillmann, & Bharucha, 2002). Our task is akin to other continuous monitoring tasks that have been used to investigate properties of tonal contexts. Berent and Perfetti (1993) showed that performance on a continuous click-detection task was impaired when sudden distant modulations occurred in a simultaneously presented harmonic sequence, but returned to normal as the new harmonic context was established. Krumhansl (1996) showed that the multifaceted percept of musical tension could be quantified by monitoring the position of a slider that was continually adjusted by subjects as they listened to excerpts of a Mozart piano sonata. The observed measure of tension correlated well with predicted locations of high and low tension that were based on theoretical analyses of the score. The method of continuous adjustment has also been used to generate probe-tone profiles that are then projected onto the surface of a torus to determine a listener's perceived tonal location (Krumhansl & Toiviainen, 2000). Although attractive because it provides a continuous measure of perceived assimilation, this method can result in long experiments because it requires listening to the same piece of music 12 times in order to obtain continuous ratings for each pitch class.

Our test-tone method could be extended to investigate melodic anchoring (Bharucha, 1984, 1996). For example, nondiatonic test tones in a segment of a continuous melody could be anchored in one version and left unanchored in another so that one could assess the likelihood of obtaining a response from the listener. The concomitant decreases in response likelihood and increases in reaction times that we observed when test tones were members of the current diatonic set corroborate numerous reports by participants that they would sometimes wait to see whether a note that sounded strange would make sense in the context of the next few notes. Thus, if a note is ultimately judged to be adequately anchored, the response likelihood will decrease. However, if the anchoring of a test tone is insufficient, a response will be made. The latency of the response then provides a measure of how salient the contextual violation was. Such measurements may provide a perspective on the temporal window and short-term memory processes that govern attentive listening to music.[2]

References

Berent, I., & Perfetti, C. A. (1993). An on-line method in studying music parsing. *Cognition, 46*, 203–222.

Bharucha, J. J. (1984). Anchoring effects in music: The resolution of dissonance. *Cognitive Psychology, 16*, 485–518.

Bharucha, J. J. (1996). Melodic anchoring. *Music Perception, 13*, 383–400.

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

Cuddy, L. L., & Thompson, W. F. (1992). Asymmetry of perceived key movement in chorale sequences: Converging evidence from a probe-tone analysis. *Psychological Research, 54*(2), 51–59.

Francès, R. (1958). *La perception de la musique.* Paris: Vrin.

---

[2] We thank Brent Jones for assistance with data acquisition. This work was supported by NIH grant P50 NS17778-18.

Halpern, A. R., Kwak, S., Bartlett, J. C., & Dowling, W. J. (1996). Effects of aging and musical experience on the representation of tonal hierarchies. *Psychology and Aging, 11*, 235–246.

Hebert, S., Peretz, I., & Gagnon, L. (1995). Perceiving the tonal ending of tune excerpts: the roles of pre-existing representation and musical expertise. *Canadian ournal of Experimental Psychology, 49*, 193–209.

Huron, D., & Parncutt, R. (1993). An improved model of tonality perception incorporating pitch salience and echoic memory. *Psychomusicology, 12*, 154–171.

Janata, P., Birk, J. L., Van Horn, J. D., Leman, M., Tillmann, B., & Bharucha, J. J. (2002). The cortical topography of tonal structures underlying Western music. *Science, 298,* 2167–2170.

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

Krumhansl, C. L. (1990). *Cognitive foundations of musical pitch.* New York: Oxford University Press.

Krumhansl, C. L. (1996). A perceptual analysis of Mozart's Piano Sonata K.282: Segmentation, tension, and musical ideas. *Music Perception, 13*, 401–432.

Krumhansl, C. L., & Kessler, E. J. (1982). Tracing the dynamic changes in 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.

Krumhansl, C. L., & Toiviainen, P. (2000). *Dynamics of tonality induction: A new method and a new model.* Paper presented at the 6th International Conference on Music Perception and Cognition, Keele, United Kingdom.

Leman, M. (2000). An auditory model of the role of short-term memory in probe-tone ratings. *Music Perception, 17*, 481–509.

Oram, N., & Cuddy, L. L. (1995). Responsiveness of Western adults to pitch-distributional information in melodic sequences. *Psychological Research, 57*, 103–118.

Parncutt, R., & Bregman, A. S. (2000). Tone profiles following short chord progressions: Top-down or bottom-up? *Music Perception, 18*, 25–57.

Thompson, W. F., & Cuddy, L. L. (1989). Sensitivity to key change in chorale sequences: A comparison of single voices and 4-voice harmony. *Music Perception, 7*, 151–168.

Thompson, W. F., & Cuddy, L. L. (1992). Perceived key movement in 4-voice harmony and single voices. *Music Perception, 9*, 427–438.

Tillmann, B., Bharucha, J. J., & Bigand, E. (2000). Implicit learning of tonality: a self-organizing approach. *Psychological Review, 107*, 885–913.