

# The detection and tracing of melodic key changes

ANTHONY J. BISHARA

Washington University, St. Louis, Missouri

and

GABRIEL A. RADVANSKY

University of Notre Dame, Notre Dame, Indiana

Adjusting to key changes in music is vital to a listener's understanding of it, although it is unclear when this adjustment occurs. In three experiments, participants heard obscure melodies that sometimes contained key changes. Both the number of notes after the change and the change distance were manipulated. For one task, the participants were explicitly asked to *detect* a change. For another task, the participants rated probe tones. These ratings allowed us to *trace* the listeners' multiple representations of keys. Results suggest that old keys were abruptly abandoned but that new keys were accepted gradually. More distant key changes (e.g., C–F#) were associated with more accurate detection of change, as well as with more time being required to adapt to the new key.

A musical key provides a framework or schema for the set of pitches that make up a passage of music. It is common for a single piece of music to have one or more key changes. Despite this, listeners readily comprehend the music, and it seems intelligible almost immediately after such a change. Detecting and adjusting to a new key appears to be a relatively rapid process. However, it is unclear exactly when this process occurs. "Situations may arise . . . where it is difficult to fix precisely the moment of change: we want to insist on the [old key], while a new key has already taken over: When did it make its appearance?" (Schenker, 1906/1954, p. 323). Because most Western music has key changes, adjusting to a new key is a vital and integral part of listening to music. The aim of the present work is to determine when this adjustment begins. Specifically, when do listeners detect that there has been a shift to a new key, and how is this affected by the nature of the key change? Can we "fix precisely the moment of change," or do listeners gradually detect key changes?

Historically, there have been two seemingly conflicting approaches to the treatment of key changes. On the one hand, some treatments of key have assumed a short or even nonexistent transition between keys. It is common in the pedagogy of music theory to label an exact

point at which a new key begins. If there is any ambiguity in the boundary between keys, this is often limited to one and only one chord (Aldwell & Schacter, 1989; Piston, 1962). Many recent efforts to develop key algorithms involve a similar assumption about the abruptness of change (e.g., Shmulevich & Yli-Harja, 2000; Temperley, 1999; for a review, see Krumhansl, 1990). These algorithms are an attempt to identify the exact boundary between an old and a new key. On the other hand, key changes can be thought of as dynamic and unfolding over time. Even when a certain key is abandoned, the listener may have to gradually accumulate more and more information to find a new schema with which to make sense of the music (Krumhansl & Kessler, 1982).

Whether the processing of key changes occurs abruptly or gradually may depend on how the question is asked. For instance, one can ask listeners directly whether they have *detected* a change in the music. This approach is akin to some paradigms within the domain of visual perception, where attention is necessary for an accurate detection of change (Rensink, 2002; Rensink, O'Regan, & Clark, 1997). In the visual domain, attention is most important within the first 80 msec after a change (Watanabe, 2003), leaving a narrow window of time for change detection to occur. A similar situation may arise with music, with listeners noticing key changes relatively rapidly, or not all. Such a pattern of results would be consistent with the treatment of key changes as abrupt occurrences.

Alternatively, one can also use a more indirect approach to *trace* listeners' representations of key across time. In this approach, a key is treated as a hierarchy of pitches that can be described as a tonal profile (Krumhansl & Shepard, 1979). A tonal profile indicates how well each pitch belongs to, or is prototypical of, a particular key. Tonal profiles for keys have been derived from probe tone studies in which listeners hear a scale, a single chord, or a

---

Experiment 1 was conducted as part of the requirements for the senior honors thesis class at the University of Notre Dame. This research was supported in part by Grant ARMY-DASW01-99-K-0001 from the Army Research Institute, awarded to the second author. We thank David Copeland for assistance with data collection, Thomas Merluzzi, Emily Skow, Jeff Zacks, and three anonymous reviewers for helpful comments on earlier drafts, Martha Storandt and Michael Strube for statistical advice, and Diane Berish and Sarah Wieber for creating numerous MIDI files. Correspondence concerning this article should be sent to A. J. Bishara, Washington University, Department of Psychology, St. Louis, MO 63130-4899 (e-mail: ajbishar@artsci.wustl.edu).

cadential chord progression and then rate how well a probe tone fits with that stimulus. The set of ratings for the 12 pitch classes within a key define the key's tonal profile, and this profile is relatively consistent across a variety of key-defining stimuli (Krumhansl & Kessler, 1982). This consistency justifies the use of a tonal profile as a basis for comparison, as a way to trace listeners' representation of key when the key is changing or otherwise ambiguous.

Krumhansl and Kessler (1982) used this method to trace representations of key after a key change. In their work, chord progressions with key changes were played and stopped at various points so that probe tones could be rated. Listeners' ratings were correlated with previously established key profiles so as to estimate the strength of key representations for both the old and the new keys across time. Krumhansl and Kessler found that listeners took a good deal of time to adapt to the change by abandoning the old key and gaining a sense of the new key. However, the pattern of results depended on the psychological distance between the keys.

It is well known that some key changes are perceived to be more distant than others (e.g., C to F# is more distant than C to G). Krumhansl and Kessler (1982) developed a model to quantify these distances by using the intercorrelations among the profiles of all major and minor keys. They generated a four-dimensional spatial map of key relations, known as the *toroidal* model. This model has provided a good description of the psychological distance between keys (Shepard, 1982; Thompson, 1993; Thompson & Cuddy, 1997), and it converges with traditional conceptions of key distance derived from music theory (Schoenberg, 1954). For changes to close keys as defined by the toroidal model, Krumhansl and Kessler (1982) observed that listeners tended to maintain some sense of the first key, while quickly gaining a representation of the new key. In contrast, for changes to distant keys, the old key was quickly abandoned, and the new one was derived over a longer period of time. Unfortunately, the available evidence only weakly supported an effect of key distance. Conclusions about the time course of key adaptation were not based on statistical hypothesis testing, but on a narrative description of results averaged across participants.

Although key distance's relation to key tracing may be complex, its impact on the detection of change ought to be more straightforward. Distant keys are generally perceived to be more dissimilar from one another. For example, it is harder to recognize a transposition of a melody as being the same as the original when it has been transposed to a distant key, rather than to a closer one (Bartlett & Dowling, 1980; Cuddy, Cohen, & Miller, 1979). Because distant transpositions are seen as more "different" from the original, distant key changes may be easier to detect.

To get a broader picture of responses to key change, our experiments involved both key change detection and key-tracing tasks. Results from these tasks can be used to address two separate *when* questions about adjustment to key changes. The first question is, when do listeners

detect a change in key? This was assessed using a discriminability measure on the key change detection task. The second question is, when do listeners gain a stable representation of the new key? This was assessed using probe tone ratings so as to trace the strength of multiple representations of a key. We also examined the impact of key change distance on both tasks. Finally, although not central to our hypotheses, musical training was used as a covariate in these analyses to assess its possible effect on key change perception.

Importantly, the two tasks need not reveal the same time courses. Detecting that the old key is no longer relevant may or may not occur simultaneously with the activation of the new key. On the one hand, converging time courses across tasks would be in line with depictions of key change in standard roman numeral notation and key-finding algorithms, in which one key is almost immediately replaced by another. These depictions imply (perhaps unintentionally) that key representation is necessary and exclusive.

Alternatively, if time courses diverge across tasks, it would suggest at least partial independence among active representations of key. At least two pieces of evidence favor the latter possibility. First, listeners are able to hold multiple active representations of key when listening to some types of polytonal music (Thompson & Mor, 1992), suggesting that key representation is not exclusive. Second, previous attempts to map key representation during key changes have shown that the listeners' sense of key sometimes jumps around wildly along the toroidal map (Krumhansl & Kessler, 1982), as if there were no real sense of key for some period of time. In other words, key representations at any given point in the music need be neither exclusive nor necessary, and this may become apparent if change detection occurs before the new key is strongly activated.

In many studies in which the probe tone method has been used, the same stimulus has been played repeatedly for each participant. However, some researchers have worried that repetition can artificially enhance tonality (Berent & Perfetti, 1993). To avoid this potential problem, in our experiments, we used a large battery of novel musical stimuli without repetition for any given participant. For the sake of simplicity, the stimuli were melodies without harmonic accompaniment. It is important to note that a melody alone can convey key change information. Thompson and Cuddy (1989) showed that listeners judge key change distance as effectively with melody-only as with harmonized stimuli. Even untrained listeners are capable of identifying the key change distance for single voices (Thompson & Cuddy, 1992). Furthermore, there are situations in which listeners identify the distance of key changes more accurately in melody-only than in harmonized music, such as when music lacks the expressiveness of a human performance (slight changes in tempo and loudness across a piece; Thompson & Cuddy, 1997).

Key changes in the present experiments were created by transposing the endings of melodic stimuli. Although this method of modulation has the disadvantage of oc-

curring rarely in everyday music, it has advantages for experimental purposes. It provides a natural and clear-cut comparison condition—that is, the same sequence without a transposition. It also avoids the interpretive difficulties associated with pivot chord key changes, where the change is prepared by one or more chords that belong to both the old and the new keys. Pivot chords make it hard to objectively define the point of change.

### EXPERIMENT 1

We expected listeners to detect change early on, or not at all. This means that discriminability on the key change detection task should peak within the first few notes after the change. Changes to distant keys should be easier to detect than changes to closely related keys. Because the first experiment focused on the detection of change, only two probe tones were used for the key-tracing task. This approach provides a rough estimate of the waxing and waning of key representations after key changes have occurred. When the probe tone pitch matches the ending key of the piece, it ought to be given a high rating, indicating that it “fit” with the music. The rating should become gradually higher as more notes are played in the new key.

#### Method

**Listeners.** Eighty-four undergraduates from the University of Notre Dame subject pool participated in the experiment. They had 0–17 years ( $M = 5.04$ ,  $SD = 4.59$ ) of formal training on an instrument or with voice. All the listeners reported normal hearing, and none reported having absolute pitch or previous contact with the source of the stimuli.

**Apparatus and Stimuli.** Ninety-one obscure folk tunes and melodies from classical music were used from *Music for Sight Singing* (Ottman, 1996), with 84 serving as critical items and 7 as practice/example items. All the melodies consisted of two- to four-measure excerpts, with a phrase ending followed by the first seven notes of the next phrase. Adjacent note repetitions at the same pitch

were considered one note. The first and last phrases always began on the same beat. Some melodies were altered slightly to preserve this feature. The starting keys of the melodies were randomized. Because the tonal hierarchies of minor keys are complicated by multiple variations of the minor scale, only melodies in major modes were selected, and only modulations to major keys were used. Melodies with nondiatonic pitches (accidentals) were excluded. Figure 1 shows an example of a melody and some manipulations of it that were used (see the appendix for information on the distribution of tones in melodies).

MIDI files were prepared with precise onset-to-onset timing, and the loudness of each note was held at a constant scale velocity. The tempo was fixed at 120 beats per minute for all the melodies. Tones were produced by a computer sound card with an OPL3 FM synthesizer. The timbre was fixed on “acoustic grand piano” from the General MIDI set.

**Procedure.** A musical background questionnaire was given to assess formal training and to ensure that no listeners had had prior experience with *Music for Sight Singing* (Ottman, 1996). The listeners were then seated in separate rooms and fitted with headphones. They were given an opportunity to adjust the volume level before the testing began. Forced-choice responses were entered on a two-button mouse, and rating responses were entered on the numerical keypad of a PC keyboard.

The concept of *key change* was explained, and three labeled examples were given. Each example was played first without and then with a key change. Afterward, each listener was exposed to 4 practice melodies and then to 84 critical melodies. The listeners were not informed that the first 4 were for practice. No melody was ever played twice for a given listener. The order of the melodies was randomized. A random set of half of the melodies modulated downward, and the other half upward. The direction of key changes was not expected to influence the results, because key change direction in music without harmonic accompaniment does not significantly influence judgments of key distance (Thompson & Cuddy, 1989, 1992).

For each participant, half of the melodies were randomly assigned to the *no key change* condition. The remaining *key change* melodies were randomly assigned to modulate two, four, or six half-steps up or down, resulting in changes to the keys of II, III, and #IV for upward shifts and to ♭VII, ♭VI, and #IV for downward shifts, respectively. According to the toroidal model (Krumhansl & Kessler, 1982), these three transpositions (II/♭VII, III/♭VI, and #IV) provide

No Shift. 7 Notes.

No Shift. 6 Notes.

6 Half-Step Shift Upward. 6 Notes.

**Figure 1.** An example of a critical stimulus used for all the experiments. The top panel shows the stimulus with seven notes in the last phrase (adjacent note repetitions are not counted). The middle panel has six notes. The bottom panel also has six notes, and the last phrase is transposed up six half-steps from the key of G to C# major.

three ordinal levels of distance between keys, going from least distant (II/VII) to most distant (#IV). In addition, each melody was randomly assigned to stop one to seven notes after the critical point, which was always the beginning of the last phrase.

For the recognition task, the participants answered either *yes* or *no* to the question "Was there a key change?" by pressing either the left or the right mouse button, respectively. After a 1-sec delay, a probe tone was played for 1 sec. The same timbre was used for the probe tones as for the melodies. The probe tone was one of two pitches. For *matching key* trials, the probe tone was the tonic of the final key. For *mismatching key* trials, the probe tone was the tonic of another key. Specifically, in the key change conditions, it was the tonic of the old key. In the no key change condition, it was the tonic of one of the possible second keys, even though no key change had occurred. We chose the tonic as our point of reference because it is perceived to be the most prototypical pitch of a key (Krumhansl & Shepard, 1979). The listeners rated how well the tone belonged to the end of the melody in a musical sense on a scale from 1 to 7, with 7 representing *fits well*, 1 representing *fits poorly*, and 4 representing *unsure*. The listeners were encouraged to use the full range of the scale. The "enter" key was pressed to move on to the next melody.

**Design and Analysis.** Assignment of melody to type of key change (including the no-change condition), number of notes, and probe pitch was rotated across participants. There were two primary dependent measures. These were (1) accuracy on the key change detection task and (2) probe tone ratings. In addition to these two primary measures, years of formal training was examined as the covariate.

For the key change detection task,  $A'$  scores were calculated (following Donaldson, 1992) for each individual participant, using hits in the key change conditions and false alarms in the no key change condition. To correct for ceiling and floor effects, .5 was added to all hit and false alarm frequencies. The result was divided by the number of trials + 1 for calculation of hit and false alarm rates (Snodgrass & Corwin, 1988).  $A'$  is a nonparametric signal detection measure of discriminability. A score of .5 represents chance performance, and a score of 1 indicates perfect discriminability.

## Results

**Detection task.** False alarm rates in the no key change condition were uniformly low ( $M = .19$ ,  $SD = .13$ ). When these data were analyzed alone, the effect of the number of notes was not significant ( $F < 1$ ). Thus, at no point were the listeners thinking that a key change had occurred.

Mean  $A'$  values for the key change conditions are displayed in Table 1. The data were first submitted to a 3 (key change: two, four, or six half-steps)  $\times$  7 (number of notes: one to seven) repeated measures analysis of variance (ANOVA). Distant key changes were more detectable than close key changes, as revealed by a significant main effect of key change [ $F(2,166) = 91.51$ ,  $MS_e = 0.020$ ,  $p < .001$ ]. Detectability did not generally increase with more notes, as the main effect of number of notes was not significant [ $F(6,498) = 1.05$ ,  $MS_e = 0.045$ ,  $p = .39$ ].

However, there was a subtle increase in change detection. It was largely confined to the first two notes, and particularly in the two and four half-step shift conditions. The key change was great enough in the six half-step condition to immediately signal a key change. This was confirmed by a significant key change  $\times$  number of notes interaction [ $F(12,996) = 2.20$ ,  $MS_e = 0.016$ ,  $p < .05$ ]. Furthermore, when the data were analyzed with the one-note conditioned removed, whereas the main effect of key change remained significant [ $F(2,166) = 76.67$ ,

$MS_e = 0.017$ ,  $p < .001$ ], the main effect of number of notes and the interaction were not [ $F < 1$  and  $F(10,830) = 1.39$ ,  $MS_e = 0.016$ ,  $p = .18$ , respectively].

**Probe tone task.** The probe tone rating data were submitted to a 4 (no change and two, four, and six half-steps)  $\times$  2 (probe key)  $\times$  7 (number of notes) repeated measures ANOVA. There was a significant main effect of key change [ $F(3,249) = 23.58$ ,  $MS_e = 3.68$ ,  $p < .001$ ], with ratings being highest in the no key change condition ( $M = 4.2$ ), followed by the two half-step ( $M = 4.0$ ), four half-step ( $M = 3.8$ ), and six half-step ( $M = 3.8$ ) conditions. Thus, the greater the key change, the lower the probe tone ratings. There was also a significant main effect of probe key [ $F(1,83) = 22.76$ ,  $MS_e = 2.44$ ,  $p < .001$ ], with the listeners providing higher ratings when the probe tone matched the most recent key ( $M = 4.3$ ) than when it mismatched ( $M = 3.9$ ). Finally, there was a significant main effect of number of notes [ $F(6,498) = 4.15$ ,  $MS_e = 3.00$ ,  $p < .001$ ]. The mean ratings for one to seven notes were 4.1, 3.9, 4.0, 4.0, 4.0, 4.2, and 4.2, respectively. Thus, the ratings seem to have increased as the number of notes increased, but this change was relatively small.

There was also a significant key change  $\times$  probe key interaction [ $F(3,249) = 29.942$ ,  $MS_e = 3.0$ ,  $p < .001$ ]. As can be seen in Table 2, overall probe tone ratings were affected by the match between the probe and the most recent key only when there was no key change. If the data for the no key change condition are analyzed separately, there is an effect of probe key [ $F(1,83) = 81.86$ ,  $MS_e = 5.32$ ,  $p < .001$ ]. In contrast, if the key change conditions are analyzed without the no key change condition, the main effect of probe key is marginally significant at best [ $F(1,83) = 3.53$ ,  $MS_e = 3.38$ ,  $p = .06$ ]. Thus, when there is no key change, it is relatively easy to assess whether the probe tone fits in the most recent key. The lack of a clearly significant effect of probe key for the shift conditions might suggest that the listeners were confused about what the key was. However, other aspects of the data suggest that they were adjusting in some complex ways.

In the main analysis, there was a significant probe key  $\times$  number of notes interaction [ $F(6,498) = 4.17$ ,  $MS_e = 3.07$ ,  $p < .001$ ]. If the no key change condition is analyzed separately, this interaction is not significant [ $F(6,498) = 1.28$ ,  $MS_e = 1.09$ ,  $p = .27$ ], suggesting that the number of additional notes is not as important when there has been no key change. However, this interaction is significant when the data for the key change conditions are analyzed without the no key change condition [ $F(6,498) = 4.22$ ,  $MS_e = 3.64$ ,  $p < .001$ ]. The data point to the idea that when there has been a key change and the probe tone matches that new key, there is a gradual increase in the probe tone ratings as the number of notes accumulates.<sup>1</sup> For the matching key data, there is a significant main effect of number of notes [ $F(6,498) = 6.84$ ,  $MS_e = 3.68$ ,  $p < .001$ ]. The rate of increase is similar across the different key change conditions, as the key change  $\times$  number of notes interaction is not significant

**Table 1**  
**A' Detection Data for Experiment 1**

Key Change	Number of Notes							<i>M</i>
	1	2	3	4	5	6	7	
Two half-step shift	.64	.69	.69	.69	.69	.73	.71	.69
Four half-step shift	.72	.78	.75	.77	.78	.77	.75	.76
Six half-step shift	.81	.80	.81	.81	.78	.79	.82	.80
<i>M</i>	.72	.76	.75	.76	.75	.76	.76	.75

**Table 2**  
**Probe Rating Data for Experiment 1**

Key Change	Number of Notes							<i>M</i>
	1	2	3	4	5	6	7	
Matching Key								
None	4.9	4.6	4.8	4.7	4.6	4.9	5.2	4.8
Two half-step shift	4.0	4.0	4.2	4.4	4.6	4.6	4.8	4.4
Four half-step shift	3.7	3.5	3.9	3.8	3.9	4.3	4.4	3.9
Six half-step shift	3.7	3.9	3.8	3.8	4.0	4.2	4.2	3.9
Shift <i>M</i>	3.8	3.8	4.0	4.0	4.2	4.4	4.5	4.1
Mismatching Key								
None	3.5	3.6	3.7	3.7	3.4	3.7	3.6	3.6
Two half-step shift	4.6	4.3	4.2	4.2	4.3	4.5	4.2	4.3
Four half-step shift	3.7	3.9	3.8	3.8	3.6	3.7	3.9	3.8
Six half-step shift	3.8	3.8	3.6	3.6	3.6	3.8	3.7	3.7
Shift <i>M</i>	4.0	4.0	3.9	3.9	3.8	4.0	3.9	3.9

( $F < 1$ ). In contrast, when there has been a key change and the probe tone mismatches the new key (i.e., it fits the original key), there is no change as the number of notes accumulates. For the mismatching key data, the main effect of number of notes is not significant ( $F < 1$ ).

To summarize, for the influence of the number of notes, there is no impact when there is no key change. When there is a key change, it appears as though people are continuously trying to build up an idea of what the new key is, and this adaptation process takes time. Concurrently, listeners appear to be retaining some idea of the original key. After a key change, acceptability of the original key is somewhat lower, but it does not gradually decline.

**Effects of training.** Musical training was positively correlated with overall performance on the detection task ( $r = .48, p < .001$ ). The more experienced a listener, the more accurate he or she was. However, when experience was used as a covariate in a 3 (key change)  $\times$  7 (number of notes) analysis of covariance (ANCOVA), it neither interacted with any of the variables nor altered the basic pattern of results.

Although musical training was not correlated with the overall probe tone ratings ( $r = -.06, p = .57$ ), when it was used as a covariate in a 4 (key change)  $\times$  2 (probe key)  $\times$  7 (number of notes) ANCOVA, it was involved in an interaction with key change and probe key [ $F(3,246) = 4.16, MS_e = 3.82$ ]. Separate analyses were then done for each key change condition. Training was not related to probe tone ratings when there had been a key change (all  $p_s > .10$ ), but it was related when there was no key change, influencing performance on the basis of which key the

probe tone came from [ $F(1,82) = 7.01, MS_e = 4.96$ ]. The more training the listeners had, the greater the difference in their ratings of probe tones from the key of the melody and their ratings of a key that was not heard. This was confirmed by a positive correlation between training and the difference between the ratings for these two probe key types ( $r = .28, p = .01$ ).

## Discussion

When the listeners detected key changes in melodies, they did so very rapidly. As was revealed by our signal detection analysis, within one or two notes, the listeners had noticed that a key change had occurred. A greater number of notes beyond the first two did not have an impact on detection of a key change. As was predicted, it was also observed that the greater the shift in key, the easier it was for a listener to detect that a key change had occurred. In the present study, the largest key change (six half-steps) produced almost immediate detection.

In terms of the adjustment to the new key, the probe tone rating data suggest that this is a relatively slow process. This adjustment continued as the number of notes after the key change increased. Thus, although detection of a key change occurs very rapidly, adapting to the new key takes much longer. Of course, one might object to making such a comparison across tasks. Perhaps detection of key changes occurs just as gradually as adapting to the new key, but this is obscured because of scale differences. The 7-point probe rating scale might just be more sensitive to gradual changes than is the 2-point yes/no detection scale. However, there are two problems

with this objection. First, there was clearly a significant difference across key change distances in the detection task, so the 2-point response scale was sensitive enough to reveal at least one effect. Second, even ignoring the yes/no data, the 7-point probe rating data also suggest that change detection is relatively abrupt. After there had been a key change, probe tones from the original key were given consistently low ratings. Ratings did not decline with the number of notes.

Musical training improved the ability to detect key changes and to correctly classify probe tones when there had not been a key shift. However, the basic pattern of results following a key change does not appear to have been altered by the listener's having had formal training, since key distance effects were not moderated by musical experience.

Contrary to our expectations, key change distance did not significantly affect tracing of the new key. This could have occurred for several reasons. First, unlike other studies, in our experiment there was an intervening task between the melody and the probe tone. The time it took for the listeners to respond on the detection task may have hindered rating performance. Second, the probe tones came from only one octave, so they sometimes fell far outside the range of the end of the stimulus. This may have reduced the probe tone's perceived similarity with the ending key and increased its similarity to the original key. Finally, the use of only two probes may simply be inadequate when an attempt is made to trace representations of key. The issue of tracing key representations via probe tones is taken up again in Experiment 3.

A more immediate concern about Experiment 1 involves the key change distances used. Transpositions of two, four, and six half-steps should represent increasing levels of distance between keys. Such an approach has been used in previous studies that involved pivot chord modulations (Thompson & Cuddy, 1989, 1992). However, because we used unharmonized melodies, these key change distances also happened to correspond to increasing distances in pitch height. The listeners could have been responding on the basis of either pitch height or key distance.

## EXPERIMENT 2

Experiment 2 was designed to disentangle the influences of pitch height and key distance on key change detection. Shifts of six and seven half-steps were used (cf. Cuddy et al., 1979). The six half-step condition was a shift to IV#, a very distant key based on the toroidal model. The seven half-step condition, although slightly larger in terms of pitch height, resulted in a key change up to V or down to IV, both of which were very close to the original key. If the listeners detected key changes primarily on the basis of pitch height in Experiment 1, they should perform better in the seven half-step condition. Conversely, if the listeners were responding primarily on the basis of key distance, they should respond less accurately in the seven half-step condition.

**Table 3**  
**A' Detection Data for Experiment 2**

Key Change	Number of Notes					<i>M</i>
	1	2	3	4	5	
Six half-step shift	.69	.70	.74	.71	.72	.71
Seven half-step shift	.59	.66	.67	.68	.67	.65
<i>M</i>	.64	.68	.71	.69	.69	.68

## Method

**Listeners.** Forty undergraduates from the University of Notre Dame subject pool participated in Experiment 2. All the listeners reported normal hearing, and none reported having absolute pitch or previous contact with the source of the stimuli. The listeners had 0–14 years ( $M = 4.20$ ,  $SD = 3.64$ ) of formal training on an instrument or with voice.

**Apparatus and Stimuli.** Of the 84 critical melodies used in Experiment 1, 20 were chosen at random for use in Experiment 2. For each participant, half of the melodies were randomly assigned to the *no key change* condition. The remaining *key change* melodies were randomly assigned to modulate six or seven half-steps up or down. Because key changes were detected so rapidly in the previous experiment, the number of notes available to the listeners in Experiment 2 was varied from one to five, rather than from one to seven.

**Procedure.** The procedure, including the practice period, was identical to that in Experiment 1, except that the probe tone task was not used.

## Results and Discussion

**Detection task.** Once again, false alarm rates in the no key change condition were uniformly low ( $M = .22$ ,  $SD = .17$ ). In an analysis using only the no key change condition, the effect of the number of notes was not significant ( $F < 1$ ). Thus, when there was no key change, the listeners were rarely claiming that there was one.

Mean  $A'$  values are displayed in Table 3. The data were first submitted to a 2 (key change: six or seven half-steps)  $\times$  5 (number of notes: one to five) ANOVA. There was a significant main effect of key change [ $F(1,39) = 17.68$ ,  $MS_e = 0.02$ ,  $p < .001$ ], but neither the main effect of number of notes nor the interaction was significant [ $F(4,156) = 1.01$ ,  $MS_e = 0.05$ ,  $p = .41$  and  $F < 1$ , respectively]. When the key change was to six half-steps away, it was detected immediately.

Although the key change in the seven half-step condition was further from the original in terms of pitch height, detection performance did not improve but, rather, worsened. This pattern of results indicates that the listeners were responding primarily on the basis of key distance, as opposed to pitch height. Moreover, detectability did not increase with more evidence (even if the seven half-step condition is analyzed alone, the effect of the number of notes is not significant;  $F = 1.30$ ). This is in contrast to Experiment 1, in which initial low levels of detectability were associated with increased performance. The generally poor performance in the seven half-step condition was due to the fact that the second key was so closely related to the first. The listeners had a harder time detecting that there had been a shift.

**Effects of training.** Although musical training was positively correlated with overall performance, this rela-

tionship was only marginally significant ( $r = .29, p = .07$ ). Moreover, when experience was used as a covariate in a 2 (key change: six or seven half-steps)  $\times$  5 (number of notes: one to five) ANCOVA, it neither interacted with any of the variables nor altered the basic pattern of results.

The results from Experiment 2 generally confirmed the findings of Experiment 1. In the change detection task, the listeners were responding primarily on the basis of changes in keys, rather than changes in pitch height. They detected key changes relatively rapidly or not at all, and they found it easier to recognize changes to more distant keys. What remained to be confirmed was the more gradual response to key changes found when probe ratings were used to trace key representation.

### EXPERIMENT 3

In Experiment 3, we sought to extend the key-tracing results from Experiment 1 showing that the listeners' representations of key changed gradually despite the narrow window of time for an explicit recognition of change. All 12 probe tones were employed this time (one for each pitch class), and this allowed us to generate profiles for each participant for each condition. The change detection task was dropped so that it would not interfere with probe tone ratings. The six and seven half-step key changes were used again so as to allow comparison with the detection task in Experiment 2. Probe tones were rated at one, three, and five notes after the point of change.

In Experiment 1, the listeners had been instructed to rate how well probe tones fit with "the end of the melody." Although this instruction had been intended to focus listeners on the ending key, it may have encouraged them to respond on the basis of closure. To avoid this possibility, the listeners in Experiment 3 were asked to rate how well the tones fit with the melodies, and not specifically with their endings.

Probe tone ratings ought to reveal a gradual adaptation to the new key. This was assessed in two ways. First, simple correlations were computed between the 12 probe tone ratings given by the participants and the standard tonal profiles for the old and new keys. This was done for each participant for each condition, providing a series of within-subjects  $r$ s. A higher  $r$  would be indicative of a stronger or more active representation of that key.

However, one concern with the use of probe tone ratings is that they may reflect how frequently the probe tone occurred within the melodic sequence itself, rather than the activation of a listener's representation of key (Butler, 1989). Some studies have shown that recent exposure to the probed pitch accounts for at least some part of the simple correlations to the key profiles (Oram & Cuddy, 1995; Parncutt & Bregman, 2000; but see Cuddy & Thompson, 1992). For this reason, the second analytical approach was to use simultaneous multiple regression to disentangle the multiple predictors of probe tone ratings (Krumhansl & Schmuckler, 1986; see also Cuddy & Thompson, 1992). This resulted in a series of within-

subjects standardized regression weights ( $\beta$ s) for each of three predictors: *old key profile*, *new key profile*, and *probe tone occurrence within the melodic sequences (sequence profile)*.

### Method

**Listeners.** Seventy-two undergraduates participated, with 36 recruited and tested at the University of Notre Dame and 36 at Washington University. All the listeners reported normal hearing, and none reported having absolute pitch. Data from 1 participant had to be replaced because she had had previous experience with the materials. The listeners had 0–16 years ( $M = 4.64, SD = 4.43$ ) of formal training on an instrument or with voice.

**Apparatus and Stimuli.** Additional melodies were selected from the same source (Ottman, 1996) and from other sight-singing texts (Cole & Lewis, 1909; Crowe, Lawton, & Whittaker, 1933), for a total of 144 critical melodies. Five practice melodies were also chosen. All the melodies were selected with the same criteria as those used in Experiment 1.

All the stimuli were rerecorded into .wav format for use with E-Prime experimental software (Schneider, Eschman, & Zuccolotto, 2002). Due to hard disk space limitations, recordings used an 8-bit, 22.5-kHz sampling rate.

Transpositions of zero (no shift), six (#IV), and seven (V/IV) half-steps were used, and there were one, three, or five notes played after the key change. Unlike in Experiment 1, all 12 probe tones were employed. Probe tones consisted of four octaves of a pitch played simultaneously so as to ensure that they overlapped with the register of both the beginning and the ending of the melodic stimulus.

Assignment of melodies to shift/no-shift status, number of notes, and probe tone was rotated across participants. Because the additional counterbalancing of melody assignment to shift type (six or seven half-steps) would have required 144 participants, this assignment was done randomly for each participant.

**Procedure.** The experiment began with 5 practice trials and then 144 critical trials. On each trial, a melody was played, which was followed by a 1,250-msec delay and then a probe tone. The listeners were instructed to rate how well the tone "fits with the melody." Ratings were given on a scale from 1 to 7, with 7 representing *very well*, 1 representing *very poorly*, and 4 representing *unsure*. The listeners were encouraged to use the full range of the scale. A 1,500-msec interval separated each trial.

### Results

**Bivariate correlations with key profiles.** For the following analyses, the sets of 12 probe tone ratings for each cell were examined for each participant. These sets of ratings were correlated with standard key profiles (Krumhansl & Kessler, 1982) for both the old and the new keys, to create a series of within-subjects  $r$ s. The resulting mean correlations are displayed in Figure 2. These correlations were the dependent variables in the analyses below. All statistics were conducted on Fisher's  $z'$  transformation of  $r$  so as to reduce skew (Cohen & Cohen, 1983).

If the listeners were representing the keys presented in the stimuli, their probe tone ratings should correspond well to the standard key profiles, especially when only one key was presented in the passage. In the no-shift condition, probe tone ratings did correlate strongly with the standard profile for that key (mean  $r = .52$ ). The correlations did not change with the number of notes presented [ $F(2,142) < 1$ ].

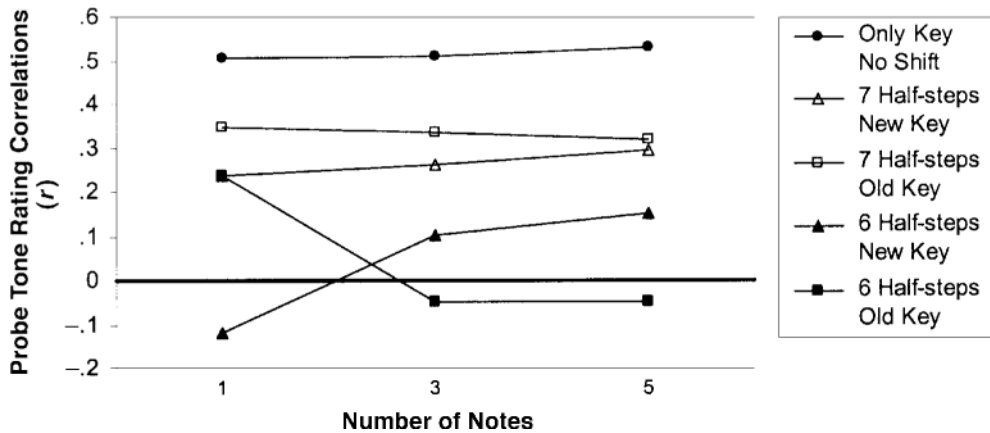


Figure 2. Bivariate correlations between probe tone ratings and key profiles for old and new keys in Experiment 3.

The remaining correlations shown in Figure 2 suggest a more complicated pattern of results for the shift conditions. As the number of notes increased, the probe tone ratings tended to correlate more with the new key profile and less with the old key profile. A 2 (key profile)  $\times$  3 (number of notes) repeated measures ANOVA revealed both a significant interaction for the six half-step shift condition [ $F(2,142) = 17.78$ ,  $MS_e = 0.23$ ,  $p < .00001$ ] and a marginally significant interaction for the seven half-step shift condition [ $F(2,142) = 2.91$ ,  $MS_e = 0.04$ ,  $p < .06$ ]. The tradeoff between the new and the old key profiles as predictors occurred earlier in the six half-step condition, with the  $r$  for the old key dropping to near chance (mean  $r = -.05$ ) by the third note after the key change. In confirmation of this, a 2  $\times$  2  $\times$  3 ANOVA revealed a significant three-way interaction between key change, key profile, and number of notes [ $F(2,142) = 12.61$ ,  $MS_e = 0.12$ ,  $p < .0001$ ].

**Simultaneous regression with probed pitch occurrence in the melody.** Three independent variables were used in simultaneous multiple regression: old key profile, new key profile, and sequence profile. The sequence profile consisted of the duration of the probed pitch occurrence within a melodic sequence divided by the total duration of the sequence. Old key and new key profiles were collapsed into a single profile for the no-shift condition. This regression analysis was applied for each participant  $\times$  key change  $\times$  number of notes cell. The resulting within-subjects  $\beta$  coefficients were used in the analyses below as the dependent measure.

As was expected, the occurrence of the probed pitch within the melodic sequence was associated with higher probe tone ratings (see Figure 3). The mean  $\beta$  weight for the sequence profile was .21, which was significantly greater than zero [ $t(71) = 8.95$ ,  $p < .00001$ ]. The impact of sequence profile was not moderated by key change type or number of notes or by their interaction (all  $ps > .10$ ).

Despite the effect of sequence profile, the pattern of results for the old and new key profile  $\beta$  coefficients was

very similar to the pattern of bivariate correlations described earlier. In the no-shift condition (Figure 3A), probe tone ratings were still predicted by the standard key profile (mean  $\beta = .30$ ), and this was not moderated by the number of notes presented [ $F(2,142) < 1$ ]. For the shift conditions, as the number of notes increased, probe tone ratings tended to be better predicted by the new key profile and less so by the old key profile. For the six half-step condition, a 2 (key profile)  $\times$  3 (number of notes) repeated measures ANOVA revealed a significant interaction [ $F(2,142) = 11.04$ ,  $MS_e = 0.11$ ,  $p < .0001$ ]. At the first note after the six half-step shift,  $\beta$  was significantly above zero for the old, but not the new, key profile. By the third note, though, this pattern had reversed (Figure 3B). Similarly for the seven half-step shift condition, the new key profile replaced the old one as a significant predictor as more notes were presented in the key, although the tradeoff occurred later here than in the six half-step shift condition (Figure 3C).

The only major change in the results that occurred with analysis of  $\beta$ s was that the profile  $\times$  number of notes interaction in the seven half-step shift condition was no longer significant, despite the crossover pattern that appeared in the mean  $\beta$  weights [ $F(2, 142) = 1.37$ ,  $MS_e = 0.27$ ,  $p = .26$ ]. This likely occurred for two reasons. First, the strong positive correlation between profiles for the starting and the ending keys in that condition ( $r = .59$ ) makes it generally difficult to disentangle their effects. Second, and particularly important for the multiple regression analysis, partialing the effect of pitch occurrence necessarily reduced the impact of the other profiles, since pitch occurrence was strongly correlated with both the old key (mean  $r = .70$ ) and the new key (mean  $r = .50$ ) profile for the seven half-step shift condition. Such multicollinearity increases the standard error of  $\beta$ , which would make it difficult to detect the sought after interaction.

**Effects of training.** Training was positively correlated with  $\beta$  for the sequence profile ( $r = .27$ ,  $p < .05$ ).



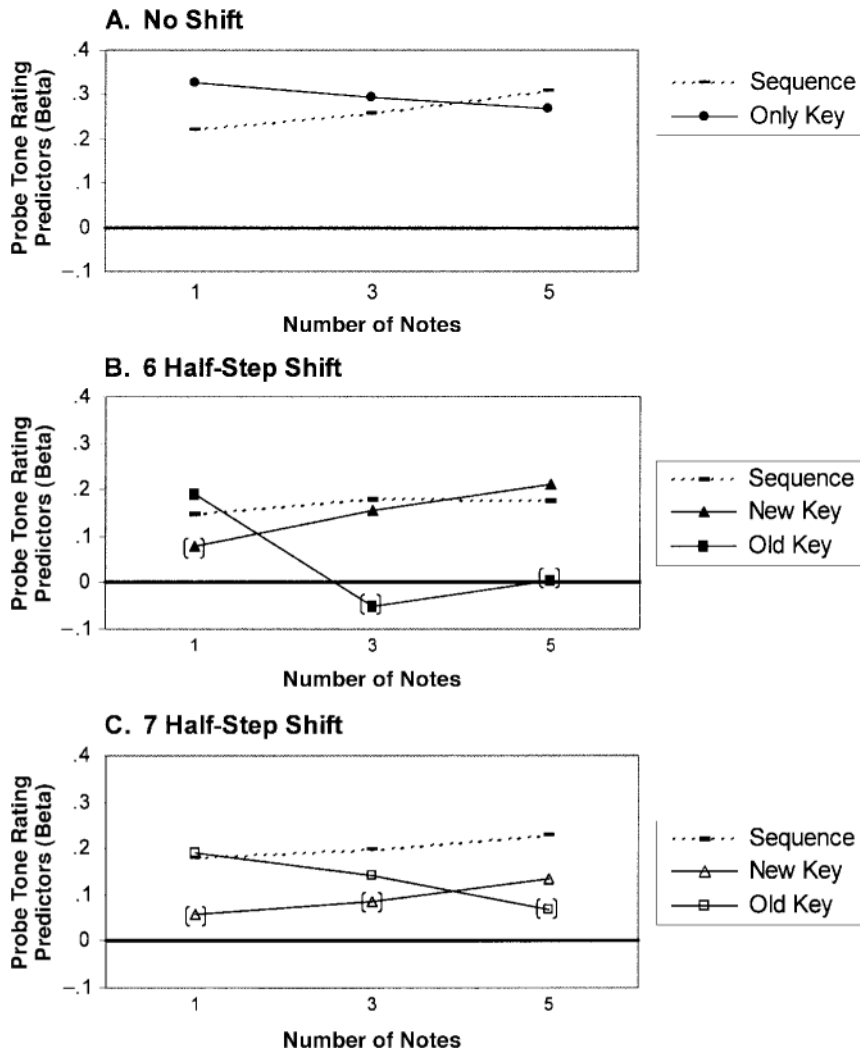


Figure 3. Standardized regression coefficients ( $\beta$ ) for sequence, old key, and new key profiles as predictors of probe tone ratings in Experiment 3. Data points in parentheses are not significantly different from zero ( $\alpha = .05$ ).

Training also showed a weak but positive correlation with  $\beta$  for the only key profile in the no-shift condition ( $r = .21, p < .08$ ). In other words, expertise increased sensitivity to individual notes within the melody, but also to a more stable long-term representation of key when there was no key change.

Otherwise, training had no observable impact on  $\beta$  coefficients. When used as a covariate in a 2 (shift type: six or seven half-steps)  $\times$  2 (key profile: old or new)  $\times$  3 (number of notes) ANCOVA, it had neither a significant main effect nor an interaction with any other variable (all  $ps > .10$ ).

## GENERAL DISCUSSION

The first aim of the present study was to assess the speed with which listeners are able to detect key changes

in melodies. This was done using a key change detection task and was assessed using a signal detection analysis. In both Experiments 1 and 2, key change detection was relatively rapid, occurring within the first one or two notes after the change had occurred. Key change detection might take place in such a narrow window of time because it requires attention at the specific moment of change (Berent & Perfetti, 1993). This is potentially important, because if attention is a unitary and amodal resource (Jolicœur, 1999), change detection in different modalities might share some of the same underlying cognitive constraints.

As was observed in Experiments 1 and 3, also occurring soon after the change was a reduction in the correspondence between probe tone ratings and the original key. This drop in the listeners' sense of the first key may indicate the detection of change. An examination of the

effect of key distance across both tasks supports this interpretation. Distant changes were associated with increased discriminability in the key change detection task (Experiments 1 and 2), as well as with a greater reduction in probe tone ratings' correlation with the original key's profile (Experiment 3). This effect of distance depended primarily on tonal relatedness to the original key, rather than on pitch height (Experiment 2).

In contrast to these abrupt changes in response, the listeners' sense of the new key appeared to develop gradually (Experiments 1 and 3). One way of describing this contrast is that a period of *key limbo* follows a key change. During this period, listeners recognize that the old key is less relevant, but they have yet to determine the new orienting schema. This is consistent with the notion that different keys may be represented at least somewhat independently of one another. Giving up one key does not necessitate the immediate and total replacement by another.

Key distance appears to have an effect on the period of key ambiguity. Despite several differences between the probe tone methodology used here and that used by Krumhansl and Kessler (1982), our results largely converge with theirs on the effect of key distance. For changes to a close key, the listeners' sense of the old key dropped, but not completely or immediately. The listeners tended to maintain at least some sense of the first key while developing a sense of the new key. In contrast, for changes to a distant key, the listeners' sense of the old key dropped off more dramatically, and the new key took longer to become acceptable.

This gradual adaptation to the new key raises a question. Results from the detection task showed that detection was confined mainly to the first one or two notes after the key change. Why did a gradual increase in stability of the new key, as indicated by probe tone ratings, not trigger detection *after* the first few notes of the new key? One possibility is that gradual accumulation of pitch information in the new key leads to change blindness because it is too subtle to automatically draw the observer's attention (Simons, Franconeri, & Reimer, 2000). Alternatively, our findings may be related to other asymmetries in which it is easier to notice a change from a stable to an unstable representation than vice versa. In music perception, this asymmetry has been observed in moving from consonance to dissonance (Schellenberg, 2002), as well as from stable to unstable pitches within a key (e.g., G to F# in the key of C; Krumhansl, 1990). As long as the initial key is, in some sense, psychologically stable, it may be easier to notice a departure from that key than an approach to a new one.

Regarding the effect of training, musically experienced listeners were better able to detect key changes than untrained listeners were. This might have occurred because musical expertise allows for more controlled allocation of attention toward music (Berent & Perfetti, 1993). For the probe tone task, training was associated with a stronger sense of the key of the music when there was no key change. Others have similarly found that

trained listeners have a stronger sense of the tonality of a musical excerpt (Krumhansl & Shephard, 1979; Oram & Cuddy, 1995). Finally, training was associated with greater use of the pitch content of the melody as a basis for judging probe tones (see also Oram & Cuddy, 1995). One possible explanation for this is that musicians simply have better memory for music and that they judge probe tones as fitting better when they remember the same tones as having occurred within the melody.

Musical training affected detection and tracing of key changes in various ways, but across all three experiments, training never moderated the effect of key distance. This result is quite consistent with previous findings (Bartlett & Dowling, 1980; Cuddy & Thompson, 1992) and suggests that the psychological distance between keys is very similar among adult listeners in our culture. More important, training never moderated the effect of the number of notes in the new key. The time course of key change perception appears to be consistent across a range of training levels.

Returning to the dilemma we started with, can we point to a specific juncture at which listeners respond to melodic key changes? It appears that the answer is both *yes* and *no*. Yes, listeners quickly notice that the old key no longer applies. However, there appears to be no specific point in time at which the new key is discovered. Listeners gradually come to identify the new key over a period of time.

## REFERENCES

- ALDWELL, E., & SCHACTER, C. (1989). *Harmony and voice leading* (2nd ed.). San Diego: Harcourt Brace Jovanovich.
- BARTLETT, J. C., & DOWLING, W. J. (1980). Recognition of transposed melodies: A key-distance effect in developmental perspective. *Journal of Experimental Psychology: Human Perception & Performance*, **6**, 501-515.
- BERENT, I., & PERFETTI, C. A. (1993). An on-line method in studying music parsing. *Cognition*, **46**, 203-222.
- BUTLER, D. (1989). Describing the perception of tonality in music: A critique of the tonal hierarchy theory and a proposal for a theory of intervallic rivalry. *Music Perception*, **6**, 219-242.
- COHEN, J., & COHEN, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- COLE, S. W., & LEWIS, L. R. (1909). *Melodia: A comprehensive course in sight-singing (solfeggio)*. Boston: Ditson.
- CROWE, E., LAWTON, A., & WHITTAKER, W. G. (Eds.) (1933). *The folk song sight singing series*. London: Oxford University Press.
- CUDDY, L. L., COHEN, A. J., & MILLER, J. (1979). Melody recognition: The experimental application of musical rules. *Canadian Journal of Psychology*, **33**, 148-157.
- 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**, 51-59.
- DONALDSON, W. (1992). Measuring recognition memory. *Journal of Experimental Psychology: General*, **121**, 275-277.
- JOLICEUR, P. (1999). Restricted attentional capacity between sensory modalities. *Psychonomic Bulletin & Review*, **6**, 87-92.
- KNOPOFF, L., & HUTCHINSON, W. (1983). Entropy as a measure of style: The influence of sample length. *Journal of Music Theory*, **27**, 75-97.
- KRUMHANSL, C. L. (1990). *Cognitive foundations of musical pitch*. New York: Oxford University Press.
- 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., & SCHMUCKLER, M. A. (1986). The Petroushka chord: A perceptual investigation. *Music Perception*, **4**, 153-184.

- KRUMHANSL, C. L., & SHEPARD, R. N. (1979). Quantification of the hierarchy of tonal functions within a diatonic context. *Journal of Experimental Psychology: Human Perception & Performance*, *5*, 579-594.
- ORAM, N., & CUDDY, L. L. (1995). Responsiveness of Western adults to pitch-distributional information in melodic sequences. *Psychological Research*, *57*, 103-118.
- OTTOMAN, R. W. (1996). *Music for sight singing* (4th ed.). Englewood Cliffs, NJ: Prentice-Hall.
- PARNCUTT, R., & BREGMAN, A. S. (2000). Tone profiles following short chord progressions: Top-down or bottom-up? *Music Perception*, *18*, 25-57.
- PISTON, W. (1962). *Harmony* (3rd ed.). New York: Norton.
- RENSINK, R. A. (2002). Change detection. *Annual Review of Psychology*, *53*, 245-277.
- RENSINK, R. A., O'REGAN, J. K., & CLARK, J. J. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science*, *8*, 368-373.
- SHELLENBERG, E. G. (2002). Asymmetries in the discrimination of musical intervals: Going out-of-tune is more noticeable than going in-tune. *Music Perception*, *19*, 223-248.
- SCHENKER, H. (1954). *Harmony* (O. Jonas, Ed., and E. M. Borgese, Trans.). Chicago: University of Chicago Press. (Original work published 1906)
- SCHNEIDER, W., ESCHMAN, A., & ZUCCOLOTTA, A. (2002). *E-Prime user's guide*. Pittsburgh: Psychological Software Tools.
- SCHOENBERG, A. (1954). *Structural functions of harmony*. New York: Norton.
- SHEPARD, R. N. (1982). Structural representations of musical pitch. In D. Deutsch (Ed.), *Psychology of music* (pp. 343-390). Orlando, FL: Academic Press.
- SHMULEVICH, I., & YLI-HARJA, O. (2000). Localized key finding: Algorithms and applications. *Music Perception*, *17*, 531-544.
- SIMONS, D. J., FRANCONERI, S. L., & REIMER, R. L. (2000). Change blindness in the absence of a visual disruption. *Perception*, *29*, 1143-1154.
- SNODGRASS, J. G., & CORWIN, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General*, *117*, 34-50.
- TEMPERLEY, D. (1999). What's key for key? The Krumhansl-Schmuckler key-finding algorithm reconsidered. *Music Perception*, *17*, 65-100.
- THOMPSON, W. F. (1993). Modeling perceived relationships between melody, harmony, and key. *Perception & Psychophysics*, *53*, 13-24.
- THOMPSON, W. F., & CUDDY, L. L. (1989). Sensitivity to key change in chorale sequences: A comparison of single voices and four-voice harmony. *Music Perception*, *7*, 151-168.
- THOMPSON, W. F., & CUDDY, L. L. (1992). Perceived key movement in four-voice harmony and single voices. *Music Perception*, *9*, 427-438.
- THOMPSON, W. F., & CUDDY, L. L. (1997). Music performance and the perception of key. *Journal of Experimental Psychology: Human Perception & Performance*, *23*, 116-135.
- THOMPSON, W. F., & MOR, S. (1992). A perceptual investigation of polytonality. *Psychological Research*, *54*, 60-71.
- WATANABE, K. (2003). Differential effect of distractor timing on localizing versus identifying visual changes. *Cognition*, *88*, 243-257.
- YOUNGBLOOD, J. E. (1958). Style as information. *Journal of Music Theory*, *2*, 24-35.

## NOTE

1. We also considered the possibility that gradual increases in probe tone ratings for this condition occurred because the probed pitch was more and more likely to be included in the melody after the key change. There was a small but significant relationship between probe tone ratings and the number of occurrences of the probe's pitch in the last seven notes played [ $r = .18$ ;  $t(1762) = 7.62$ ,  $p < .001$ ]. However, none of our findings changed when this was statistically controlled for.

## APPENDIX

Table A1 displays the distribution of tone durations in the melodic sequences of the three experiments. Nondiatonic pitches never occurred and so are not shown in the table. The distribution after the point of change is taken from the no-shift condition. Before the point of change, the first, third, and fifth degrees were especially frequent. After this point, the first degree became less prominent, the third and fifth degrees stayed the same, and the remaining diatonic tones became more prominent.

**Table A1**  
Mean Distributions of Pitch Durations in  
Melodies Shown with Reference to a Tonic of C

Pitch	Experiment			<i>M</i>
	1	2	3	
	Before Point of Change			
C	.27	.27	.27	.27
D	.13	.13	.13	.13
E	.20	.25	.21	.22
F	.07	.07	.07	.07
G	.23	.22	.23	.23
A	.05	.03	.05	.04
B	.05	.02	.04	.04
	After Point of Change			
C	.15	.11	.13	.13
D	.15	.15	.14	.15
E	.19	.21	.22	.20
F	.12	.14	.13	.13
G	.21	.25	.21	.22
A	.09	.07	.09	.08
B	.09	.07	.08	.08

## APPENDIX (Continued)

Overall, though, the distributions of tones in the melody corresponded closely to the tonal profile for the tonic key both before and after the point of change. As is shown in Table A2, correlations ranged from .75 to .93 and tended to be higher before the change. All the correlations were significantly different from zero (all  $ps < .01$ ). Others have found a similar correspondence between melodic lines and standard tonal profiles. Krumhansl (1990) compared standard key profiles with the frequency of occurrence of pitches in 18th- and 19th-century vocal melodies (as tabulated by Knopoff & Hutchinson, 1983; Youngblood, 1958). Correlations there ranged from .84 to .93.

**Table A2**  
**Correlations Between Mean Distributions of**  
**Pitch Durations and Major Key Profiles for**  
**Both Before and After the Point of Change**  
**(in the No-Shift Condition)**

	Experiment			<i>M</i>
	1	2	3	
Before	.93	.91	.93	.92
After	.85	.75	.81	.80

(Manuscript received October 22, 2003;  
revision accepted for publication March 16, 2004.)