

Perceptual Tests of an Algorithm for Musical Key-Finding

Mark A. Schmuckler and Robert Tomovski
University of Toronto at Scarborough

Perceiving the tonality of a musical passage is a fundamental aspect of the experience of hearing music. Models for determining tonality have thus occupied a central place in music cognition research. Three experiments investigated 1 well-known model of tonal determination: the Krumhansl–Schmuckler key-finding algorithm. In Experiment 1, listeners' percepts of tonality following short musical fragments derived from preludes by Bach and Chopin were compared with predictions of tonality produced by the algorithm; these predictions were very accurate for the Bach preludes but considerably less so for the Chopin preludes. Experiment 2 explored a subset of the Chopin preludes, finding that the algorithm could predict tonal percepts on a measure-by-measure basis. In Experiment 3, the algorithm predicted listeners' percepts of tonal movement throughout a complete Chopin prelude. These studies support the viability of the Krumhansl–Schmuckler key-finding algorithm as well as a model of listeners' tonal perceptions of musical passages.

Keywords: music cognition, tonality, key-finding

As a subdiscipline with cognitive psychology, music perception provides a microcosm for investigating general psychological functioning (Schmuckler, 1997b), including basic psychophysical processing, complex cognitive behavior (e.g., priming, memory, category formation), issues of motor control and performance, and even social and emotional influences on musical behavior. Within this broad range, the apprehension of music has been most vigorously studied from a cognitive standpoint. One example of the type of insights into basic cognitive function afforded by such research is found in work on the psychological representation of pitch in a tonal context (e.g., Krumhansl, 1990; Schmuckler, 2004).

A fundamental characteristic of Western music is that the individual tones making up a piece of music are organized around a central reference pitch, called the *tonic* or *tonal center*, with music organized in this fashion said to be in a musical *key* or *tonality*. Tonality is interesting psychologically in that it represents a very general cognitive principle: that certain perceptual and/or conceptual objects have special psychological status (Krumhansl, 1990). Within psychological categories, for example, there is a gradient of representativeness of category membership, such that some members are seen as central to the category, functioning as reference points, whereas other members are seen as more peripheral to the category and, hence, function as lesser examples (see Rosch, 1975). This characterization is particularly apt for describing musical tonality. In a tonal context, the tonic note is considered the

best exemplar of the key, with the importance of this tone indicated by the fact that it is this pitch that gives the tonality its name. The remaining pitches (with the complete set composed of 12 pitches called the *chromatic scale*) then vary in terms of how representative they are of this tonality relative to this reference pitch.

Within Western tonal music, there are two categories of musical key—*major* and *minor* tonalities. For any given reference pitch, it is possible to produce a major and a minor tonality, with each tonality establishing a unique hierarchical pattern of relations among the tones. Moreover, major and minor tonalities can be built on any chromatic scale tone; thus, there are 24 (12 major and 12 minor) tonalities used in Western music. For all of these keys, however, the theoretical pattern of note relations holds, with the tonic functioning as the reference pitch and the remaining tones varying in their relatedness to this reference pitch. Table 1 presents this theoretical hierarchy for the 12 *semitones* (the smallest unit of pitch change in Western music) for both major and minor tonalities.

Krumhansl and colleagues (Krumhansl, 1990; Krumhansl, Bharucha, & Castellano, 1982; Krumhansl, Bharucha, & Kessler, 1982; Krumhansl & Kessler, 1982; Krumhansl & Shepard, 1979) have provided psychological verification of this theoretical hierarchy, using what is known as the *probe-tone procedure*. In this procedure, a listener hears a context passage that unambiguously instantiates a given tonality. The context is then followed by a probe tone, which is a single tone from the chromatic scale, and listeners provide a goodness-of-fit rating for this probe relative to the tonality of the previous context. By sampling the entire chromatic set, one can arrive at an exhaustive description of the perceptual stability of these individual musical elements vis-à-vis a given tonality. The ratings for these 12 events with reference to a key is known as the *tonal hierarchy* (Krumhansl, 1990).

The top panel of Figure 1 shows the ratings of the probe tones relative to a major and minor context, with the note C as the tonic (C major and C minor tonalities, respectively); for comparison with Table 1, the semitone numbering is also given. These ratings

Mark A. Schmuckler and Robert Tomovski, Department of Life Sciences, University of Toronto at Scarborough, Toronto, Ontario, Canada.

Portions of this work were presented at the 38th Annual Meeting of the Psychonomic Society, Philadelphia, Pennsylvania, November 1997. This research was supported by a grant from the Natural Sciences and Engineering Research Council of Canada to Mark A. Schmuckler.

Correspondence concerning this article should be addressed to Mark A. Schmuckler, Department of Life Sciences, University of Toronto, 1265 Military Trail, Scarborough, Ontario M1C 1A4, Canada. E-mail: marksch@utsc.utoronto.ca

Table 1
The Theoretical Hierarchy of Importance for a Major and Minor Tonality

Hierarchy level	Major hierarchy	Minor hierarchy
Tonic tone	0	0
Tonic triad	4 7	3 7
Diatonic set	2 5 9 11	2 5 8 10
Nondiatonic set	1 3 6 8 10	1 4 6 9 11

Note. Semitones are numbered 0–11.

represent the average ratings for the probe tones, taken from Krumhansl and Kessler (1982), and generally conform to the music-theoretic predictions of importance given in Table 1. So, for example, the tonic note for these tonalities, the tone C (semitone 0), received the highest stability rating in both contexts. At the next level of importance were ratings for the notes G (semitone 7) and E (semitone 4) in major and D#/E \flat (semitone 3) in minor (see Table 1 and Figure 1). Without describing the remaining levels in detail, the ratings for the rest of the chromatic scale notes map directly onto their music-theoretic descriptions of importance. In addition, Krumhansl and Kessler (1982) found that these patterns of stability generalized to keys built on other tonics. Accordingly, the profiles for other keys can be generated by shifting the patterns of Figure 1 to other tonics. The bottom panel of Figure 1 shows this shift, graphing the profiles for C and F# major tonalities. Overall, these findings provide strong evidence of the psychological reality of the hierarchy of importance for tones within a tonal context.

One issue that has been raised with this work involves concerns over the somewhat subjective nature of these ratings as a viable model of perceivers' internal representations of tonal structure. Although it is true that these ratings are subjective, the probe-tone technique itself has proved to be a reliable means of assessing the perceived stability of the tones of the chromatic scale with reference to a tonality (see Smith & Schmuckler, 2004, for a review). For instance, the ratings arising from this procedure have been found to match with the tonal consonance of tone pairs (see Krumhansl, 1990, pp. 50–62) and mimic the statistical distributions of tone durations or frequency of occurrence of tones within actual musical contexts (see Krumhansl, 1990, pp. 62–76). Thus, the tones that are the most important from music-theoretic and psychological perspectives are those that most frequently occur or are heard for the longest duration. Conversely, theoretically and psychologically unimportant tones are both less frequent and occur for shorter durations.

Probe-tone data have been found to be related to a number of psychological processes, including memory confusions between notes in a recognition memory context (Krumhansl, 1979) and memory confusions of chords within a musical context (Bharucha & Krumhansl, 1983; Justus & Bharucha, 2002). Moreover, probe-tone findings are related to reaction times for perceptual judgments of tones (Janata & Reisberg, 1988), listeners' abilities to spontaneously detect "wrong" notes in an ongoing musical context (Janata, Birk, Tillman, & Bharucha, 2003), and tonality has also been found to influence the speed of expectancy judgments (Schmuckler & Boltz, 1994) and the intonation and/or dissonance judgments of chords (Bharucha & Stoeckig, 1986, 1987; Bigand,

Madurell, Tillman, & Pineau, 1999; Bigand, Poulin, Tillman, Madurell, & D'Adamo, 2003; Tillman & Bharucha, 2002; Tillman, Bharucha, & Bigand, 2000; Tillman & Bigand, 2001; Tillman, Janata, Birk, & Bharucha, 2003). Thus, the probe-tone procedure, and the ratings to which it gives rise, is a psychologically robust assessor or assessment of tonality.

Models of Musical Key-Finding

Given the importance of tonality, it is not surprising that modeling the process of musical key-finding has, over the years, played a prominent role in music-theoretic and psychological research, resulting in models of key determination from artificial intelligence (e.g., Holtzman, 1977; Meehan, 1980; Smoliar, 1980; Winograd, 1968), neural network (Leman, 1995a, 1995b; Shmulevich & Yli-Harja, 2000; Vos & Van Geenan, 1996), musicological (Brown, 1988; Brown & Butler, 1981; Brown, Butler, & Jones,

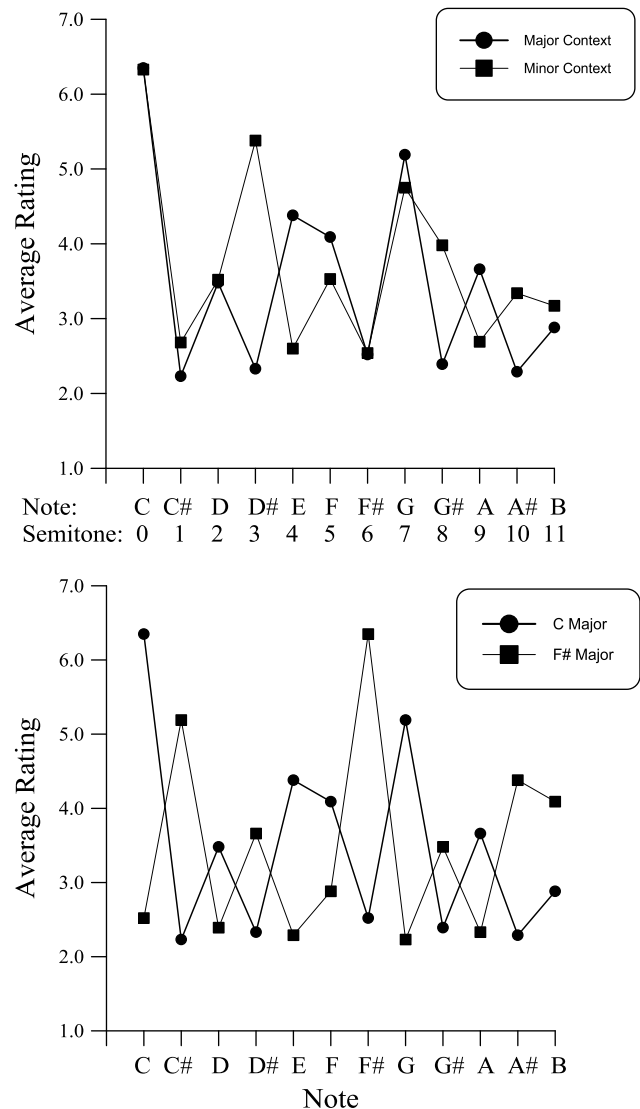


Figure 1. Krumhansl and Kessler (1982) tonal hierarchies: C major and minor (top) and C and F# major (bottom).

1994; Browne, 1981; Butler, 1989, 1990; Butler & Brown, 1994), and psychological (Huron & Parncutt, 1993; Krumhansl, 1990, 2000b; Krumhansl & Schmuckler, 1986; Krumhansl & Toiviainen, 2000, 2001; Longuet-Higgins & Steedman, 1971; Toiviainen & Krumhansl, 2003) perspectives. Although a thorough review of this work would require a study in and of itself (see Krumhansl, 2000a; Toiviainen & Krumhansl, 2003, for such reviews), currently, one of the most influential approaches to key-finding focuses on variation in the pitch content of a passage (e.g., Huron & Parncutt, 1993; Krumhansl & Schmuckler, 1986; Longuet-Higgins & Steedman, 1971; Parncutt, 1989; Shmulevich & Yli-Harja, 2000; Vos & Van Geenan, 1996). Perhaps the best known of such models is the key-finding algorithm of Krumhansl and Schmuckler (1986, described in Krumhansl, 1990). This algorithm operates by pattern matching the tonal hierarchy values for the different tonalities with statistical properties of a musical sequence related to note occurrence and/or duration.

Specifically, the key-finding algorithm begins with the creation of an input vector, which consists of a 12-element array representing values assigned to the 12 chromatic scale tones. The input vector is then compared with stored representations of the 24 tonal hierarchy vectors, taken from Krumhansl and Kessler (1982), resulting in an output vector quantifying the degree of match between the input vector and the tonal hierarchy vectors. Although the most obvious measure of match is a simple correlation, other measures, such as absolute value deviation scores, are possible. Correlation coefficients are convenient in that they are invariant with respect to the range of the input vector and have associated statistical significance tables.

One of the strengths of the key-finding algorithm is that parameters pertaining to the input and output vectors can vary depending on the analytic application. For example, if the goal is to assign a single key to a piece of music, then the input vector can be based on the first few elements of the piece, the ending of that same piece, or the summed duration of all notes in the piece. Possible output measures are the key with the highest correlation in the output vector, along with the magnitude of this correlation, as well as other keys with significant correlations. Alternatively, if the goal is to detect a change in tonality across sections of a piece, then input vectors could contain either the beginnings of the different sections or the sections in their entirety, with the output again consisting of the identity and magnitude of the key with the highest correlation. The algorithm can also be used to trace key movement throughout a piece. In this case, the input could be based on various *windows*, made up of summed durations of notes occurring within these windows. Again, the output measure of interest might be either the single highest correlation or the pattern of correlations across keys. Finally, the algorithm can be used to determine listeners' percepts of tonality arising from a psychological study. In this case, the input could be goodness-of-fit ratings, memory scores for tones, or even reaction times to tones, with the output being some combination of the keys with significant correlations or the pattern of correlations across keys.

Initially, Krumhansl and Schmuckler explored the key-finding algorithm's ability to determine tonality in three applications (see Krumhansl, 1990). In the first application, the algorithm predicted tonalities for sets of preludes by Bach, Shostakovich, and Chopin, using only the initial segments of each prelude. In the second application, the algorithm determined the tonality of fugue subjects by

Bach and Shostakovich on a sequential, note-by-note basis. Finally, in the third application, the algorithm traced tonal modulation throughout an entire Bach prelude (No. 2 in C Minor; *Well-Tempered Clavier, Book II*), comparing the algorithm's behavior with analyses of key movement provided by two expert music theorists.

Without going into detail concerning the results of these tests (see Krumhansl, 1990, pp. 77–110, for such a review), in general the algorithm performed successfully in all three applications. For example (and deferring a more specific description until later), in Application I, the algorithm easily determined the designated key of the prelude on the basis of the first few note events for the Bach and Shostakovich preludes. Success on the Chopin preludes was more limited, with this modest performance providing support for stylistic differences typically associated with Chopin's music. When determining the key of the fugue subjects (Application II), the algorithm was able to find the key in fewer notes than other models that have been applied to this same corpus (e.g., Longuet-Higgins & Steedman, 1977).¹ Finally, the key-finding algorithm closely mirrored the key judgments of the two music theorists (Application III), although the fit between the theorists' judgments was stronger than was the algorithm to either theorist. Thus, in its initial applications, the key-finding algorithm proved effective in determining the tonality of musical excerpts varying in length, position in the musical score, and musical style.

Although successful as a (quasi-) music-theoretic analytic tool, it is another question whether the key-finding algorithm can capture the psychological experience of tonality. Although not directly answering this question, a number of psychological studies over the years have used the Krumhansl–Schmuckler algorithm to quantify the tonal implications of the stimuli in their experiments. For example, Cuddy and Badertscher (1987) used a variant of the Krumhansl–Schmuckler algorithm to determine the tonal structure of ratings provided to the 12 tones of the chromatic scale when these tones were presented as the final note of melodic sequences. In a similar vein, Takeuchi (1994) used the maximum key-profile correlation (MKC) of the Krumhansl–Schmuckler algorithm to (successfully) predict tonality ratings and (unsuccessfully) predict memory errors for a set of melodies.² In a more thorough test of the algorithm's ability to predict memory errors, Frankland and Cohen (1996) used the key-finding procedure to quantify the degree of tonality of a short sequence of tones interpolated between a standard and comparison tone. They found that memory accuracy and, to a lesser extent, reaction time were well predicted by models of the stimuli that incorporated the implied tonality and tonal strength of the initial standard tone and the interpolated sequence.

¹ Vos and Van Geenan (1996) have also used Bach's fugue subjects as a basis for comparing their key-finding model with the Krumhansl–Schmuckler algorithm. Despite their claims to the contrary, Vos and Van Geenan's model performs, at best, on par with the Krumhansl–Schmuckler algorithm and, in many ways, not as efficiently.

² Takeuchi (1994) does note that design parameters may have limited the algorithm's effectiveness with reference to memory effects. Specifically, in Takeuchi's study, altered versions of standard melodies were created by changing one tone in a seven-note sequence, with this change constrained such that it did not modify either the pitch contour or the MKC for the altered version. As such, the ability of the MKC for predicting memory errors was assessed only indirectly, by having a range of melodies that varied in their MKC.

Unfortunately, none of this work directly tested the algorithm's ability to model listeners' percepts of tonality (but see Toiviainen & Krumhansl, 2003, for recent work on this issue). In this regard, any number of questions could be considered. For example, given a corpus of short musical passages varying in their tonality, do the algorithm's predictions of musical key reflect listeners' percepts of tonality with these same segments? Or, when presented with a longer, more extended musical passage, can the algorithm track listeners' percepts of key movement, or what is called *modulation*? In this latter case, if the algorithm can model listeners' percepts of tonality, is there an optimum window size for creating the input vector? Given its nature, too small a window runs the risk of producing highly variable judgments of tonality (Krumhansl, 1990; Temperley, 1999), whereas too large a window does violence to temporal order information (e.g., Butler, 1989) and potentially masks key movement or secondary key influences.

The goal of these studies was to examine such questions concerning the key-finding algorithm's ability to predict listeners' percepts of tonality. It should be noted, though, that, despite being presented within the context of musical cognition research, the results of these studies have broad implications for issues pertaining to basic perceptual processing. For example, the algorithm's performance speaks to issues concerning the viability of pattern matching in perceptual recognition and identification. Pattern matching has been, over the years, much maligned in perceptual theory, and as such, evidence that a pattern-matching process successfully models musical understanding provides renewed support for this seemingly (too) simple process. In a different vein, issues revolving around the window for producing the input vector provide insight into constraints on perceptual organization and the amount and type of information that can be integrated into a psychological unit. These issues, as well as others, are returned to in the General Discussion.

Experiment 1: Perceived Tonality of the Initial Segments of Bach and Chopin Preludes

As already described, the initial application of the key-finding algorithm involved determining the tonality of the beginning segments of preludes by Bach, Shostakovich, and Chopin, with the algorithm successfully determining the key of the Bach and Shostakovich preludes but having some difficulty with the Chopin excerpts. This variation in performance provides an ideal situation for a test of the algorithm's ability to predict listeners' percepts of musical key. Quite generally, two issues can be examined. First, does the algorithm effectively mimic listeners' perceptions of key, predicting both correct tonal identifications as well as failures of tonal determination? Second, when listeners and the algorithm fail to identify the correct key, do both fail in the same manner? That is, does the algorithm predict the content of listeners' incorrect identifications? To explore these questions, this experiment looked at percepts of tonality for the first few notes of preludes by Bach and Chopin.

An additional impetus for looking at such musical segments is that there is, in fact, some evidence on listeners' percepts of tonality as induced by the opening passage of the Bach preludes. Cohen (1991) presented listeners with passages from the first 12 (of 24) preludes from Bach's *Well-Tempered Clavier, Book I*, and had listeners sing the first musical scale that came to mind after

hearing these segments. Cohen used four different excerpts from each prelude, including the first four note events (i.e., successive note onsets) in the prelude, the first four measures, the first eight measures, and the final four measures. Analyses of these vocal responses indicated that, for the first four note events, listeners chose the "correct" tonality (the tonality as indicated by Bach) as their dominant response mode. When hearing either the first four or eight measures, however, there was considerably less congruence in perceived tonality, suggesting that listeners might have been perceiving key movement in these segments. Finally, the last four measures again produced strong percepts of the correct key, at least for the preludes in a major key. For the minor preludes, the last four measures produced a strong response for the major tonality built on the same tonic (e.g., a response of C major for the last four measures of the prelude in C minor). Far from being an error, this finding likely represents listeners' sensitivity to what music theorists call the *Picardy third*, or the practice of baroque composers of ending pieces written in a minor key with a major triad.

Thus, Cohen's (1991) study suggests that listeners can perceive the tonality of musical segments based on only the first few notes. Given that this study was limited in the number of tonalities examined (only 12 of the 24 preludes were tested), the musical style explored (the baroque music of Bach), and through the use of a production, as opposed to a perception, response measure (see Schmuckler, 1989, for a discussion of the limitations of production measures), the current experiment provided a more thorough assessment of listeners' perceptions of key in response to short musical fragments. As such, ratings of tonality in response to the first few notes of 24 preludes by Bach and 24 preludes by Chopin were gathered. These ratings were then compared with tonality judgments predicted by the algorithm.

Method

Participants. Twenty-four listeners participated in this study. All listeners were recruited from the student population (mean age = 20.4 years, $SD = 1.7$) at the University of Toronto at Scarborough, receiving either course credit in introductory psychology or \$7.00 for participating. All listeners were musically sophisticated, having been playing an instrument (or singing) for an average of 7.4 years ($SD = 3.1$), with an average of 6.2 years ($SD = 3.4$) of formal instruction. Most were currently involved in music making ($M = 3.7$ hr/week, $SD = 4.4$), and all listened to music regularly ($M = 17.5$ hr/week, $SD = 14.6$). All listeners reported normal hearing, and none reported perfect pitch. Some of the listeners did report familiarity with some of the passages used in this study, although subsequent analyses of these listeners' data did not reveal any systematic differences from those listeners reporting no familiarity with the excerpts.

Stimuli and equipment. Stimuli were generated with a Yamaha TX816 synthesizer, controlled by a 486-MHz computer, using a Roland MPU-401 MIDI interface. All stimuli were fed into a Mackie 1202 mixer and were amplified and presented to listeners by means of a pair of Boss MA-12 micromonitors.

All stimuli were derived from the preludes of Bach's *Well-Tempered Clavier, Book I*, and the Chopin preludes (Op. 28) and consisted of a series of context passages, each with a set of probe tones. Sample contexts for some of these preludes are shown in Figure 2. All 24 preludes (12 major and 12 minor) for Bach and Chopin were used as stimuli, producing 48 contexts in all. Based on Application I of the Krumhansl-Schmuckler algorithm, the context passages consisted of the first four (or so) notes of

Design and procedure. Listeners were informed that there would be a series of trials in which they would hear a short context passage followed 1 s later by a 600-ms probe note. Their task was to rate, using a 7-point scale, how well the probe tone fit with the tonality of the context in a musical sense (1 = *tone fit very poorly*, 7 = *tone fit very well*). Listeners typed their response on the computer keyboard and then pressed the enter key to record their response. After listeners responded, the next trial began automatically.

Time constraints prevented presenting listeners with all 576 context–probe pairings. Accordingly, half of the listeners heard the Bach contexts–probes, and the remaining listeners heard the Chopin contexts–probes, resulting in 288 (24 preludes \times 12 probes) experimental trials per listener. These trials were arbitrarily divided into four blocks of 72 trials, with a different random ordering of trials for all listeners. Prior to the first block, listeners had a practice session to familiarize them with the rating task and the structure of the experimental trials. For the practice trials, two idealized major (C and F#) and minor (E and Bb) contexts (e.g., diatonic scales) were created, with the typical 12 probe tones associated with each context. For the practice block, 10 of the possible 48 trials (4 tonalities \times 12 probes) were picked at random. The total session of 298 trials took about 1 hr. After the study, listeners completed a music background questionnaire and were debriefed as to the purposes of the study.

Results

As a first step in data analysis, intersubject correlation matrices were calculated, aggregating across the 12 major and 12 minor preludes for each composer. For the Bach contexts, listeners' ratings were generally interrelated, mean intersubject $r(286) = .236$, $p < .001$. Of the 66 intersubject correlations, only 1 was negative, and 12 were statistically insignificant (2 additional correlations were marginally significant; $ps < .07$). Although statistically significant, on an absolute basis these correlations are worryingly low. It should be remembered, though, that listeners only provided a single rating for each context–probe pairing; hence, a certain degree of variability is to be expected on an individual subject basis. Ratings for the Chopin preludes were considerably more variable, mean intersubject $r(286) = .077$, *ns*. Of the 66 intersubject correlations, 12 were negative and 43 were statistically insignificant (with 3 additional correlations marginal). Although also worrisome, that the Chopin intersubject correlations were less reliable than the Bach correlations could indicate that these stimuli were more ambiguous in their tonal implications.

As a next step in data analysis, the probe-tone ratings were analyzed using an analysis of variance (ANOVA) to determine whether the different tonal contexts in fact induced systematically varying ratings on the different probe tones. Such an analysis is particularly important given the variability of the Bach and Chopin ratings on an individual-subject basis. Toward this end, a four-way ANOVA, with the between-subjects factor of composer (Bach vs. Chopin) and the within-subject factors of mode (major vs. minor), tonic note (C, C#, D, D#, E, F, F#, G, G#, A, A#, B), and probe tone (C, C#, D . . . , B) was conducted on listeners' ratings. The output of this omnibus ANOVA is complex, revealing myriad results. In terms of the main effects, the only significant result was for probe tone, $F(11, 242) = 5.81$, $MSE = 4.18$, $p < .01$. None of the remaining main effects were significant, although the main effect of composer was marginal, $F(1, 22) = 3.87$, $MSE = 93.68$, $p < .06$. Of the two-way effects, the interactions between composer and tonic note, $F(11, 242) = 2.14$, $MSE = 3.55$, $p < .05$, composer and probe tone, $F(11, 242) = 4.40$, $MSE = 4.18$, $p <$

$.001$, and tonic note and probe tone, $F(121, 2552) = 7.16$, $MSE = 2.27$, $p < .01$, were all significant. For the three-way effects, the interactions between composer, tonic note, and probe tone, $F(121, 2662) = 3.24$, $MSE = 2.27$, $p < .01$, and composer, mode, and probe tone, $F(121, 2662) = 2.26$, $MSE = 2.19$, $p < .001$, were both significant. Finally, and most important, all of these effects are qualified by the significant four-way interaction among all factors, $F(121, 2662) = 1.84$, $MSE = 2.19$, $p < .001$. Essentially, this result reveals that despite the variability in ratings for individual participants, listeners' ratings for the probe tones did, in fact, vary systematically as a function of the tonality of the context passage regardless of whether the context passage was in a major or minor mode and whether the context was drawn from the Bach or Chopin preludes. Thus, this finding validates that the different context passages, even though short in duration, nevertheless induced different tonal hierarchies in listeners. The nature, including the identity (e.g., which tonality) and the strength of these tonal percepts, is another question, one that is addressed by the next analyses.

The primary question to be answered by the next analyses involves whether listeners' tonal percepts, as indicated by their probe-tone ratings, were predictable from the tonal implications of these same contexts as quantified by the key-finding algorithm. As a first step, probe-tone ratings for each key context for each composer (e.g., all ratings following Bach's C major prelude, Bach's C# major prelude, Chopin's C major prelude, etc.) were averaged across listeners and then used as the input vector to the key-finding algorithm. Accordingly, these averaged ratings were correlated with the 24 tonal hierarchy vectors, resulting in a series of output vectors (for each context for each composer) indicating the degree of fit between the listeners' probe-tone ratings and the various major and minor tonalities. These rating output vectors can then be scrutinized for any number of properties.

The first aspect to be considered is whether or not listeners perceived the "intended" tonality for each prelude. *Intended* refers to the key that the composer meant to be invoked by the prelude and is indicated by the key signature of the composition. To examine this question, the correlation corresponding to the intended key in the listeners' rating output vector was compared with the correlation for the same key taken from the algorithm's output vector. Table 2 presents the results of the correlations with the intended key for all of the Bach and Chopin contexts.

Table 2 reveals that for the Bach preludes, listeners were quite successful at determining the intended tonality of the excerpts. For 23 of the 24 preludes, the correlation with the intended key was significant, and for 21 of the 24 preludes this correlation was the highest positive value in the output vector. The algorithm was similarly successful in picking out the intended key. All 24 correlations with the intended key were significant; this correlation was highest in the output vector for 23 of 24 contexts.

In contrast, performance with the Chopin preludes was much more variable. Listeners' probe-tone ratings correlated significantly with the intended key for 8 of the 24 preludes; this value was the highest key correlation in only 6 of 24 contexts. The algorithm was similarly limited in its ability to pick out the tonality, determining the intended key in 13 of 24 cases; this correlation was the maximum value for 11 of 24 preludes.

Given that both the key-finding algorithm and the listeners experienced difficulty in determining the intended key of the

Table 2

Correlations for the Intended Key, as Indicated by Composer, Taken From the Rating and Algorithm's Output Vectors for Both Bach and Chopin

Prelude	Bach		Chopin		Chopin output
	Algorithm	Listeners	Algorithm	Listeners	Vector correlations
C major	.81***	.91***	.81***	.69**	.91***
C minor	.92***	.87***	.88***	.83***	.92***
C#/D♭ major	.83***	.88***	.82***	.68**	.88***
C#/D♭ minor	.87***	.72***	.25	-.09	.37
D major	.73***	.82***	.27	.61**	.48**
D minor	.81***	.91***	.76***	.32	.40*
D#/E♭ major	.87***	.78***	.59**	.39	.38
D#/E♭ minor	.92***	.90***	.71***	.42	.69***
E major	.83***	.85***	.88***	.41	.61***
E minor	.92***	.80***	.55	.45	.68***
F major	.83***	.75***	.76***	.74***	.73***
F minor	.85***	.57**	.00	-.10	.30
F#/G♭ major	.67**	.76***	.88***	.56	.69***
F#/G♭ minor	.93***	.61**	.38	.58**	.87***
G major	.83***	.74***	.79***	.67**	.91***
G minor	.85***	.84***	.21	-.07	.10
G#/A♭ major	.87***	.73***	.76***	.30	.71***
G#/A♭ minor	.83***	.67**	.85***	-.01	.10
A major	.73***	.74***	.49	.58**	.76***
A minor	.82***	.88***	-.08	.41	-.23
A#/B♭ major	.88***	.52	.53	.55	.73***
A#/B♭ minor	.91***	.88***	.18	.00	-.05
B major	.68**	.71***	.38	.03	.69***
B minor	.83***	.60**	.92***	.14	.51**

* $p = .05$. ** $p < .05$. *** $p < .01$.

Chopin preludes, it is of interest to examine more closely both sets of output vectors to determine whether the listeners and the algorithm behaved similarly irrespective of whether or not the intended key was found. This issue was explored in two ways. First, the correlations with the intended key for the algorithm and the listeners (see, e.g., the third and fourth columns of Table 2) were themselves correlated and were found to be significantly related, $r(22) = .46, p < .05$. Figure 3 presents this relation graphically and reveals that those preludes in which the algorithm failed to find the key tended to be those in which listeners similarly failed to determine the key, and vice versa,³ although there are clearly deviations between the two sets of correlations.

Second, the complete pattern of tonal implications for each Chopin prelude was examined by correlating the algorithm's output vector, which represents the goodness of fit between the musical input and all major and minor keys, with the listeners' output vector. Although such correlations must be treated with some degree of caution, given that the values in the output vectors are not wholly independent of one another (this issue is explored further in the *Discussion* section), the results of these correlations (see Table 2) are nonetheless intriguing. Of the 24 correlations between output vectors, 16 were significant, with 1 additional marginally significant correlation; aggregating across all preludes, the output vectors for the algorithm and the listeners were significantly related, $r(574) = .58, p < .001$. Not surprisingly, the strength of the relation between the algorithm's and listeners' output vectors was related to the strength of the correlation for the intended key, with the value of the intended key correlation for algorithm and listeners significantly predicting the output vector

correlation ($R = .76, p < .001$); both factors contributed significantly to this prediction, β (algorithm) = .38, $p < .05$, and β (listeners) = .51, $p < .01$.

Discussion

The primary result arising from this study is that the Krumhansl-Schmuckler key-finding algorithm provided, at least for the Bach preludes, a good model of listeners' tonal percepts of short musical fragments, with the intended key of the preludes both predicted by the algorithm and perceived by the listeners. What is most striking about this finding is that both algorithm and listeners (in the aggregate) made accurate tonal responses based on remarkably scant information: the first four to five note events, lasting (on average) less than 1 s. Such a result speaks to the speed with which tonal percepts establish themselves.

The results from the Chopin preludes are also illuminating albeit more complex. At first glance, it seems that these findings represent poor tonal determination by the algorithm and hence underscore a significant weakness in this approach. It is important to remember, though, that even though the intended key was not

³ A comparable analysis of the Bach preludes failed to reveal any relation between intended key correlations for the listeners and the algorithm, $r(22) = .0001, ns$. Although worrisome at first blush, it must be remembered that both listeners and algorithm were uniformly successful in tonal determinations with these preludes; thus, the lack of a correlation in the strength of these determinations might simply reflect ceiling performance.

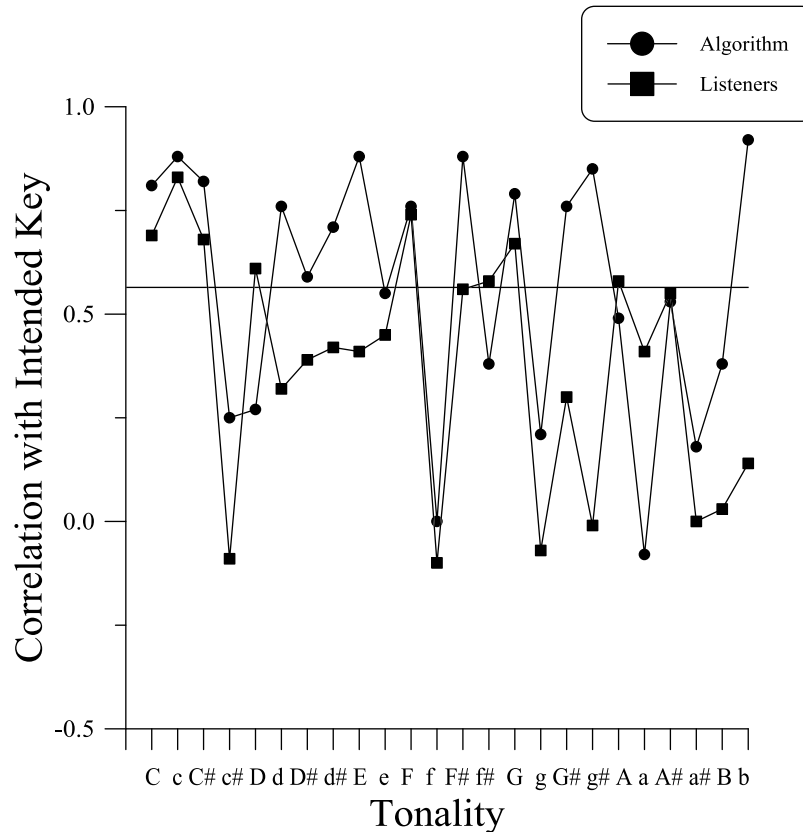


Figure 3. Intended key correlations for algorithm and listeners' probe-tone ratings for the Chopin contexts.

picked out by the algorithm it still performed comparably to listeners' actual judgments of tonality. Thus, situations in which the algorithm failed to determine the intended key were also those in which listeners failed to perceive the intended tonality. Accordingly, the Krumhansl–Schmuckler key-finding algorithm might be actually picking up on what are truly tonally ambiguous musical passages.

Before accepting such an interpretation, however, it is important to discuss a problematic issue with the means used for assessing the fit between listeners' tonal perceptions and the algorithm's tonal predictions for the Chopin preludes. Specifically, this fit was measured by comparing the output vector value for the intended key for both algorithm and listeners and by correlating both listeners' and algorithm's output vectors. Although the first of these measures is noncontroversial, the problem with this second assessment procedure is that, because the ideal key profiles are simply permutations of the same set of ratings, the correlations with each individual key making up the output vector are themselves largely nonindependent.⁴ This lack of independence between the individual values of the output vector has the rather unfortunate consequence of spuriously raising the correlation coefficients when two output vectors are themselves correlated.

One method of addressing this issue would be to adopt a stricter criterion for statistical significance when comparing the correlation between two output vectors. Thus, for instance, an alpha level of .01, or even .001, might be used for indicating statistical significance. A drawback to this solution, though, is that this new

alpha level is chosen somewhat arbitrarily, not based on any principled reason.

As an alternative means of assessing the importance of output vector correlations, one can compare the obtained output vector correlations with correlations for two input vectors with a known level of musical relatedness. Although this method relies on an intuitive judgment for assessing the fit between output vectors, comparing output vector correlations with output vectors for well-known musical relations is nonetheless informative. Accordingly, output vectors were generated using different ideal key profiles themselves as input vectors. These output vectors were then correlated to give a sense of how strong a correlation might be expected for two input vectors embodying a well-known musical relation. Table 3 lists the results of a number of such comparisons. As seen in Table 3, the strongest correlation between output vectors, .804, is found for input vectors comprising a major tonic and its relative minor (or a minor tonic and its relative major), such as C major and A minor (or C minor and E \flat major). Next comes the correlation for input vectors comprising a major tonic and its dominant, such as C major and G major, with an output vector correlation of .736. Using correlations such as these as a yardstick,

⁴ These values do have some degree of independence, given that the output vector consists of correlations for both major and minor keys. The ideal profiles for the major and minor keys, although cyclical permutations within each set, are themselves independent.

Table 3
Correlations Between Output Vectors Based on Calculating the Krumhansl–Schmuckler Key-Finding Algorithm Using Different Ideal Key Profiles as the Input Vectors

Input vector 1	Input vector 2	Correlation between output vectors
Major tonic	Major dominant	.736
Minor tonic	Minor dominant	.614
Major (minor) tonic	Relative minor (major)	.804
Major (minor) tonic	Parallel minor (major)	.514

it might be understood that output vector correlations of greater than .736 or .804, then, represent a reasonably compelling degree of fit.⁵

Using these correlations as a basis of comparison, it is instructive to note that all of the significant output vector correlations shown in Table 2 were at least as strongly related as the relation between a major key and its parallel minor (two reasonably strongly related musical tonalities), and a good many were more strongly correlated than the two closest musical relations: that between a major tonality and its relative minor and a major tonality and its dominant. Thus, this comparison indicates that even though the output vector correlations are inflated because of lack of independence, they are nevertheless still informative as to the relative strength of the relation between input vectors giving rise to the output vectors.

The question remains, of course, as to what it was about some of the Chopin preludes that led both listeners and algorithm to fail in determining the key. Clearly, one potentially important factor is stylistic: Chopin's music makes heavy use of the chromatic set (all 12 notes of the scale). Accordingly, it could well be that the tonal implications of Chopin's music might be quite subtle, with the consequence that more information is thus needed (by both listeners and algorithm) to clearly delineate tonality, although it must be recognized that the length of these contexts makes the communication of significant stylistic information somewhat improbable. Nevertheless, it still might be that four to five note events are simply not sufficient to reliably determine tonality for Chopin's music. Examining this issue is one of the goals of Experiment 2.

Experiment 2: Perceived Tonality in Extended Chopin Contexts

Experiment 1 demonstrated that tonality could be determined based on limited musical information, at least for the Bach preludes. In contrast, key determination was much more difficult for the Chopin preludes, presumably because of stylistic factors that necessitate more information for tonal determination with the Chopin preludes.

To make matters even more complex, a closer examination of the Chopin preludes reveals that these preludes do not uniformly instantiate the tonality of the intended, or home, key in their initial segments. Rather, the instantiation of the intended key is more complex, showing a variety of emerging patterns. One pattern involves the straightforward, consistent instantiation of the home key. Preludes such as the A \flat major and B minor, shown in the top panel of Figure 4, provide exemplars of this pattern. As an aside, given that these preludes do present clear key-defining elements, it

is a mystery as to why listeners failed to perceive this key in Experiment 1. One potential explanation for this result might simply have to do with the fact that the actual context patterns heard by listeners for these preludes were, in fact, quite short. Based on the tempo of these preludes, each context actually lasted less than 500 ms. Thus, it might be that listeners simply required more time to apprehend these otherwise clear key contexts.

Different patterns of tonal instantiation are also possible. The contexts for the A major and B major preludes appear in the second panel of Figure 4. Inspection of these passages reveals that, although the intended key eventually becomes clear, it does not do so immediately. Thus, for these preludes, it might be that listeners need a longer context to determine the key. A third pattern of tonal instantiation can be seen in the third panel of Figure 4 and represents the converse of the previous example. In this case, although the home key is presented initially, the tonal implications of the music quickly move toward a different key region. As described earlier, such movement is called modulation and is a common aspect of Western tonal music. Chopin's E major and C# minor preludes provide examples of this pattern.

A final pattern of tonal implications appears in the bottom panel of Figure 4 and is exemplified by the preludes in A minor and F minor. For these preludes, there truly does not appear to be any clear sense of key that develops, at least initially. This suggests that, for contexts such as these, listeners' tonal percepts might indeed remain ambiguous as to key.

Application of the key-finding algorithm to these contexts confirms these intuitions of how the tonality of these passages develops over time. For this application, the input vector included all of the duration information for each musical measure (all of the musical notes occurring between the vertical lines) at one time, with no overlap of duration information between the measures. Figure 4 also shows the results of the key-finding algorithm's tonal analysis of each of these measures in terms of the correlations with the intended key and reveals reasonably good correspondence with the earlier descriptions of key development.

This application of the key-finding algorithm suggests an interesting extension of the earlier test, with the algorithm now predicting how listeners' sense of key might change as the music

⁵ An analogous analysis involves mapping different input vectors onto a two-dimensional representation of Krumhansl and Kessler's (1982) four-dimensional torus model of key space, using Fourier analysis procedures (see Krumhansl, 1990; Krumhansl & Schmuckler, 1986, for examples of this approach). Such an analysis reveals that, for instance, the distance between a major tonic and its dominant is 79.2°, whereas the distance between a major tonic and its relative minor is 49.6° (the maximum distance in key space is 180°). In fact, these two procedures (correlating output vectors of ideal key profiles and comparing distance in key space) produce comparable patterns. Mapping into key space is intriguing in that it avoids the problem of the nonindependence of the output vector values, given that this analysis operates on the input vectors themselves. As such, it does provide some evidence that the correlational values, although inflated by their lack of independence, nevertheless represent a reasonable means of assessing relative degrees of relatedness between the key implications of various input patterns. Unfortunately, though, distance in key space is not as easy to grasp intuitively as correlating output vectors, nor does it have any means for assessing fit in a statistical sense. As such, the output vector correlation approach was adopted in the current context.

Chopin, A \flat Major Prelude

Algorithm: .76** .76** .44 .74**

Chopin, B Minor Prelude

Algorithm: .90** .92** .92** .89**

Chopin, A Major Prelude

Algorithm: .42 .37 .83** .88**

Chopin, B Major Prelude

Algorithm: .45 .43 .92** .86**

Chopin, E Major Prelude

Algorithm: .92** .22 .88** .16

Chopin, C# Minor Prelude

Algorithm: .90** .95** .32 .36

Chopin, A Minor Prelude

Algorithm: -.30 -.30 .13 -.15

Chopin, F Minor Prelude

Algorithm: .54 .54 .23 .22

** $p < .01$

Figure 4. Four-measure contexts for Chopin's A \flat major, B minor, A major, B major, E major, C# minor, A minor, and F minor preludes. Correlations with the intended key for each measure of the musical score, based on the Krumhansl-Schmuckler key-finding algorithm, are shown above each excerpt.

progresses. The goal of this experiment was to assess these predictions of tonal development using these contexts as stimuli in a probe-tone experiment.

Method

Participants. The data from a total of 16 listeners were ultimately used in this experiment; the data from 1 additional listener was discarded

because he failed to complete the study (as a result of experimental ennui). All listeners were recruited from the student population at the University of Toronto at Scarborough (mean age = 20.9 years, $SD = 1.8$) and received course credit in introductory psychology for participating. All listeners had some musical training, having played an instrument or sang for a mean of 7.9 years ($SD = 4.6$) and received formal instruction for a mean of 6.1 years ($SD = 3.3$). Most ($n = 12$) were currently involved in music making ($M = 2.8$ hr/week, $SD = 3.6$), and all listened regularly to music ($M = 16.2$

hr/week, $SD = 14.6$). All listeners reported normal hearing, and none reported perfect pitch. A few listeners reported that they did recognize some of the musical passages, although none named the actual pieces presented.

Stimuli, equipment, design, and procedure. All stimuli were generated and presented to listeners using the same equipment as in the previous study. Stimuli were derived from the Chopin preludes (Op. 28) and again consisted of a set of contexts and associated probe tones. Eight preludes were chosen for study (see Figure 4), including the preludes in $A\flat$ major, B minor, A major, B major, E major, C# minor, A minor, and F minor. For each prelude, four different contexts were produced. The first set of contexts consisted of Measure 1 of each prelude; the second set of contexts consisted of Measures 1 and 2 from each prelude; the third set of contexts consisted of Measures 1, 2, and 3 of each prelude; and the fourth set of contexts consisted of Measures 1, 2, 3, and 4 of each prelude. Figure 5 presents a sample of the contexts for the B minor prelude. Each context was associated with the 12 probe tones. The four different context sets (i.e., Measure 1, Measures 1–2, Measures 1–3, Measures 1–4) were considered four experimental conditions and were presented in a blocked fashion to

listeners in increasing order of length (i.e., all Measure 1 contexts in the first block, Measures 1–2 contexts in the second block, Measures 1–3 contexts in the third block, and Measures 1–4 contexts in the fourth block). Within each block the order of the different context and probe pairings for all preludes was randomized for each listener. Accordingly, listeners received four blocks of 96 trials, producing 388 trials in all. Listeners were given the same instructions as in the previous study and typed in their response (on a 1–7 scale) using the computer keyboard. The entire experimental session lasted approximately 30–45 min, after which listeners were debriefed.

Results

The first step in data analysis involved calculating intersubject correlations. As before, it is important to remember that, because listeners received only a single repetition of each context passage and probe, variability is expected in these rating profiles. Aggregating across the four blocks of trials, listeners' ratings were

Chopin's B Minor Prelude

Figure 5. Sample contexts for Chopin's B minor prelude, Measures 1–4, as a function of the different length conditions.

interrelated, and intersubject correlations ranged from $-.04$ to $.49$, mean intersubject $r(384) = .14$, $p < .006$. Of the 120 intersubject correlations, 79 were significant, and 9 were negative.

Next, listeners' probe-tone ratings were analyzed using a three-way ANOVA, with the within-subjects factor of context length (Measure 1, Measures 1–2, Measures 1–3, Measures 1–4), key (Ab major, B minor, A major, B major, E major, C# minor, A minor, and F minor), and probe tone (C, C#, D, D#, E, F, F#, G, G#, A, A#, B). As with the earlier study, the output of this omnibus ANOVA is quite complex. There were main effects for context length, $F(3, 45) = 3.96$, $MSE = 11.11$, $p < .05$; key, $F(7, 105) = 6.76$, $MSE = 14.17$, $p < .001$; and probe tone, $F(11, 165) = 11.25$, $MSE = 7.36$, $p < .001$. The two-way interaction between context length and key was also significant, $F(21, 315) = 3.84$, $MSE = 3.69$, $p < .001$, as was the interaction between key and probe tone, $F(77, 1155) = 4.84$, $MSE = 3.27$, $p < .001$. The interaction between context length and probe tone was not significant, $F(33, 495) = 0.69$, $MSE = 2.2$, *ns*. All of these results, however, were qualified by the significant three-way interaction among context length, key, and probe tone, $F(231, 3465) = 1.40$, $MSE = 2.10$, $p < .001$, which indicates that the ratings for the probe tones varied because of the length and tonality of the context. Again, this finding simply validates that the different contexts did indeed induce varying percepts in listeners. The next series of analyses investigated these percepts more closely.

As with Experiment 1, average probe-tone ratings were created for each key context (i.e., Ab major, B minor, A major, B major, E major, C# minor, A minor, and F minor) for each context length (i.e., Measure 1, Measures 1–2, Measures 1–3, Measures 1–4) and were then used as input vectors to the key-finding algorithm. Examination of the tonal implications of these passages then proceeded along two lines. First, the correlations with the intended key were compared on a measure-by-measure basis for the algorithm and the probe-tone ratings. Table 4 presents these correlations for all preludes, and Figure 6 graphs these correlations for the Ab major, B major, C# minor, and A minor preludes; each of these preludes represents one of the patterns of key influences described earlier. Overall, both Table 4 and Figure 6 show a reasonable correspondence between the predictions of the algorithm and the listeners' percepts of the intended key; aggregating across the four context lengths for each prelude, these correlations were themselves correlated, $r(30) = .78$, $p < .001$.

Second, the complete patterns of tonal implications for the preludes were examined by correlating the output vectors for the algorithm and the listeners' ratings. The results of these correlations, as a function of the different context lengths, appear in Table 5. Overall, there was strong correspondence between the output vectors, with all correlations significant. Moreover, output vectors were just as related for situations in which both listeners and algorithm found the intended key (e.g., Measures 1–3 and Measures 1–4 for the A and B major preludes, based on Table 3) as they were for situations in which one or the other found the intended key (e.g., Measure 1 for the B minor or E major; Measures 1–4 for the C# minor) and in which neither algorithm nor listeners found the intended key (e.g., all contexts for the A minor prelude). Finally, and with respect to the concern about the non-independence of values of the output vector, 22 of 32 of these correlations were greater than what would occur between a major tonic and its dominant (two highly related keys), and 31 of the 32 values exceeded the correlation between a tonic and its parallel minor (see Table 3). As such, it seems reasonable to conclude that, irrespective of the variation in the ability to perceive the intended key, both listeners and the algorithm picked up on comparable tonal implications.

Discussion

A number of intriguing findings arise from this study. First and foremost, the algorithm again was able to model listeners' percepts of tonality, both in situations in which a given tonality was clearly instantiated and when a passage was more unclear as to its key. Thus, the inability of the algorithm to determine the (intended) key of a specific passage does not necessarily imply a failing of the algorithm but rather indicates a true tonal ambiguity in the passage. Together, Experiments 1 and 2 strongly suggest that the key-finding algorithm can model listeners' percepts of tonality in situations of both tonal clarity and tonal ambiguity.

Second, the algorithm was able to track changes in listeners' percepts of musical key across extended passages. Of course, an important caveat to this result is that the contexts used here were still short (only four musical measures), and as such it is unclear whether the observed changes truly represent modulations of the global tonality or rather reflect changes in more local underlying harmonies without true key modulation.

Table 4
Correlations Between Listeners' Probe-Tone Ratings and the Algorithm's Key Predictions With the Intended Key for the Contexts of Experiment 2

Key	Measure 1		Measures 1–2		Measures 1–3		Measures 1–4	
	Algo	List	Algo	List	Algo	List	Algo	List
Ab major	.76***	.55	.76***	.60**	.44	.76***	.74***	.50
B minor	.90***	.48	.92***	.56	.92***	.76***	.89***	.55
A major	.42	.39	.37	.47	.83***	.85***	.88***	.68**
B major	.45	.42	.43	.46	.92***	.81***	.86***	.79***
E major	.92***	.50	.22	.03	.88***	.44	.16	.34
C# minor	.90***	.88***	.95***	.82***	.32	.52	.36	.72***
A minor	-.30	-.12	-.30	-.06	.13	.09	-.15	.04
F minor	.54	.50	.54	.47	.23	.54	.22	.61**

Note. Algo = algorithm; List = listener.
** $p < .05$. *** $p < .01$.

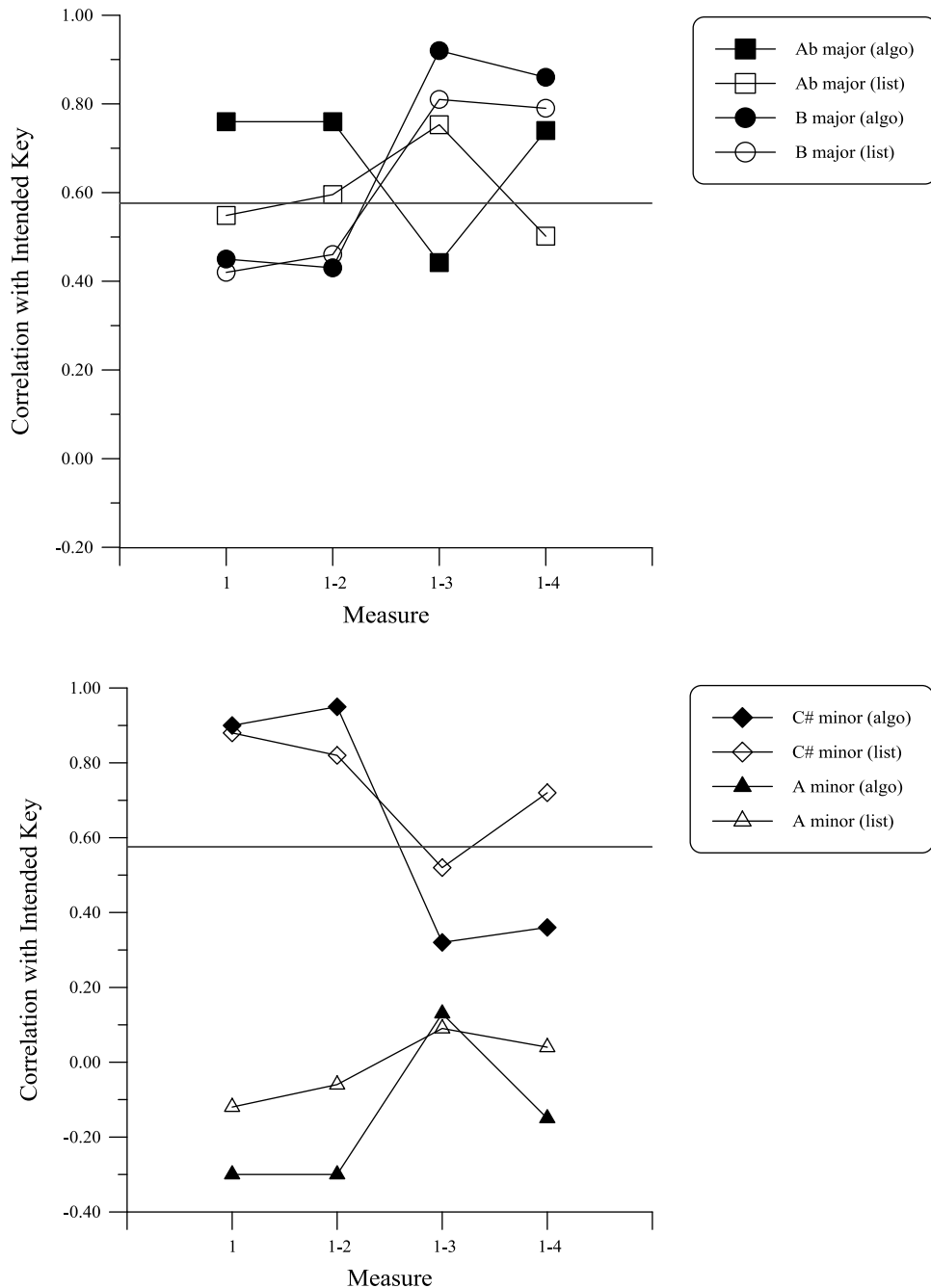


Figure 6. Correlations with the intended key for the algorithm (algo) and listeners' (list) ratings for the Ab major and B major (top) and C# minor and A minor (bottom) preludes. The .05 significance level is indicated.

Disentangling these two alternatives is, in fact, tricky, raising issues as to whether tonality is a global property of musical passages or whether it is better thought of as a local phenomenon, only involving percepts of relatively recent materials. The answer to this question has implications for how best to apply the algorithm. Should, for example, the algorithm operate on the entire set of music up to a given point in time, or is it more appropriate to use some form of weighted moving window of note durations, with the

algorithm only taking into account local note information? Exploring these questions was the goal of the final study in this series.

Experiment 3: Key Modulation in Chopin's E Minor Prelude

The primary goal of Experiment 3 was to examine the key-finding algorithm's ability to track listeners' percepts of key mod-

Table 5
Correlations Between the Output Vector Generated From the Listeners' Probe-Tone Ratings and the Output Vector Generated by the Key-Finding Algorithm for the Preludes of Experiment 2

Key	Measure 1	Measures 1–2	Measures 1–3	Measures 1–4
A♭ major	.54***	.78***	.45**	.87***
B minor	.82***	.70***	.92***	.80***
A major	.90***	.67***	.98***	.90***
B major	.90***	.71***	.88***	.95***
E major	.86***	.60***	.86***	.68***
C# minor	.95***	.92***	.80***	.86***
A minor	.74***	.58***	.80***	.84***
F minor	.51**	.71***	.89***	.80***

** $p < .05$. *** $p < .01$.

ulation throughout a single piece of music. Although Krumhansl and Schmuckler's third application of this algorithm (see Krumhansl, 1990) was specifically designed to investigate key modulation in an extended musical context, the algorithm was compared with two music theorists' key judgments and not with perceptual judgments. Thus, just as with Experiment 1, this study extends the key-finding algorithm's performance to listeners' explicit perceptual judgments.

Examination of the algorithm's ability to track the perception of key modulation is important in that a number of authors have specifically questioned whether a structural approach is robust enough to model this aspect of musical behavior. Butler (1989), for instance, although not critiquing the key-finding algorithm explicitly (likely because Butler's comments preceded the publication of the key-finding algorithm in Krumhansl, 1990), has pointedly questioned whether tonal hierarchy information (on which the key-finding algorithm is based) is useful in key-finding in general and in tracking modulation in particular: The key profile "does not describe how we hear moment-to-moment harmonic successions within a key; indeed, the theory offers no precise description of how we initially identify the key itself, or recognize key modulations" (Butler, 1989, p. 224). Other authors, such as Huron and Parncutt (1993) and Temperley (1999), have expressed similar, albeit more muted, reservations concerning the algorithm's ability to track key modulation. Temperley (1999), for instance, suggests that the algorithm is limited in that it only produces a single key judgment for a passage (as opposed to being sensitive to multiple key influences) and that it is insensitive to factors such as inertia in key movement (e.g., a tendency to remain in an already established key).

To date, there are two published accounts of such a test of the key-finding algorithm. Using the probe-tone method, Smith and Cuddy (Cuddy & Smith, 2000; Smith & Cuddy, 2003) utilized the key-finding algorithm as a model of tonality in their intensive investigation of multiple aspects of the opening phrase of Beethoven's *Waldstein* piano sonata (Op. 53) and found that the algorithm was able to model listeners' tonal perceptions across this phrase. Similarly, Krumhansl and Toiviainen (Krumhansl, 2000a, 2000b; Krumhansl & Toiviainen, 2001; Toiviainen & Krumhansl, 2003) developed a variant of the probe-tone procedure in which listeners provided a continuous rating (through the use of a computer slider) for an individual probe tone relative to an ongoing

passage of music. Using this technique, these authors demonstrated that a modified version of the key-finding algorithm (described in more detail subsequently) provided a good fit to listeners' perceptions of key movement in Bach's organ *Duetto* (BWV805). In this regard, the current study replicates these earlier results and extends these findings to a more chromatic musical style.

A secondary goal of this study was to explore the issue of global versus local influences on tonal perception. This question was examined by systematically varying the amount of musical information used to make up the input vector for the key-finding algorithm. Such a question is critical in that the choice of how the input vector is constructed (i.e., what segment of music is used to generate this vector) is the only free parameter in this model.

In exploring an extended musical passage, there are, in fact, any number of candidates for the segment of music that could make up the input vector. At one extreme, the input vector could consist of an aggregation of all of the tone durations that have occurred up to a given point in time. Such an approach provides a very global level of tonal analysis, one that takes into equal account everything that has occurred prior to a specific position. The disadvantage to this approach is that no special significance is attached to any prior event. Intuitively, though, it seems that recently occurring musical information should receive some priority in tonal determination, based on psychological influences such as recency effects (e.g., Murdock, 1962). Moreover, this form of aggregated information will also likely gloss over points of modulations. Such positions are interesting in that it is at these points that listeners' tonal sense might be strongly driven either by current or even upcoming events.

Accordingly, a more local method of generating the input vector may prove more effective. In this case, the input vector could be based on only the immediately preceding musical materials, although the question still remains as to how much information is to count as "immediately preceding." One could, for instance, use only the most recent musical events (i.e., the note or chord strictly preceding the position in question), or one could include a slightly larger amount of immediately preceding material. This latter strategy was used in Experiment 2, in which the input vector was based on all of the note durations of the single measure preceding the probe and not strictly on the final event(s) of that measure.

The previous two models represent extremes on a global versus local continuum; combinations of these two are also conceivable. One compromise would be to use more than just local information but less than the cumulative information. Accordingly, the input vector could consist of information from the previous few chunks of music, with the contributions of earlier chunks weighted to mimic deterioration or loss of information as a result of memory processes. Again, the size of the window for these chunks of musical information, as well as the weighting function used, are open questions.

All of these approaches have involved the inclusion of previously heard information. As a general category, such models can be considered memory models. Along with an influence from memory, music cognition research has highlighted the importance of the anticipation of upcoming events in the apprehension of music. Such work on musical expectancy (e.g., Bharucha, 1994; Bharucha & Stoeckig, 1986, 1987; Bigand et al., 1999; Bigand & Pineau, 1997; Bigand et al., 2003; Cuddy & Lunney, 1995; Jones, 1981, 1982, 1990; Krumhansl, 1995; Meyer, 1956; Narmour,

1989, 1990, 1992; Schmuckler, 1989, 1990, 1997a; Schmuckler & Boltz, 1994) has demonstrated myriad influences on the apprehension of, response to, and memory for musical materials (see Schmuckler, 1997a, for a review). Moreover, work on musical expectancy has demonstrated that tonality is a major determinant on the formation of expectations (e.g., Schmuckler, 1989, 1990). Accordingly, upcoming (not-yet-heard) information seems a likely candidate for inclusion into a model of tonal determination.

Finally, of course, both memory and expectancy factors can be combined with current information for the determination of tonality. Once again, the exact size of the windows for memory, current, and expectancy information, along with the relative weighting of these components, is unclear. Nevertheless, a model combining these features may well prove quite powerful in predicting listeners' percepts of tonality. Exploring these issues involving the modeling of tonality was a secondary goal of this study.

Method

Participants. A total of 16 listeners participated in this study. All listeners were recruited from the student population at the University of Toronto at Scarborough (mean age = 20.3 years, $SD = 1.5$) and received either course credit in introductory psychology or \$10.00 for participating. All listeners had some musical training, having played an instrument or sang for a mean of 10.1 years ($SD = 3.6$) and received formal instruction for a mean of 7.3 years ($SD = 3.1$). Ten of the 16 listeners were currently involved in music making ($M = 5.3$ hr/week, $SD = 2.5$), and all listened regularly to music ($M = 16.9$ hr/week, $SD = 14.9$). All listeners reported normal hearing, and none reported perfect pitch. In contrast to the previous studies, a majority of the listeners ($n = 11$) reported that they recognized the stimulus passage.

Stimuli, equipment, design, and procedure. All stimuli were generated and presented to listeners using the same equipment as in the previous studies. Stimuli were derived from Chopin's E minor prelude (Op. 28, no. 4; see Figure 7). In this passage, a quartertone (a single beat) had a duration of 1,000 ms. Eight probe positions were chosen for study within this prelude, based on interesting harmonic or tonal changes occurring at these points; these positions are notated in Figure 7. Eight contexts were thus generated on the basis of these probe positions, with the entire musical passage up to, but not including, the labeled probe position constituting the different contexts. One second after the context, listeners heard 1 of the 12 possible probe tones; these probes were sounded for 600 ms. The timbre of the context passage was the piano sound used in the previous two studies, whereas the timbre for the probe tones consisted of octavely related sine waves, which produced a well-defined sense of pitch chroma but not pitch height.

Listeners first heard all of the context-probe pairs for Probe Position 1, then those for Probe Position 2, and so on, up to Probe Position 8; each probe position constituted a single experimental block. All listeners received different random orderings of trials within each block. Prior to the experimental blocks, listeners heard 5 practice trials consisting of a single octave ascending and descending major scale followed by 5 randomly chosen probe tones. Overall, listeners received 101 trials (5 practice trials plus 8 blocks of 12 probe tones).

Listeners rated how well the probe tone fit with the tonality of the context passage using the same rating scale as in the previous studies. Listeners wrote their ratings on an answer sheet and were given 4 s to make each response. The experimental session lasted about 75–90 min, after which listeners were debriefed as to the purposes of the study.

Results

Once again, preliminary data analysis involved calculating intersubject correlations and exploring probe-tone ratings as a func-

tion of the various experimental manipulations. Aggregating across the eight blocks of trials, listeners' ratings were only weakly related, with intersubject correlations ranging from $-.22$ to $.45$, mean of $r(94) = .08$, *ns*. Of the 120 correlations, only 23 were significant. Although alarmingly low, it should once again be remembered that listeners heard each context-probe pair only once; hence, a fair degree of intersubject variability, although unfortunate, is to be expected.

Probe-tone ratings were then examined in a two-way ANOVA, with the within-subject factors of context (Probe Position 1 through Probe Position 8) and probe tone (C, C#, D . . . , B). This ANOVA revealed main effects for both context, $F(7, 105) = 3.53$, $MSE = 4.02$, $p < .01$, and probe tone, $F(11, 165) = 6.18$, $MSE = 5.01$, $p < .01$, as well as a (marginally) significant interaction between the two, $F(77, 1155) = 1.27$, $MSE = 2.75$, $p = .06$. Once again, this interaction indicates that the various experimental contexts induced different patterns of tonal stability for the probe tones. One reason that this interaction might have been only marginally significant is that, unlike the previous two experiments, the patterns of tonal stability across probe positions were much more related to one another in this study; hence, differences as a function of context would be expected to be more subtle.

As discussed, the primary purpose of this study was to examine the algorithm's ability to model listeners' changing tonal percepts throughout this piece, with the secondary goal of assessing the different ways in which the input vector for the algorithm could be generated. Toward these ends, different input vectors were created, based on varying a few underlying factors. The first of these factors was the type of window used to delineate duration information to be included into the input vector; Table 6 presents a summary of these model types. At one extreme, the input vector could be based on all duration information occurring up to a specific point in time (e.g., all note durations up to Probe Position 1); such a model is a *cumulative* model in that it simply accumulates note duration information. At the other extreme, the input vector could be based on note duration occurring only immediately prior to a specific point in time; this formulation is a *current* model based exclusively on local information. Finally, there are models in which input vector durations are based on combinations of current information, remotely past (i.e., memory) information, and upcoming, to-be-heard (i.e., expectancy) information. Based on previous work (i.e., Application III of the Krumhansl-Schmuckler algorithm; Krumhansl, 1990), it was decided that, along with current information, note durations from two adjacent chunks would serve as memory and expectancy components. Thus, memory models looked backward in time by two chunks (and combined this information with current information), expectancy models looked forward two chunks, and combined memory-expectancy models looked backward and forward by one chunk each. As an aside, it is intriguing that expectancy here is equated with what is actually to occur in the future as opposed to what listeners may or may not expect to occur; the implications of this assumption are explored further in the *Discussion* section.

For all models (except for the cumulative model), the amount of information comprising a chunk has been left unspecified. This parameter of the input vector, called the size of the input vector window, represents the second factor that was varied. Five different window sizes were explored (see Table 6), ranging from a window containing note durations from a single beat

The figure displays a musical score for Chopin's E minor prelude (Op. 28, no. 4) in 3/4 time. The score is divided into eight sections, each labeled with a Probe Position (PP1 through PP8) above the treble clef staff. The bass clef staff provides a harmonic accompaniment. PP1 is at the beginning. PP2, PP3, and PP4 are located in the first system. PP5 is at the start of the second system. PP6, PP7, and PP8 are in the third system. The piece concludes with a double bar line and a repeat sign.

Figure 7. Chopin's E minor prelude (Op. 28, no. 4). The probe positions (PPs) for Experiment 3 are noted.

to a window containing durations from 8 beats. It should be noted that it is not possible to fully construct some of the different combinations of model type and window size, depending on the specific probe position in question. So, for example, for Probe Position 1, although one can create memory models with 1-beat information, one cannot fully include all of the requisite memory information for larger window sizes (e.g.,

2-beat up to 8-beat windows) because this information simply does not exist (e.g., there are no note durations 6 beats prior to Probe Position 1). Likewise, no expectancy information can be incorporated for Probe Position 8. Whenever possible, the different models and window sizes were fully constructed; no attempt was made to compensate for missing information when the requisite note information did not exist.

Table 6
Schematic Listing of the Parameters Associated With the Formation of the Different Model Types for Analysis of Chopin's E Minor Prelude

Model type	Weight	Window size (beats)				
Current	—	1	2	4	6	8
Memory	1:1	1	2	4	6	8
	2:1					
	3:1					
Expectancy	1:1	1	2	4	6	8
	2:1					
	3:1					
Memory–expectancy	1:1	1	2	4	6	8
	2:1					
	3:1					
Cumulative	—	—	—	—	—	—

The third and final aspect varied in model formation involved the weighting of the memory and expectancy components relative to the current component. Although complex weighting schemes can be used (e.g., Toivianen & Krumhansl, 2003, discussed subsequently), a simplistic approach involves modifying the different components by integer weights ranging from 1:1 (current vs. memory/expectancy information at equal strength) to 3:1 (current information weighted 3 times more than memory/expectancy information). The different weights used are also summarized in Table 6.

On the basis of the three factors of model type (cumulative, current, memory, expectancy, and memory–expectancy), window size (1 beat, 2 beats, 4 beats, 6 beats, and 8 beats), and weighting (1:1, 2:1, and 3:1), a family of note duration vectors was created for the eight probe positions of this study. These note duration vectors were then used as input vectors to the algorithm, with the corresponding output vectors compared with the output vectors produced by using listeners' probe-tone ratings as the input vectors.

Listeners' tonal percepts and the models' tonal predictions were compared by correlating the rating output vectors with the output vectors generated by the different models. As with the previous studies, these correlations represent the goodness of fit between predictions and percepts of tonality. Theoretically, it is possible to create output vectors based on individual listeners' ratings as well as the average listeners' ratings and to use either in these analyses. Although each method has advantages and disadvantages, the correlations between the averaged listeners' output vectors and the various models were used as the dependent variable.⁶ Initially, the different versions of the memory, expectancy, and memory–expectancy models were compared in a three-way ANOVA, including a between-subjects factor of model type (memory, expectancy, and memory–expectancy) along with within-subjects factors of window size (1 beat, 2 beats, 4 beats, 6 beats, and 8 beats) and weighting (1:1, 2:1, and 3:1). Because the current model does not vary weighting, and the cumulative model does not vary either weighting or window size, neither of these models could be included in this analysis. These models are explored subsequently.

In this three-way ANOVA, neither the main effects for model type, $F(2, 21) = 0.03$, $MSE = 0.84$, *ns*, nor window size, $F(4,$

84) = 1.82, $MSE = 0.01$, *ns*, were significant. In contrast, there was a significant main effect for weighting, $F(2, 42) = 8.04$, $MSE = 0.002$, $p = .001$; the average correlation between output vectors for the 1:1 weighting ($M = .62$, $SD = .23$) was less than the correlations for the 2:1 ($M = .64$, $SD = .23$) or the 3:1 ($M = .64$, $SD = .23$) weightings. Of the two-way effects, the only significant interaction was between window size and weighting, $F(6, 168) = 6.12$, $MSE = 0.001$, $p < .001$. Inspection of this interaction revealed a peaked function for the correlation between output vectors, with these correlations increasing as window size increased from 1 to 2 to 4 beats and then decreasing from 4 to 6 to 8 beats. This pattern, however, held only for the 2:1 and 3:1 weightings; the 1:1 weighting was much less systematic.

All of these results were qualified, however, by the significant three-way interaction, $F(16, 168) = 3.53$, $MSE = 0.001$, $p < .001$. Inspection of this interaction revealed that the peaked pattern just described characterized results for the 2:1 and 3:1 weightings across all three models (memory, expectancy, and memory–expectancy); in contrast, the pattern for the 1:1 weighting was more variable across the models.

Up to now, these analyses have used the correlations between the models' and the listeners' output vectors simply as a convenient dependent measure, with no real regard for the absolute strength of the relation between the two. As such, what these analyses have masked is the fact that, regardless of any variation across model type, window size, or weighting, all of the correlations (averaged across the eight probe positions) were statistically significant. In other words, the actual predictions of listeners' average ratings were consistently good. Subsequent explorations aimed to examine the predictive power of the models more fully, now incorporating the remaining cumulative and current models.

As a first pass, the predictive performance of the memory, expectancy, and memory–expectancy models were compared with the current model in a two-way ANOVA, again using probe position as the random-subject variable, with the factors of model type (memory, expectancy, memory–expectancy, and current) and window size (range of 1 beat to 8 beats); to simplify the comparison, only the 3:1 weighting for the memory, expectancy, and memory–expectancy models was examined. The only significant result arising from this analysis was a main effect of window size,

⁶ The primary advantage to using individual listener ratings is that it provides a natural random-subject variable that enables comparisons within an ANOVA framework and allows for a complete testing of all of the different factors manipulated in this study. Unfortunately, and as might be anticipated based on the intersubject correlations, the actual correlations between individual listeners' output vectors and the various models were quite variable. In contrast, averaging the ratings across listeners provides some much needed stability to these ratings, thus producing better output vector correlations. Unfortunately, this approach eliminates the random-subject factor rendering simple ANOVA comparisons difficult. One way of circumventing this last problem is to treat the eight probe positions as a quasi-random factor, thereby enabling testing of all factors (except for probe position) in an ANOVA framework. Although this latter approach was ultimately adopted in the current situation, it should be noted that all of the analyses presented here were also run using the former procedure (i.e., using individual listener output vector correlations). By and large, the pattern of findings, if not the levels of statistical significance, mimicked that for the averaged ratings.

$F(4, 112) = 7.56$, $MSE = .006$, $p < .001$; the correlations once again showed a peaked function, with the acme at a 4-beat window size. Neither the model type main effect, $F(3, 28) = 0.08$, $MSE = 0.28$, ns , nor the Model Type \times Window Size interaction, $F(12, 112) = 1.42$, $MSE = 0.006$, ns , were significant.

Finally, the predictive power for all five models was examined, using the 3:1 weighting for memory, expectancy, and memory–expectancy models and the 4-beat window size for the memory, expectancy, memory–expectancy, and current models. Figure 8 graphs the correlation between the models and the averaged probe-tone ratings output vectors across the eight probe positions. By and large, the algorithm was accurate in predicting listeners' tonal percepts. There were two notable exceptions to this successful performance, however. First, inspection of Figure 8 reveals that at Probe Position 2 there was significant divergence between listeners' percepts and the algorithm's predictions of tonality, with all the model's predictions nonsignificant. Second, Figure 8 also reveals that predictions from the cumulative model seem to deteriorate for the final three probe positions.

Discussion

In regard to the primary goal of this study, this experiment demonstrates that the Krumhansl–Schmuckler key-finding algorithm can model listeners' percepts of key movement across an extended musical context. This finding thus complements the initial application of the algorithm (Krumhansl, 1990) and replicates and extends the results of Smith and Cuddy (Cuddy & Smith, 2000; Smith & Cuddy, 2003) and Krumhansl and Toivianinen (Krumhansl & Toivianinen, 2000, 2001; Toivianinen & Krumhansl, 2003).

These findings do, however, suggest some important caveats regarding the performance of the key-finding algorithm in that listeners' judgments of tonality were not equally well predicted across all probe positions. Specifically, at Probe Position 2 (Measure 4), listeners generated decidedly different tonal percepts than those predicted by the algorithm. For this probe position, inspection of both listeners' and the algorithm's output vectors reveals that, whereas listeners were perceiving a tonality in the region of

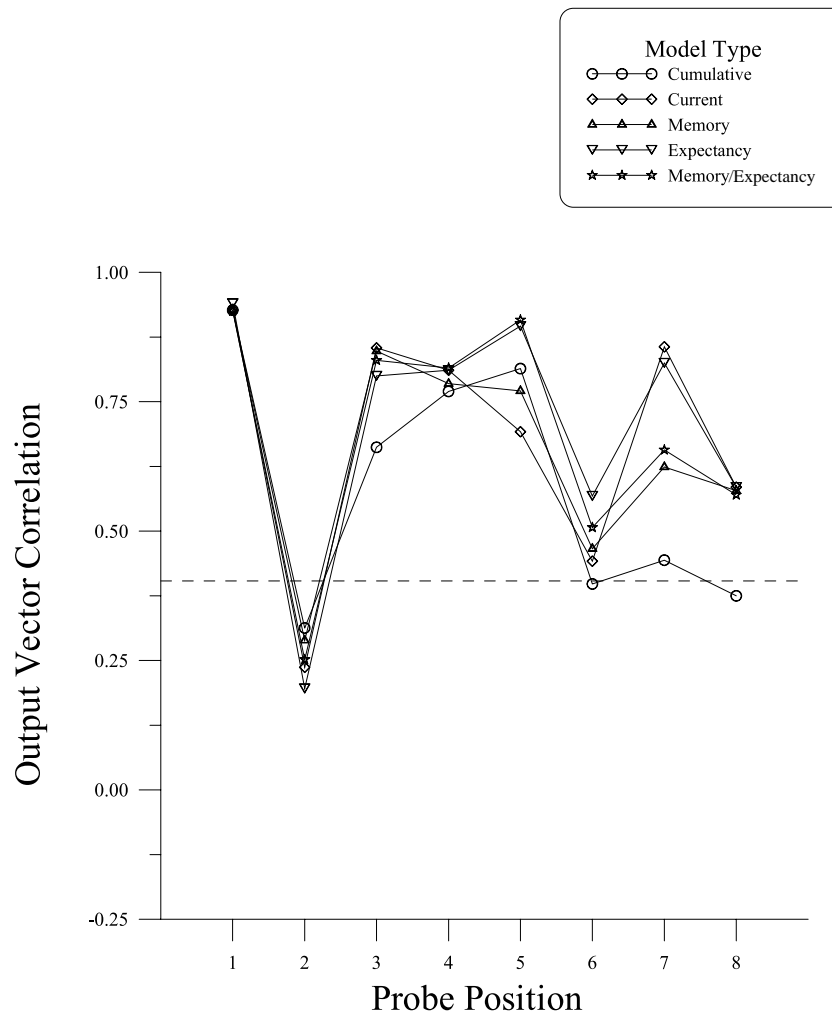


Figure 8. Correlations between the ratings output vectors and the memory (4 beats, 3:1 weighting), expectancy (4 beats, 3:1 weighting), memory/expectancy (4 beats, 3:1 weighting), current (4 beats), and cumulative models' output vectors as a function of probe position. The .05 significance level is indicated.

B major and G# minor, the algorithm predicted different tonal percepts, in the region of E minor to B minor.

This point in the passage presents a somewhat unusual harmonic event, in which the piece moves from the chord built on the 5th scale degree to the chord built on the 2nd scale degree, which (in a minor key) is an extremely rare diminished chord. As an aside, this particular Chopin piece has been the subject of considerable music analytic study and discussion (Kresky, 1994; Lerdahl, 1992; Schacter, 1994), with this phrase in particular highlighted as containing a fairly novel and unusual harmonic progression. Lerdahl (1992), for instance, points to this particular passage as one in which a number of harmonies are passed through with no clear instantiation of their respective tonics; in Lerdahl's words, "the harmonies are unresolved; everything is implication" (p. 185).

One possibility is that this point in the prelude might well constitute a location of musical surprise, one at which listeners' expectations or, correspondingly, the algorithm's predictions for what is to come are disconfirmed in a unique fashion. Schmuckler (1989), for example, observed a similar point of musical surprise and violation of expectations with the occurrence of a unique harmonic progression in a Schumann song. This finding suggests that the key-finding algorithm might be useful in picking out points of musical surprise by, for instance, looking for significant mismatches in localized tonality between consecutive harmonic events. Of course, such a hypothesis requires empirical testing, as does the delineation of what exactly would constitute a significant mismatch in localized tonality, but the ability to potentially distinguish a priori positions of violations of expectations (and hence musical surprise) is especially intriguing.

This study also resolves many of the issues related to the secondary goal of this project: the specification of factors important in forming the input vector. Three different parameters were initially identified as potentially being critical: the type of note duration information entered into the input vector window (i.e., cumulative, current, memory, or expectancy information), the weighting of this information (i.e., equivalent vs. differential weighting), and the size of the input vector window (i.e., 1 beat vs. 2 beats, etc.).

With regard to the first of these factors, this study convincingly demonstrates that there is little variation between the different models used. Thus, it appears to make little difference whether, along with current information, memory or expectancy information is included into the input vector. In many ways, this finding seems counterintuitive, in that one would have anticipated some difference between at least some of the models tested.

With regard to this finding, a few points are worth noting. First, it should be remembered that the predictions of the cumulative model did drop off as the piece progressed, although the "poor performance" of this model at the end was still at roughly the level of statistical significance (an admittedly questionable virtue given the issues regarding the meaning of statistical significance with output vector correlations). Still, the fact that the cumulative model does worse later in the piece is significant in that it suggests that percepts of tonality do not take into account everything that has gone before but instead are restricted to a smaller span of time. In other words, tonality seems to be a localized phenomenon (the size of this localized window is discussed shortly). This interpretation runs counter to the (often implicit) assumption in music-theoretic analyses that there is always an influence of the intended tonality

of a piece regardless of localized tonicizations as well as to the evidence of global context effects in harmonic priming (Bigand et al., 1999; Tillman & Bigand, 2001). What these results do suggest, however, is that at least in listeners' experiences any global tonal influence might be negligible. As an avenue for future research, the question of whether or not listeners actually do maintain a global sense of tonality that persists despite more localized harmonic implications is an intriguing issue but one that might not be easily quantifiable by means of the probe-tone technique, which does not explicitly differentiate local versus global influences.

Second, and with respect to at least the expectancy model, it should be remembered that the information for this model was based on what was to actually occur in the future. Now, within the current context, such an assumption is not especially limiting, in that a majority of the listeners in this study were, in fact, familiar with this piece of music. More generally, though, this does raise the possibility that future model building that wishes to include expectancy might wish to use other means for generating expectancy information. One possibility might be to try including Schellenberg's two-factor reduction (Schellenberg, 1997) of Narmour's implication-realization model (Narmour, 1990, 1992), although such an approach is hampered by the fact that such expectancy information is limited to an extremely local (i.e., the single next note event) level. Regardless, this approach might show a more impressive role for expectancy information in tonal modeling.

In contrast with the lack of differences between model types, both the weighting and window size factors did influence the algorithm's predictive powers. These factors interacted, with the best predictions occurring for a 4-beat window for any form of differential (i.e., 2:1 or 3:1) weighting scheme. The fact that no difference was seen between the current model and the more extended (expectancy, memory, memory-expectancy) models at this window size is telling in that it suggests that little is gained by the addition of information further away than 4 beats from any particular probe position. What is unclear about this result is whether this cutoff represents a constraint in the amount of information that can be processed by listeners at any one time or a constraint in terms of a time window in which information can be stored and integrated by listeners; this issue is more fully explored in the General Discussion. Regardless, these effects do suggest that such factors need be carefully considered when creating the input vector.

General Discussion

The results of this series of experiments have a number of implications for the processing of musical materials in particular and for aspects of cognitive processing more generally. These two sets of implications are explored in turn.

In terms of the implications for musical processing, one finding is that the Krumhansl-Schmuckler key-finding algorithm does seem operable across a range of analytical and psychological contexts. In Krumhansl and Schmuckler's original formulation (described in Krumhansl, 1990), the algorithm was successfully applied to an array of analytic tasks, from determining the composer's designated key for a passage based on very brief segments (Application I) to a note-by-note determination of the designated key in melodic fugue subjects (Application II) to the modeling of music-theoretic expert judgments of tonal influences on a

measure-by-measure basis of an entire Bach prelude (Application III). What the current studies add to this list is the ability to model listeners' percepts of tonality across situations such as brief initial segments of a range of musical excerpts (Experiment 1), on an incremental measure-by-measure basis (Experiment 2) and to track perceived tonality and modulation throughout an extended piece of music (Experiment 3).

Thus, the key-finding algorithm represents a useful tool in both psychological and musicological contexts, providing a direct link between the analytic processes used to highlight musical structure in music-theoretic applications and the viability of these analytic processes as an actual model of the psychological processes operative during the apprehension of music. Such direct links between music-analytic tools on the one hand and models of cognitive processes on the other hand, although much sought after, have been noticeably scant in the music cognition literature. To use a well-known example, one of the more influential theories of musical structure, Lerdahl and Jackendoff's (1983) *Generative Theory of Tonal Music*, explicitly assumes that the music-theoretic rules used to analyze musical passages coincide with the operation of comparable psychological processes during the perception of such passages. In other words, the theoretical principles proposed for guiding the analysis of music are assumed to have psychological reality. In fact, a fair amount of research over the past few decades has been devoted to testing this assumption of Lerdahl and Jackendoff's theory, with admittedly mixed results (Deliège, 1987; Dibben, 1994; Palmer & Krumhansl, 1987a, 1987b). The key-finding algorithm, then, represents a significant addition here, providing a process that is equally potent in both music-theoretic and psychological contexts.

Accordingly, the key-finding algorithm has the potential to be profitably used across a range of psychological and music analytic contexts. On the psychological side, the key-finding algorithm can be used to quantitatively assess the tonal implications of musical stimuli, thus enabling researchers to evaluate the influence of tonal factors in different experimental contexts. Frankland and Cohen (1996) provide an example of such an application in their use of the algorithm to quantify the effects of tonality in a pitch-comparison task. Similarly, Schmuckler and colleagues have routinely used the key-finding algorithm to assess the tonal content of stimuli in studies on short-term memory (STM) of musical phrases (Chiappe & Schmuckler, 1997), the impact of tonality on derived similarity judgments of musical contour (Schmuckler, 1999, 2004), and even tonal influences on listeners' perceptions (Schmuckler, 1989) and pianists' productions (Schmuckler, 1990) of musical expectations.

In terms of music-analytic uses, the key-finding algorithm could be similarly useful. Theories and models of harmonic analysis (e.g., Gjerdingen, 1987; Lerdahl & Jackendoff, 1983; Meehan, 1980; Smoliar, 1980; Temperley, 1997), for instance, often take as a starting point knowledge of the fundamental tonality of a piece of music, with this reference tonality frequently given either externally or explicitly to the model prior to the beginning of harmonic analysis. Smoliar's (1980) computer program for Schenkerian analysis, for instance, assigns a reference tonality at the initial step in the running of the program, and, even more recently, sophisticated algorithms for harmonic analysis (e.g., Temperley, 1997) explicitly divide the task of harmonic analysis into the two subprocesses of key-finding (determining the underlying tonality

of a particular passage) and root-finding (identifying the root triad for the passage).

Thus, the key-finding algorithm represents a potentially powerful tool for many different forms of analysis. Of course, one ironic consequence of the algorithm's success in key-finding is that multiple researchers have since suggested ways by which the algorithm's performance might be enhanced. Some approaches, such as those of Temperley (1999) and Huron and Parncutt (1993), have focused primarily (albeit not exclusively) on improving the operation of the existing algorithm by modifying key parameters of the algorithm (e.g., the input vector, the means of comparing input vectors and idealized key vectors). Huron and Parncutt (1993), for instance, suggest that the input vector can be sharpened by applying two additional factors to its calculation: an explicit, continuous exponential decay function for previous information (based on the rate of decay in echoic memory), and a psychoacoustic factor based on Terhardt et al.'s (Terhardt, 1974; Terhardt, Stoll, & Seewann, 1982a, 1982b) theory of pitch perception (see Parncutt, 1988, 1989).

Probably the most extensive set of modifications has been proposed by Temperley (1999, 2001, 2002, 2004). This work has highlighted a number of potential flaws in the key-finding algorithm, including problems in the actual values used as the ideal key profile, the use of a correlational technique for comparing the input vector with the idealized values, and the hypersensitivity of the algorithm to variations in duration and frequency of occurrence information. To address these concerns, Temperley has proposed substitute values for the key profiles and an alternative means of calculating fit that makes use of a flat profile consisting of a series of 1s and 0s, which represent presence versus absence, respectively, of the chromatic pitches in the context under examination. This flat profile is then multiplied by the revised key profiles, and the resulting values are summed to produce a single key score. The summed scores across all 12 major and 12 minor keys can then be compared, with the maximum summed value indicative of listeners' perception of key.

Although intriguing, one can question whether changes such as these significantly enhance key-finding assessments, particularly with reference to listeners' actual percepts of key. Huron and Parncutt (1993), for instance, compare their revised model with the findings from Krumhansl and Kessler's (1982) investigations of changes in perceived tonality over the course of nonmodulating and modulating sequences. For the nonmodulating sequence, the revised model does not appear to produce better performance than what would be produced by the original key-finding algorithm,⁷ and for the modulating sequence the performance of the key-finding algorithm is not even compared with Huron and Parncutt's (1993) revised version. Temperley (1999) provides a detailed comparison of his modified algorithm with other models of key-

⁷ Huron and Parncutt (1993) do not test their model against predictions based on the actual Krumhansl-Schmuckler key-finding algorithm but rather on the correlations between listeners' ratings and the idealized key profile for the nonmodulating sequences (i.e., the C major key profile), as originally presented in Krumhansl and Kessler (1982). Given that these correlations would simply fall out of applying the key-finding algorithm to this passage, however, it is reasonable to take this comparison as indicative of one instance of how the algorithm might perform with this sequence.

finding (e.g., Holtzman, 1977; Longuet-Higgins & Steedman, 1971; Vos & Van Geenan, 1996) and demonstrates that this version of the algorithm performs favorably relative to these other models. Again, though, this comparison did not include the original Krumhansl–Schmuckler algorithm and was done with reference to the author’s (admittedly expert) musical intuitions of tonality and not on the basis of listeners’ percepts of tonality. Thus, such comparisons still leave open the question of whether or not these modifications actually add to what the original algorithm was able to accomplish.

In an attempt to at least partially address this issue, two versions of Temperley’s (1999, 2001, 2002, 2004) modified algorithm were applied to the contexts and data of the first two experiments in this series of studies. The first of these models involved substituting Temperley’s revised key profile values for the Krumhansl and Kessler (1982) values and running the Krumhansl–Schmuckler algorithm as it normally would operate (i.e., correlating duration based input vectors with these revised profiles). The second model involved applying Temperley’s flat input vector approach, with each pitch class given a value of 1 if present or 0 if absent, multiplying this flat vector with the modified key profile values, and then summing the results to yield a score for each key.

Although the results of this analysis are not presented formally,⁸ the findings are nonetheless informative. Comparison of the two models based on Temperley’s approach with the original Krumhansl–Schmuckler algorithm failed to reveal increased predictive power in terms of predicting either the intended key of the context passages or listeners’ tonal percepts for both Experiments 1 and 2. For example, for the Chopin contexts of Experiment 1, the Krumhansl–Schmuckler algorithm was able to predict the intended key for 13 of 24 contexts. In comparison, using Temperley’s alternative profile values resulted in 12 of 24 significant predictions, and using the flat input vector resulted in the intended key having the maximum key score in 15 of 24 contexts compared with a maximum key correlation for 11 of 24 contexts for the key-finding algorithm. Relative to listeners’ tonal percepts, output vector correlations were significant for 16 of 24 preludes for the Krumhansl–Schmuckler algorithm, 19 of 24 preludes for the revised profile values, and 16 of 24 preludes for the flat input vector model. Finally, the Krumhansl–Schmuckler algorithm produced a higher output vector correlation than the other two models for 12 of the preludes compared with 3 and 6 for the modified values and flat input vector models, respectively.

Given such findings, it is tempting to conclude that the modifications proposed by Temperley are unnecessary to the operation of the Krumhansl–Schmuckler key-finding algorithm. Such a conclusion however, seems unnecessarily hasty. First, it must be remembered that Temperley (1999, 2001) has applied his modified algorithm to a substantial corpus of materials, with the findings showing an advantage for his approach. Although these results must be interpreted cautiously, because they are not based on perceptual data, they nevertheless do represent positive evidence for his approach. Second, comparing the performance of the various models on the current contexts is likely not the most illuminating test of the distinctions between these models. All these comparisons actually demonstrate is that the different models make comparable predictions for these stimuli. A much more informative test would be to choose contexts in which the models make contrasting predictions and determine which approach best

predicts listeners’ percepts in response to such passages. A similar strategy was applied by Schmuckler and Tomovski (2000) in comparing the Krumhansl–Schmuckler algorithm and the intervallic rivalry model of Brown and Butler (Brown, 1988; Brown & Butler, 1981; Butler & Brown, 1994).

Along with modifying the original algorithm, other authors have proposed additional processes that could be run in conjunction with the key-finding algorithm. Krumhansl and Toivainen (Krumhansl & Toivainen, 2000, 2001; Toivainen & Krumhansl, 2003),⁹ using their continuous perceptual judgment task, compared the performance of the original key-finding algorithm, modified by the inclusion of a pitch memory vector, with a related model using two-note transition information. Unexpectedly, these authors found that both the original and tone-transition models were equally effective at modeling listeners’ tonal percepts. As such, it is simply not clear exactly what there is to be gained by adding this extra component.

Smith and Schmuckler (2004) suggested a different addition to the key-finding algorithm, proposing a separate mechanism for assessing key strength in general, regardless of how tonality itself is determined. Specifically, Smith and Schmuckler (2004) investigated the tonal magnitude of a sequence. *Tonal magnitude* refers to the degree to which the pattern of differences between the values of the component pitches of the input vector is either accentuated or deaccentuated relative to what would be expected based on the Krumhansl and Kessler (1982) profiles. Accentuated input (or key) vectors are ones in which the psychologically stable tones are heard for even longer (and, correspondingly, the unstable tones are heard for shorter) than would be expected based on a literal translation of the idealized profile into proportional note durations, and deaccentuated input vectors are ones in which the relative differences between tones, although still retaining the patterning of the idealized profiles, are more or less flattened. Although both accentuated and deaccentuated input vectors strongly correlate with the idealized key profiles, Smith and Schmuckler (2004) observed that listeners perceived tonality in random note sequences only at higher tonal magnitude values (e.g., tonal magnitude values of about 1.5) than what would be predicted based on the original Krumhansl and Kessler values (a tonal magnitude value of 1.0).

These results suggest that the tonal magnitude of a sequence might provide a measure of tonal strength, with increasing deviation away from tonal magnitude values of 1.5 (roughly) indicating increasingly weaker tonal strength. Although speculative, this idea is promising in that tonal magnitude represents an independent way of predicting tonal strength and thus provides a method for determining passages that, although clear in their potential key indications, might nevertheless instantiate a key only weakly.

Finally, along with enhancing tonal assessment, other models might be run in conjunction with the algorithm to provide a more complete characterization of the passage in question, one that

⁸ Details regarding this analysis are available upon request from Mark A. Schmuckler.

⁹ Although Krumhansl and Toivainen present their work in the form of a self-organizing map neural network, the actual details of its method of key determination are not far off from that used in the original Krumhansl–Schmuckler key-finding algorithm.

incorporates the perception of tonality along with the perception of other musically important dimensions. Smith and Cuddy (2003) provide an excellent example of such a global approach by simultaneously examining four dimensions of listeners' percepts—phrase structure, tonality, tension, and consonance and dissonance—throughout a short excerpt from Beethoven's *Waldstein* sonata. Similarly, Toiviainen and Krumhansl (2003) studied the perception of tension, in conjunction with tonality, in their analyses, and Krumhansl (1996) assessed listeners' perceptions of segmentation, tension, and musical ideas throughout the first movement of Mozart's Eb major piano sonata, K. 282. All of these studies demonstrate how assessing multiple musical dimensions simultaneously can provide a more complete description of the listening experience.

Turning now to the implications of this work for general cognitive processing, one intriguing finding is that this research provides evidence for one of the more maligned models of perceptual processing: the idea of pattern matching to a template. Typically, the fundamental problem with template matching is that perceivers can recognize patterns despite the fact that such patterns can and do occur in a virtually unlimited number of sizes and orientations (see Neisser, 1967, for a classic discussion of this problem). Although others, such as Reed (1973), have tried to salvage this approach, by and large strict pattern matching has been considered unworkable as a serious theory of perceptual recognition for decades. It is within this context, then, that the current evidence for the viability of a template-matching process is most noteworthy. In the current situation, the template matching that is occurring is, in fact, more along the lines proposed by Reed (1973), in which the to-be-matched template is actually an abstract prototype (e.g., the tonal hierarchy pattern) rather than a copy of the to-be-recognized pattern (e.g., a typical tonal pattern).

Of course, a successful pattern-matching process assumes that listeners are, on some level, sensitive to the statistical variations in note durations of the to-be-heard sequences. Although it has generally been thought that sensitivity to such subtle properties of auditory or visual sequences is difficult for perceivers, research suggests that sensitivity to the statistical properties of sequences may not be as problematic as previously supposed. Some of the most dramatic evidence in this regard has come from research on word segmentation (Aslin, Saffran, & Newport, 1998; Saffran, 2001; Saffran, Aslin, & Newport, 1996; Saffran, Newport, & Aslin, 1996; Saffran, Newport, Aslin, Tunick, & Barrueco, 1997), in which both adults and infants make use of statistical probabilities between neighboring speech sounds to extract "words" from continuous streams of nonsense syllables. Moreover, both infants and adults are similarly sensitive to the conditional probabilities between tones for both atonal and tonal sequences of notes (Saffran, 2003; Saffran & Griepentrog, 2001; Saffran, Johnson, Aslin, & Newport, 1999) and for sequences of musical timbres (e.g., Tillman & McAdams, 2004). More directly related to key-finding, both Smith and Schmuckler (2004) and Lantz (2002) have demonstrated that listeners are sensitive to the actual patterns of relative duration differences between tones that underlie tonal sequences. All of this research, then, indicates that, under suitable conditions, listeners are sensitive to statistical regularities in auditory sequences. Accordingly, abstraction of a pattern of durational differences that would adhere to the tonal hierarchy is clearly

possible by listeners and could thus easily form the basis for listeners' apprehension of tonality (see also Tillman et al., 2000).

A second intriguing finding arising from this work involves the result that about 4 beats worth of musical information appeared to produce the optimum predictions of tonality throughout this musical passage. As mentioned, it is not entirely clear whether the constraint here is information based, resulting from there being some optimum amount of musical information (e.g., some number of note events) that can be integrated in listeners' memories or whether this limit is more temporally based (e.g., a given time window regardless of the amount of musical information occurring during this interval). In this regard, it is noteworthy that Krumhansl and Toiviainen (2000, 2001; Toiviainen & Krumhansl, 2003) modified the original input vector by including an explicit 3-s pitch memory vector based on estimates of echoic memory (Darwin, Turvey, & Crowder, 1972; Treisman, 1964). What is intriguing is that Krumhansl and Toiviainen's pitch memory vector of 3 s matches reasonably closely with the optimum time window of this work, which was 4 s. Accordingly, this unexpected convergence between these two sources suggests that the constraining factor may indeed be time and not information.

Assuming, of course, that it is time that is key here, the question is raised of why the optimum window might be 3–4 s. One obvious answer has already been alluded to: This amount of time is roughly the length of echoic memory. Accordingly, musical information occurring earlier than this window simply fades from memory and thus cannot be effectively integrated into any form of ongoing, albeit unconscious, tabulation of event frequency.

Other authors have, in fact, given a central role to such STM processes in tonal induction generally and as a possibly important factor in the operation of the key-finding algorithm in particular. As described earlier, Huron and Parncutt's (1993) modifications to the key-finding algorithm advocated the adoption of an exponential decay function, based on the rate of decay in echoic memory, as a means of sharpening the input vector. Even more dramatically, Leman's work on tone center perception through the use of self-organizing networks (Leman, 1995a, 1995b, 2000) shows that a memory model, working from echoic memory of pitch periodicity, may be sufficient to explain listeners' ratings in probe-tone-type situations. If true, such a notion questions the entire assumption of the role of long-term internalized tonal hierarchy information in musical listening. Although exploration of this latter issue is beyond the resources of the current work, all of these findings highlight the potential importance of STM processes in tonal perception.

There are, however, reasons why one should look for factors other than memory storage processes in explaining musical key-finding. There is ample evidence that listeners can and do make use of musical information that has occurred earlier than 3–4 s. Perception of melodic contour, for instance, often requires the integration of more than 3–4 s worth of musical information. Moreover, the apprehension of large-scale musical structure, such as the recognition of theme and variations (Pollard-Gott, 1983; Werker, 1982), clearly makes use of more extended musical information. Accordingly, to suggest that it is solely memory that limits the use of more extended musical information in tonal perception seems unrealistic.

What else might be constraining tonal percepts to a 3–4-s time window? One possibility is that tonality is inherently a local

phenomenon, with listeners having an inherent bias to use only relatively recent information in such judgments. Of course, this still does not truly resolve the issue, in that the question now shifts to trying to understand why listeners have such a bias. One possibility is that fairly low-level psychoacoustic factors, such as the consonance and dissonance of tones, come into play. Although a thorough review of consonance and dissonance cannot be undertaken here (see Rasch & Plomp, 1999, for such a review), there are any number of reasons to think that such a factor might play a role in this context. First, the discrimination of consonant (the subjective experience of two or more simultaneous frequencies sounding pleasant) versus dissonant (the subjective experience of two or more simultaneous frequencies sounding unpleasant) sounds is one that can be seen even in infancy (e.g., Schellenberg & Trainor, 1996; Schellenberg & Trehub, 1996; Trainor & Heinmiller, 1998; Zentner & Kagan, 1996, 1998), thus rendering it a very fundamental property of musical sequences. Second, the consonance versus dissonance of tone pairs actually corresponds quite strongly with the Krumhansl and Kessler (1982) tonal hierarchy ratings (Krumhansl, 1990). In this case, listeners' judgments of psychological stability are (at least in part) determined by their percepts of the consonance of the probe tone with respect to the tonic. Because such consonance judgments require some working representation of the to-be-judged tones (either actual soundings of tones or an active memory trace), tonal judgments would thus be limited to information from just a prior few seconds.

Along with psychoacoustic factors, another influence might reside in a different psychological concept, specifically the idea of the psychological present (e.g., Block, 1979; Fraisse, 1984; Pöppel, 1972, 1978; Rammsayer, 2001). Generally, the psychological present represents the time period that is considered to be a single unit by observers, with the processing of durations within the psychological present possibly making use of different mechanisms than processing outside of the psychological present (Block, 1990; Rammsayer, 2001). According to Block, Zakay, and Hancock (1999), the typical use of temporal information in everyday life occurs within the psychological present, which is thought to be on the order of 3–5 s. This, then, might be the underlying mechanism of how tonal determinations are made, with listeners desiring information to reside within a single temporal unit for their judgments of psychological stability. Although speculative, the idea of temporal constraints on tonal determination, regardless of whether they arise from memory factors, psychoacoustic bases, or everyday experiences of temporal units, is an intriguing possibility and one that should be explored in future work.

Finally, the current results have an intriguing implication in that they provide evidence for the use of simultaneous cognitive reference points during musical processing. Although research in cognitive psychology has repeatedly demonstrated the importance of cognitive reference points in, for the most part, all behavior, little attention has been paid to how observers might make use of multiple reference points in their activities. Nevertheless, it seems clear that the processing of visual or auditory objects involves determining their status not only relative to the most obvious categorization hierarchy (e.g., how well a given chair fits with the concept of chair or furniture in general) but also with reference to other, operative reference frames that might be important in a given situation (e.g., how well that chair fits with the category of furniture to fit into a specific architectural niche in one's home or

fit into the color scheme of one's other furniture, etc.). One of the more compelling findings from this work is that perceivers can easily make use of multiple cognitive points when processing auditory information and can move back and forth between different reference points (in this case, different tonalities) rather effortlessly. It would be of interest to know whether the use of multiple reference schemes simultaneously is a very general ability, applicable to the processing of any number of visual and auditory objects, or whether multiple reference schemes can only be used within fairly constrained contexts in which the different reference schemes are themselves highly related (such as occurs with musical tonality).

In sum, the current studies of the organization of pitch within a tonal context have provided an interesting window into a number of topics on general perceptual and cognitive functioning. Presumably, the factors highlighted in this investigation of key-finding represent specific instantiations of much more general cognitive processes, ones in which the study of music has afforded some more privileged insight. The goal, of course, is to not only understand the psychological processes involved in the apprehension of music in particular but to gain a better insight into perceptual and cognitive functioning more generally.

References

- Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science, 9*, 321–324.
- Bharucha, J. J. (1994). Tonality and expectation. In R. Aiello & J. Sloboda (Eds.), *Musical perceptions* (pp. 213–239). Oxford, England: Oxford University Press.
- Bharucha, J. J., & Krumhansl, C. L. (1983). The representation of harmonic structure in music: Hierarchies of stability as a function of context. *Cognition, 13*, 63–102.
- Bharucha, J. J., & Stoeckig, K. (1986). Reaction time and musical expectancy. *Journal of Experimental Psychology: Human Perception and Performance, 12*, 403–410.
- Bharucha, J. J., & Stoeckig, K. (1987). Priming of chords: Spreading activation or overlapping frequency spectra? *Perception & Psychophysics, 41*, 519–524.
- Bigand, E., Madurell, F., Tillman, B., & Pineau, M. (1999). Effect of global structure and temporal organization on chord processing. *Journal of Experimental Psychology: Human Perception and Performance, 25*, 184–197.
- Bigand, E., & Pineau, M. (1997). Global context effects on musical expectancy. *Perception & Psychophysics, 59*, 1098–1107.
- Bigand, E., Poulin, B., Tillman, B., Madurell, F., & D'Adamo, D. A. (2003). Sensory versus cognitive components in harmonic priming. *Journal of Experimental Psychology: Human Perception and Performance, 29*, 159–171.
- Block, R. A. (1979). Time and consciousness. In G. Underwood & R. Stevens (Eds.), *Aspects of consciousness: Vol. 1. Psychological issues* (pp. 179–217). London: Academic Press.
- Block, R. A. (1990). Models of psychological time. In R. A. Block (Ed.), *Cognitive models of psychological time* (pp. 1–35). Hillsdale, NJ: Erlbaum.
- Block, R. A., Zakay, D., & Hancock, P. A. (1999). Developmental changes in human duration judgments: A meta-analytic review. *Developmental Review, 19*, 183–211.
- Brown, H. (1988). The interplay of set content and temporal context in a functional theory of tonality perception. *Music Perception, 5*, 219–249.
- Brown, H., & Butler, D. (1981). Diatonic trichords as minimal tonal cue-cells. *In Theory Only, 5*, 37–55.

- Brown, H., Butler, D., & Jones, M. R. (1994). Musical and temporal influences on key discovery. *Music Perception, 11*, 371–407.
- Browne, R. (1981). Tonal implications of the diatonic set. *In Theory Only, 5*, 3–21.
- 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*, 219–241.
- Butler, D. (1990). A study of event hierarchies in tonal and post-tonal music. *Music Perception, 18*, 4–17.
- Butler, D., & Brown, H. (1994). Describing the mental representation of tonality in music. In R. Aiello & J. A. Sloboda (Eds.), *Musical perceptions* (pp. 191–212). London: Oxford University Press.
- Chiappe, P., & Schmuckler, M. A. (1997). Phrasing influences the recognition of melodies. *Psychonomic Bulletin & Review, 4*, 254–259.
- Cohen, A. J. (1991). Tonality and perception: Musical scales primed by excerpts from *The Well-Tempered Clavier* of J. S. Bach. *Psychological Research, 53*, 305–314.
- 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., & Lunney, C. A. (1995). Expectancies generated by melodic intervals: Perceptual judgments of melodic continuity. *Perception & Psychophysics, 57*, 451–462.
- Cuddy, L. L., & Smith, N. A. (2000). Perception of tonal pitch space and tonal tension. In D. Greer (Ed.), *Musiology and sister disciplines* (pp. 47–59). Oxford, England: Oxford University Press.
- Darwin, C. J., Turvey, M. T., & Crowder, R. G. (1972). An auditory analogue of the Sperling partial report procedure: Evidence for brief auditory storage. *Cognitive Psychology, 3*, 255–267.
- Deliège, I. (1987). Grouping conditions in listening to music: An approach to Lerdahl and Jackendoff's grouping preference rules. *Music Perception, 4*, 325–359.
- Dibben, N. (1994). The cognitive reality of hierarchic structure in tonal and atonal music. *Music Perception, 12*, 1–26.
- Fraisse, P. (1984). Perception and estimation of time. *Annual Review of Psychology, 35*, 1–36.
- Frankland, B. W., & Cohen, A. J. (1996). Using the Krumhansl and Schmuckler key-finding algorithm to quantify the effects of tonality in the interpolated-tone pitch-comparison task. *Music Perception, 14*, 57–83.
- Gjerdingen, R. O. (1987). *A classic turn of phrase*. Philadelphia: University of Pennsylvania Press.
- Holtzman, S. R. (1977). A program for key determination. *Interface, 6*, 29–56.
- 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., Tillman, B., & Bharucha, J. J. (2003). Online detection of tonal pop-out in modulating contexts. *Music Perception, 20*, 283–305.
- Janata, P., & Reisberg, D. (1988). Response-time measures as a means of exploring tonal hierarchies. *Music Perception, 6*, 161–172.
- Jones, M. R. (1981). Music as a stimulus for psychological motion: Part I. Some determinants of expectancies. *Psychomusicology, 1*, 34–51.
- Jones, M. R. (1982). Music as a stimulus for psychological motion: Part II. An expectancy model. *Psychomusicology, 2*, 1–13.
- Jones, M. R. (1990). Learning and the development of expectancies: An interactionist approach. *Psychomusicology, 9*, 193–228.
- Justus, T. C., & Bharucha, J. J. (2002). Music perception and cognition. In S. Yantis (Ed.), *Stevens' handbook of experimental psychology: Vol. 1: Sensation and perception* (3rd ed., pp. 453–492). New York: Wiley.
- Kresky, J. (1994). *A reader's guide to the Chopin preludes*. Westport, CT: Greenwood Press.
- Krumhansl, C. L. (1979). The psychological representation of musical pitch in a tonal context. *Cognitive Psychology, 11*, 346–374.
- Krumhansl, C. L. (1990). *Cognitive foundations of musical pitch*. London: Oxford University Press.
- Krumhansl, C. L. (1995). Music psychology and music theory: Problems and prospects. *Music Theory Spectrum, 17*, 53–80.
- 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. (2000a). Rhythm and pitch in music cognition. *Psychological Bulletin, 126*, 159–179.
- Krumhansl, C. L. (2000b). Tonality induction: A statistical approach applied cross-culturally. *Music Perception, 17*, 461–480.
- Krumhansl, C. L., Bharucha, J. J., & Castellano, M. A. (1982). Key distance effects on perceived harmonic structure in music. *Perception & Psychophysics, 32*, 96–108.
- Krumhansl, C. L., Bharucha, J. J., & Kessler, E. J. (1982). Perceived harmonic structure of chords in three related musical keys. *Journal of Experimental Psychology: Human Perception and Performance, 8*, 24–36.
- 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). *Key-finding in music: An algorithm based on pattern matching to tonal hierarchies*. Paper presented at the 19th Annual Meeting of the Society of Mathematical Psychology, Cambridge, MA.
- 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. In C. Woods, B. B. Luck, R. Rochard, S. A. O'Neil & J. A. Sloboda (Eds.), *Proceedings of the Sixth International Conference on Music Perception and Cognition* (pp. 1504–1513). Keele, England: Keele University.
- Krumhansl, C. L., & Toiviainen, P. (2001). Tonal cognition. In R. J. Zatorre & I. Peretz (Eds.), *The biological foundations of music: Annals of the New York Academy of Sciences* (Vol. 930, pp. 77–91). New York: New York Academy of Sciences.
- Lantz, M. E. (2002). *The role of duration and frequency of occurrence in perceived pitch structure*. Kingston, Ontario, Canada: Queen's University.
- Leman, M. (1995a). A model of retroactive tone-center perception. *Music Perception, 12*, 430–471.
- Leman, M. (1995b). *Music and schema theory: Cognitive foundations of systematic musicology*. Berlin: Springer-Verlag.
- Leman, M. (2000). An auditory model of the role of short-term memory in probe-tone ratings. *Music Perception, 17*, 481–509.
- Lerdahl, F. (1992). Pitch-space journeys in two Chopin preludes. In M. R. Jones & S. Holleran (Eds.), *Cognitive bases of musical communication* (pp. 171–191). Washington, DC: American Psychological Association.
- Lerdahl, F., & Jackendoff, R. (1983). *A generative theory of tonal music*. Cambridge, MA: MIT Press.
- Longuet-Higgins, H. C., & Steedman, M. J. (1971). On interpreting Bach. *Machine Intelligence, 6*, 221–241.
- Meehan, J. R. (1980). An artificial intelligence approach to tonal music. *Computer Music Journal, 4*, 60–65.
- Meyer, L. B. (1956). *Emotion and meaning in music*. Chicago: University of Chicago Press.
- Murdock, B. B. J. (1962). The serial position effect on free recall. *Journal of Experimental Psychology, 64*, 482–488.
- Narmour, E. (1989). The “genetic code” of melody: Cognitive structures generated by the implication–realization model. *Contemporary Music Review, 4*, 45–63.

- Narmour, E. (1990). *The analysis and cognition of basic melodic structures*. Chicago: University of Chicago Press.
- Narmour, E. (1992). *The analysis and cognition of melodic complexity*. Chicago: University of Chicago Press.
- Neisser, U. (1967). *Cognitive psychology*. New York: Appleton-Century-Crofts.
- Palmer, C., & Krumhansl, C. L. (1987a). Independent temporal and pitch structures in determination of musical phrases. *Journal of Experimental Psychology: Human Perception and Performance*, *13*, 116–126.
- Palmer, C., & Krumhansl, C. L. (1987b). Pitch and temporal contributions to musical phrases: Effects of harmony, performance timing, and familiarity. *Perception & Psychophysics*, *51*, 505–518.
- Parncutt, R. (1988). Revision of Terhardt's psychoacoustical model of the root(s) of a musical chord. *Music Perception*, *6*, 65–94.
- Parncutt, R. (1989). *Harmony: A psychoacoustical approach*. Berlin: Springer-Verlag.
- Pollard-Gott, L. (1983). Emergence of thematic concepts in repeated listening to music. *Cognitive Psychology*, *15*, 66–94.
- Pöppel, E. (1972). Oscillations as possible basis for time perception. In J. T. Fraser, F. C. Haber, & G. H. Müller (Eds.), *The study of time* (pp. 219–241). Berlin: Springer-Verlag.
- Pöppel, E. (1978). Time perception. In R. Held, H. W. Leibowitz, & H.-L. Teuber (Eds.), *Handbook of sensory physiology* (pp. 713–729). Berlin: Springer-Verlag.
- Rammesayer, T. H. (2001). Ageing and temporal processing of durations within the psychological present. *European Journal of Cognitive Psychology*, *13*, 549–565.
- Rasch, R., & Plomp, R. (1999). The perception of musical tones. In D. Deutsch (Ed.), *The psychology of music* (2nd ed., pp. 89–112). San Diego, CA: Academic Press.
- Reed, S. K. (1973). *Psychological processes in pattern recognition*. New York: Academic Press.
- Rosch, E. (1975). Cognitive reference points. *Cognitive Psychology*, *7*, 532–547.
- Saffran, J. R. (2001). Words in a sea of sounds: The output of infant statistical learning. *Cognition*, *81*, 149–169.
- Saffran, J. R. (2003). Absolute pitch in infancy and adulthood: The role of tonal structure. *Developmental Science*, *6*, 35–45.
- Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996, December 13). Statistical learning by 8-month-old infants. *Science*, *274*, 1926–1928.
- Saffran, J. R., & Griepentrog, G. J. (2001). Absolute pitch in infant auditory learning: Evidence for developmental reorganization. *Developmental Psychology*, *37*, 74–85.
- Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, *70*, 27–52.
- Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language*, *35*, 606–621.
- Saffran, J. R., Newport, E. L., Aslin, R. N., Tunick, R. A., & Barrueco, S. (1997). Incidental language learning. *Psychological Science*, *8*, 101–105.
- Schacter, C. (1994). The Prelude in E Minor op. 28 no. 4: Autograph sources and interpretation. In J. Rink & J. Samson (Eds.), *Chopin studies* (Vol. 2, pp. 161–182). New York: Cambridge University Press.
- Schellenberg, E. G. (1997). Simplifying the implication-realization model of musical expectancy. *Music Perception*, *14*, 295–318.
- Schellenberg, E. G., & Trainor, L. J. (1996). Sensory consonance and the perceptual similarity of complex-tone harmonic intervals: Tests of adult and infant listeners. *Journal of the Acoustical Society of America*, *100*, 3321–3328.
- Schellenberg, E. G., & Trehub, S. E. (1996). Natural musical intervals: Evidence from infant listeners. *Psychological Science*, *7*, 272–277.
- Schmuckler, M. A. (1989). Expectation in music: Investigation of melodic and harmonic processes. *Music Perception*, *7*, 109–150.
- Schmuckler, M. A. (1990). The performance of global expectations. *Psychomusicology*, *9*, 122–147.
- Schmuckler, M. A. (1997a). Expectancy effects in memory for melodies. *Canadian Journal of Experimental Psychology*, *51*, 292–305.
- Schmuckler, M. A. (1997b). Music cognition and performance: An introduction. *Canadian Journal of Experimental Psychology*, *51*, 265–267.
- Schmuckler, M. A. (1999). Testing models of melodic contour similarity. *Music Perception*, *16*, 295–326.
- Schmuckler, M. A. (2004). Pitch and pitch structures. In J. Neuhoff (Ed.), *Ecological psychoacoustics* (pp. 271–315). San Diego, CA: Academic Press.
- Schmuckler, M. A., & Boltz, M. G. (1994). Harmonic and rhythmic influences on musical expectancy. *Perception & Psychophysics*, *56*, 313–325.
- Schmuckler, M. A., & Tomovski, R. (2000, November). *Tonal hierarchies and intervallic rivalries in musical key-finding*. Paper presented at the meeting of the Society for Music Perception and Cognition, Toronto, Ontario, Canada.
- Shepard, R. N. (1964). Circularity in judgments of relative pitch. *Journal of the Acoustical Society of America*, *36*, 2346–2353.
- Shmulevich, I., & Yli-Harja, O. (2000). Localized key-finding: Algorithms and applications. *Music Perception*, *17*, 531–544.
- Smith, N. A., & Cuddy, L. L. (2003). Perceptions of musical dimensions in Beethoven's *Waldstein* sonata: An application of tonal pitch space theory. *Musicae Scientiae*, *7*, 7–24.
- Smith, N. A., & Schmuckler, M. A. (2004). The perception of tonal structure by the differentiation and organization of pitches. *Journal of Experimental Psychology: Human Perception and Performance*, *30*, 268–286.
- Smoliar, S. W. (1980). A computer aid for Schenkerian analysis. *Computer Music Journal*, *4*, 41–59.
- Takeuchi, A. (1994). Maximum key-profile correlation (MKC) as a measure of tonal structure in music. *Perception & Psychophysics*, *56*, 335–346.
- Temperley, D. (1997). An algorithm for harmonic analysis. *Music Perception*, *15*, 31–68.
- Temperley, D. (1999). What's key for key? The Krumhansl-Schmuckler key-finding algorithm reconsidered. *Music Perception*, *17*, 65–100.
- Temperley, D. (2001). *The cognition of basic musical structures*. Cambridge, MA: MIT Press.
- Temperley, D. (2002). A Bayesian approach to key-finding. In C. Anagnostopoulou, M. Merrand, & A. Smaill (Eds.), *Music and artificial intelligence* (pp. 195–206). Berlin: Springer-Verlag.
- Temperley, D. (2004). Bayesian models of musical structure and cognition. *Musicae Scientiae*, *8*, 175–205.
- Terhardt, E. (1974). Pitch, consonance, and harmony. *Journal of the Acoustical Society of America*, *55*, 1061–1069.
- Terhardt, E., Stoll, G., & Seewann, M. (1982a). Algorithm for extraction of pitch and pitch salience from complex tonal signals. *Journal of the Acoustical Society of America*, *71*, 679–688.
- Terhardt, E., Stoll, G., & Seewann, M. (1982b). Pitch of complex signals according to virtual-pitch theory: Test, examples, and predictions. *Journal of the Acoustical Society of America*, *71*, 671–678.
- Tillman, B., & Bharucha, J. J. (2002). Effect of harmonic relatedness on the detection of temporal asynchronies. *Perception & Psychophysics*, *64*, 640–649.
- Tillman, B., Bharucha, J. J., & Bigand, E. (2000). Implicit learning of tonality: A self-organizing approach. *Psychological Review*, *107*, 885–913.
- Tillman, B., & Bigand, E. (2001). Global context effect in normal and scrambled musical sequences. *Journal of Experimental Psychology: Human Perception and Performance*, *27*, 1185–1196.

- Tillman, B., Janata, P., Birk, J. L., & Bharucha, J. J. (2003). The costs and benefits of tonal centers for chord processing. *Journal of Experimental Psychology: Human Perception and Performance*, 29, 470–482.
- Tillman, B., & McAdams, S. (2004). Implicit learning of musical timbre sequences: Statistical regularities confronted with acoustical (dis)similarities. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30, 1131–1142.
- Toiviainen, P., & Krumhansl, C. L. (2003). Measuring and modeling real-time responses to music: The dynamics of tonality induction. *Perception*, 32, 741–766.
- Trainor, L. J., & Heinmiller, B. (1998). The development of evaluative responses to music: Infants prefer to listen to consonance over dissonance. *Infant Behavior and Development*, 21, 77–88.
- Treisman, A. M. (1964). Verbal cues, language, and meaning in selective attention. *American Journal of Psychology*, 77, 206–219.
- Vos, P. G., & Van Geenan, E. W. (1996). A parallel-processing key-finding model. *Music Perception*, 14, 185–223.
- Werker, R. L. (1982). Abstraction of themes from melodic variations. *Journal of Experimental Psychology: Human Perception and Performance*, 8, 435–447.
- Winograd, T. (1968). Linguistics and the computer analysis of tonal harmony. *Journal of Music Theory*, 12, 2–49.
- Zentner, M. R., & Kagan, J. (1996, September 5). Perception of music by infants [Letter]. *Nature*, 383, 29.
- Zentner, M. R., & Kagan, J. (1998). Infants' perception of consonance and dissonance in music. *Infant Behavior and Development*, 21, 483–492.

Received September 2, 2003

Revision received December 14, 2004

Accepted April 19, 2005 ■