

AN IMPROVED MODEL OF TONALITY PERCEPTION INCORPORATING PITCH SALIENCE AND ECHOIC MEMORY

David Huron
Conrad Grebel College
University of Waterloo

Richard Parncutt
McGill University

The tone profile method of key determination (Krumhansl, 1990) predicts key and key changes in a range of western tonal styles. However, the tone profile method fails to account for certain important effects in tonality perception (Butler, 1989). A modified version of Krumhansl's method of key determination is described that takes into account (a) subsidiary pitches and pitch salience according to Terhardt, Stoll, and Seewann (1982a, 1982b), and (b) the effect of sensory memory decay. Both modifications are shown to improve the correlation between model predictions and experimental data gathered by Krumhansl and Kessler (1982) on the tonality of harmonic progressions. However, the new model described here fails to account for Brown's (1988) experimental findings on the tonality of melodies. The results here are consistent with the view that both *structural* and *functional* factors play a role in the perception of tonality.

Introduction

Brown (1988) has distinguished two approaches to tonality perception, the *structural* and *functional* approaches. According to Brown's distinction, structural approaches to tonality assume that listeners identify tonal centers by integrating the pitch content of a passage and deciding what key best accounts for the distribution of pitch-class material. In typical diatonic music, the most prevalent pitches are the tonic and dominant (diatonic steps 1 and 5), followed by steps 2, 3, 4, and 6, followed by step 7, with non-scale tones least common. The *prevalence* of a pitch-class depends primarily on its accumulated duration. Secondary factors that contribute to the perceptual salience of a pitch-class include repetition and metric stress. Structural approaches assume that the local ordering of pitches is relatively unimportant, and that the passage of time affects tonality perception only to the extent that the decay of sensory memory renders distant pitches less important than recent pitches in the determination of tonal center.

In contrast to this structural approach, Brown distinguishes functional approaches to tonality as placing little emphasis on the pitch-class content of a passage, and more emphasis on sequential intervallic implications. According to functional approaches, it is the implicative context of pitches that determines the listener's perception of tonal center. As Brown has argued, "A single pitch, interval, chord, or scalar passage cannot reign as tonic without a musical context that provides functional relations sufficient to determine its tonal meaning." (Brown, 1988, p. 245)

Experimental evidence can be cited in support of both the structural and functional views of tonality perception. The structural view has received empirical

support through the work of Krumhansl and Kessler (1982). Krumhansl (1990) has shown that the correlation of pitch-class duration/prevalence distributions with empirically-determined key “profiles” provides an effective way of predicting the perceived tonality. Krumhansl regards this correlation as a pertinent observation, but not in itself a “model” or “theory” of tonality. The functional view has received empirical support through work by Butler (1983, 1989) and Brown (1988). Butler and Brown have shown that simply rearranging the order of pitches in a short passage of music is sufficient to evoke quite different key perceptions. Any theory or model of tonality based solely on pitch content without regard to pitch order will be unable to account for the results of Butler and Brown.

The present investigators set out to devise a structural model of tonality by accounting for temporal factors, and by including a more sophisticated model of pitch-class prevalence based on Parncutt (1989). Specifically, we sought to emulate the effects of echoic (or sensory) memory, and to take into account subsidiary pitch perceptions using a modified version of Terhardt’s model of pitch perception (Terhardt, 1979; Terhardt, Stoll, & Seewann, 1982a, 1982b). To foreshadow our conclusions, we will show that (a) simulating the effects of sensory memory does indeed result in a significant improvement to Krumhansl’s key-tracking algorithm, that (b) further significant improvements arise when a psychoacoustic model of pitch salience is added, but that (c) the ensuing model is still unable to account for differing key perceptions arising from the re-ordering of pitch-classes in monophonic pitch sequences. In evaluating our model against experimental data collected by Brown (1988), we will show that our model is able to predict listener perceptions only for those stimuli in which pitch-classes are ordered so as to evoke ambiguous key perceptions. On the basis of our results, we suggest that tonality perception is determined by *both* structural and functional factors.

Improving the Structural Approach to Tonality Perception

Krumhansl (1990) has described a key-finding algorithm that exemplifies the structural approach to tonality perception. Krumhansl’s algorithm correlates the prevalence (summed duration) of notated pitch-classes with two experimentally determined key prototypes (one each for major and minor modes; see Figure 1). These *key profiles* may be regarded as templates against which a given pitch-class distribution can be correlated. By shifting the key profiles through all twelve tonic pitch classes, coefficients of correlation can be determined for each of 24 major and minor keys. The key that shows the highest positive correlation is deemed to be the best key estimate for the given passage.

Butler (1983, 1989) has pointed out that since the key profile method is sensitive only to the pitch content of a passage, this method is unable to account for the harmonic (and tonal) implications of music passages such as shown in Figure 2. Although Figures 2a and 2b contain the same aggregate pitch content, the two passages evoke quite different tonality perceptions. Moreover, even if the content of each vertical sonority is preserved, Butler has pointed out that functional harmonic implications are dependent upon the order of the sonorities. For example, a G major chord followed by a C major chord produces a strong suggestion of a C major tonality. However, a C major chord followed by a G major chord is more

ambiguous in its tonal implications—suggesting either a plagal cadence in G major, or a half cadence in C major.

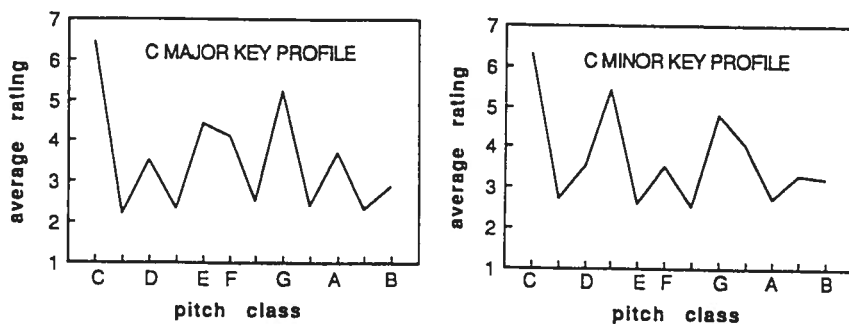


Figure 1. Major and minor key profiles of Krumhansl & Kessler (1982). The vertical axis indicates how well, on average, each pitch-class (realized as a Shepard tone) is judged to follow or fit with a preceding diatonic chord progression in C major or C minor.

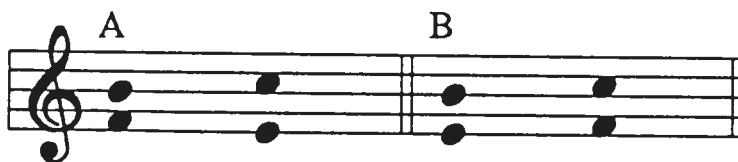


Figure 2. Pairs of dyads including the same four pitch classes (E, F, B, C) but evoking different key implications. After Butler (1983, 1989).

The Role of Sensory Memory in Models of Tonality Perception

The influence of memory on the perception of tonality has been demonstrated by Cook (1987). Cook carried out an experiment in which listeners were asked to judge the *tonal closure* or sense of completion of a passage. Stimuli consisted of performed piano works/movements selected from period-of-common-practice tonal repertoire. In addition to each original music passage, a modified version was prepared in which the work ended in a key other than the original tonic. For passages shorter than about 90 s, listeners were adept at identifying that the tonal center at the end of the passage differed from that at the beginning. Specifically, listeners rated the original tonic versions as more coherent, more expressive, more pleasurable, and providing a better sense of completion than the modified versions. For passages longer than 2 min, however, Cook found that listeners were unable to distinguish between tonic and non-tonic endings. Without the benefit of absolute pitch, it

appeared that listeners' long-term key memory was quite weak.

Cook's work affirms the obvious intuition that more recent sonorities have a greater influence on key determination than sonorities from the distant past. This accords with experimental evidence showing that auditory memory is disrupted by the occurrence of subsequent sounds (Butler & Ward, 1988; Deutsch, 1972; Dewar, Cuddy, & Mewhort, 1977; Massaro, 1970a; Olsen & Hanson, 1977; Wickelgren, 1966)

Psychologists have distinguished a variety of types of memory. Types of memory may be distinguished according to their duration (short term vs. long term), and according to whether the memory is categorical or continuous, willful or spontaneous. In addition, memory types can be distinguished according to the predominant sensory mode (if applicable); that is, visual, auditory, tactile, etc. Sensory memory is usually defined as low level spontaneous memory that occurs without recourse to conscious thought, conscious awareness, abstraction, or semantic processing. Nonsensory memory, by contrast, is often linked to cognitive strategies such as the iterative mental rehearsal of symbolic information. According to common usage of the term "memory," sensory memory is not memory at all, but merely a lingering or resonance of sensory experiences for a brief period of time—typically for no more than a few seconds. Neisser (1967) has dubbed auditory sensory memory "echoic memory" and Massaro (1970b) has spoken of the duration of preperceptual auditory images. Empirical measures of the duration of echoic memory will be reviewed later in this paper, but typical measures lie between 0.5 and 2 s. The duration of visual sensory memory (iconic memory) is an order of magnitude smaller than this; that is, 0.1 to 0.2 s (Averbach & Coriell, 1961).

In addition to being influenced by past events, the perception of key in music may also be influenced by future expectations, especially when the music is familiar to the listener (Jones, 1976). In an analysis of J.S. Bach's *C minor Prelude BWV 871 (Well-Tempered Clavier Book II)*, Krumhansl (1990, p.105) summed together the durations of pitches in three consecutive measures in order to predict the perceived key in the second of these measures. Pitch-class prevalences in the three measures were weighted in the ratio 1:2:1, hence the pitch-class prevalences for a given measure were weighted twice as important as the preceding and subsequent measures. Krumhansl's results compared favorably with judgments of key strength made by music theorists, suggesting that theorists take into account future pitches when making their ratings of key strength. To the extent that this theoretical practice reflects the perceptual experience of listeners, it implies that long-term memory (familiarity) or vivid tonal expectations influence tonality perceptions. In this paper, we propose to limit our efforts to the modeling of key perceptions for unfamiliar pieces where listeners have no overt knowledge or expectations regarding future pitches.

Thompson and Parncutt (1988) simulated the decay of echoic memory for pitch through a simple continuous exponential function. In the present formulation of echoic memory decay, the rate of decay is specified by a half-life variable. Specifically, the perceptual salience for a given pitch is made to decline by the factor $2^{-(\Delta t/\tau)}$ where Δt is the elapsed time between successive sonorities, and τ is the half-life (in seconds). So if the elapsed time is equal to the half-life ($\Delta t = \tau$) then the decay

factor becomes 0.5. Where the half-life value is relatively short (i.e., less than a few seconds) we would presume its effect to be akin to echoic memory. Conversely, where the half-life value is relatively long (say, more than a minute), its effect might be presumed to be akin to a long-term memory.

The Role of Pitch in Models of Tonality Perception

Theoretical accounts of tonality or key perception generally presume that perceived key depends on the perception of pitch. That is, key is assumed to be an emergent experience arising from the relationship between pitches—and that keys cannot be perceived without the prior perception of pitch. Given the importance of pitch information in current models of tonality, it is appropriate to evaluate the quality of the pitch information provided as input to these models. Music scores would appear to provide a ready source of pitch information—and so supply all the necessary cues contributing to the perception of key. However, a number of psychoacoustic phenomena influencing pitch perceptions are not to be found in scores. In the first instance, it is possible for some tones to mask (partially or completely) other concurrently sounding tones—especially in the case of sonorities containing many tones. In the second instance, the salience of pitches is known to change with respect to frequency. Pitches of tones near the center of the music range tend to be more salient than very high or very low pitches (Terhardt, Stoll, & Seewann, 1982a). In addition, the noticeability of pitches is also known to depend on their relative position within a sonority. Experiments have shown that inner voices are less noticeable than outer voices (Huron, 1989), and that this reduced noticeability has influenced the organization of works of music (Huron & Fantini, 1989). For all these reasons, notated pitches may differ with respect to their perceptual saliences.

In addition, subsidiary pitches are known to arise through the interaction of tones—accounting for such phenomena as *residual pitch* (where a pitch is heard that corresponds to a missing fundamental). In the case of music, subsidiary bass pitches arising from complex sonorities often correspond to the root of the sonority (Terhardt, 1974; Parncutt, 1989) and so may play a role in the perception of functional harmony. In short, some notated pitches may remain unheard, whereas other pitch perceptions may arise that are not notated.

These psychoacoustical effects are predicted quantitatively by a model of pitch perception developed by Ernst Terhardt and his colleagues (Terhardt, 1979; Terhardt, Stoll, & Seewann, 1982a, 1982b). Using Terhardt's model, it is possible to adjust the score information so that it better reflects the listener's experience of the pitch content of a passage.

Of course, it is possible that such low level psychoacoustical phenomena have no effect on higher level perceptions, such as the perception of tonality. In order to test whether or not subsidiary pitches affect tonality, we modified Krumhansl's algorithm by incorporating a simplified version of Terhardt's pitch model (Parncutt, 1989) at the input. Scores were reconstituted to better reflect the saliences of the various pitch perceptions. The resulting pitch data provided the basis for tabulating pitch-class prevalence. The results of our modified key tracking model were then compared with corresponding results from Krumhansl's original algorithm.

Pitch Salience

Although Terhardt's model of pitch perception has been highly influential in auditory perception research, his work remains poorly understood among researchers working in the field of music perception. There is some merit, therefore, in reviewing the basic elements of the model. Broadly speaking, Terhardt's model of pitch entails three stages of analysis. The first stage analyzes the input signal and determines the audibility of the various frequency components. The second stage looks for frequency components that may be perceived as harmonics of complex tones, then determines the pitches of the corresponding fundamentals. The third stage combines the various pitch components into a single spectrum representing the listener's subjective experience of the sound.

More specifically, stage 1 accepts some spectrum specifying frequency and sound pressure level information (in dB SPL). In order to account for masking, each spectral component is taken in turn, and the masking effect of all other components is estimated and combined. In the present study, a masking algorithm was implemented in which critical band rate was specified by Moore and Glasberg's ERB-rate scale (1983), and each component was assigned a symmetrical triangular masking pattern with upper and lower slopes of 12 dB per critical band. Analyzing both the influence of the threshold of hearing and the masking of concurrent spectral components results in a measure Terhardt calls "SPL excess," that is, sound pressure level (in dB) above masked threshold. If SPL excess is positive, then the corresponding component is predicted to be audible. If it is not positive, it is assumed inaudible, and is omitted from all further calculations.

Two further factors influence the audibility of pure tone components. First, the audibility (or perceptual clarity, or salience) of a pure tone (or pure tone component) does not vary linearly with respect to changes in SPL excess, except at very low levels (under about 20 dB). At higher levels, the audibility of a tone becomes saturated; that is, continuing to increase the sound pressure level will not make the tone any more audible. In Terhardt's model, the SPL excess above masked threshold is thus modified by a *saturation function*. The saturation function scales audibility values so that they lie between 0 (inaudible) and 1 (maximum audibility, or maximum perceptual clarity). At low values of SPL excess, the audibility of pure tone components grows proportionally; at high values of SPL excess, audibility gravitates toward the saturation level of 1.0.

Listener sensitivity to the pitch of pure tones is also known to change with frequency. The greatest sensitivity occurs in a broad region centered at about 700 Hz (F5). This is accommodated in the model by weighting the output of the above saturation function according to a "spectral dominance" filter. The effect of spectral dominance on the pitch of complex tones was studied experimentally by Plomp (1967) and Ritsma (1967).

The result of these calculations is a measure which Terhardt calls the "spectral pitch weight" for each pure tone component in the spectrum. Elsewhere, this has been referred to as "pure tone audibility" (Parncutt, 1989). In summary, stage 1 of Terhardt's pitch model determines the relative audibility (or relative salience) of each of the pure tone components in the spectrum. The result of this processing is a graph relating audibility (pitch weight) to spectral pitch for pure tone components

in the signal.

Stage 2 of the pitch model takes the results of stage 1 and attempts to identify complex tones that may be present in the signal. This is done using a harmonic pitch-pattern recognition procedure. The spectrum is scanned for the presence of audible tone components whose frequencies correspond approximately to the harmonic series. When an approximately harmonic set of components is found, a *virtual pitch* is assigned to the fundamental of the series. In addition, each virtual pitch is assigned a weight—a measure of goodness of fit to the harmonic series “template.”

In stage 3, a composite pitch spectrum is created by combining the spectral (stage 1) and virtual (stage 2) pitch information. Sometimes, spectral and virtual pitches coincide or lie very close to each other. For example, the spectral pitch of the first harmonic of a complex tone coincides with the overall virtual pitch of the tone. When a spectral and a virtual pitch lie very close to each other, the pitch having the higher pitch weight is assumed to dominate the pitch having the lower pitch weight. This effect is dependent upon the degree of “analytic listening;” that is, the degree to which the listener’s attention is drawn to spectral rather than virtual pitches.

The result of stage 3 analysis is a composite portrait of all pitches which *could* be perceived in a given acoustical spectrum. For a given listener, in a given context, only a particular subset of these possible pitches will be heard—typically those pitches, or that pitch, having the highest pitch weight.

Implementation and Initial Tests

We revised Krumhansl’s tone-profile method of key determination so that notated pitch inputs were modified according to the above psychoacoustic model. In effect, we “performed” music scores using complex tones and then calculated the ensuing pitch perceptions and their respective saliences. To the extent that Terhardt’s model of pitch perception is successful, adjusting the score information according to this model ought to better reflect the listener’s experience of pitch-class prevalence.

As can be seen in Table 1, the “best” chord progression (from the point of view

Figure 3 shows the results of applying this revised model to Butler’s F-B, E-C dyads (shown previously in Figure 2). Pitch salience is represented in Figure 3 by the size of the noteheads. In the case of example 3a, subsidiary pitch perceptions for the F-B dyad include Db, B, and G (the presumed theoretical root). The ensuing E-C dyad includes the subsidiary pitches A and C (again the presumed theoretical root). In contrast, example 3b displays no subsidiary perceptions for the pitch G and only weak subsidiary perceptions for the pitch C; instead, the presumed theoretical roots E and F show a greater salience. The predicted key correlations from the model (Figure 3c) tend to reflect our musical intuitions about the tonality perceptions evoked by these re-ordered pitches. Progression (a) implies C major more strongly than progression (b). This informal result suggests that carefully emulating the perception of pitch may improve models of tonality perception.¹

Recall that Butler (1983, 1989) also identified the influence of order on tonality perceptions. For example, it was noted that a G chord followed by a C chord evokes significantly different tonality perceptions than the reverse ordering of these two

chords. In order to account for such order effects, the psychoacoustic pitch model was augmented by the addition of the model of echoic memory described earlier. By weighting the perceptual salience of recent events more heavily than past events, any hypothetical tonal sequence (A→B) will typically evoke a different experience than its reverse ordering (B→A).

Using this combined model, we asked what two-chord progression (ending with C) best suggests a C major tonality? That is, we asked what preceding chord is best able to define a given tonic. In order to avoid the confounding influence of chord spelling (inversions, spacings, doubling) and voice-leading, we used chords consisting of octave-spaced (Shepard) tones as input to the model. Each chord was assigned a duration of 0.5 s, with an echoic memory half-life of 1.0 s. The pitch saliences of the first chord were thus multiplied by the factor $2^{-0.5/1} = 0.71$ before being input to the pitch model; the second (most recent) chord was unaffected by memory decay. The weighted pitch saliences for the two chords were added, and the totals were compared with the 24 key profiles of Krumhansl and Kessler (1982). The results are shown in Table 1. Each coefficient indicates the strength of the correlation between the calculated pitch content and Krumhansl's key prototypes.

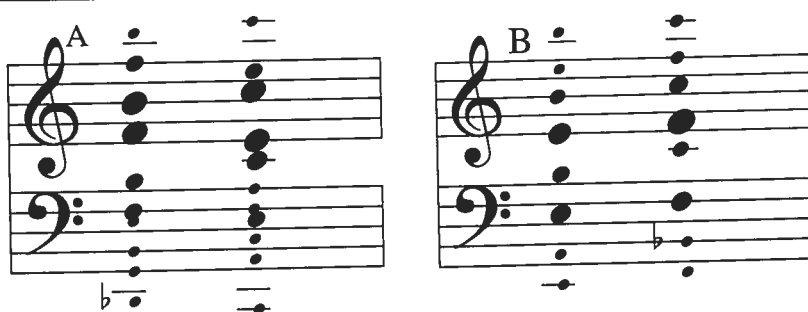


Figure 3a,b. Schematic representation of the pitch content for the dyads shown in Figure 2, calculated using the model of Parncutt (1989) with complex tone inputs. Notehead area is roughly proportional to calculated pitch salience. Only pitches in range C2 to C6 with calculated saliences exceeding 0.05 are included. No account has been taken of sensory memory decay.

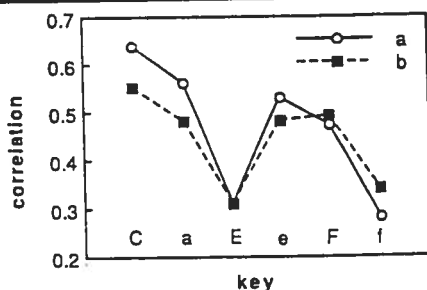


Figure 3c. Correlation coefficients between calculated tone profiles for progressions (a) and (b) and key profiles for Krumhansl & Kessler (1982) for the keys shown. Major and minor keys are indicated by upper- and lower-case letters respectively.

of defining a given key) is the dominant-tonic (G-C) progression. This result is not surprising, of course, as Krumhansl measured the key profiles using V-I progressions. The next best key defining progression is a three-way tie between the I-I, ii-I and vii^o-I progressions. Of the progressions involving diatonic chords from the major key (I, ii, iii, IV, V, vi, vii^o), the IV-I (F-C) and vi-I (a-C) progressions are the weakest (according to the model) in terms of defining the tonic key. In addition, both progressions are relatively ambiguous in their key implications: the F-C progression implies both the key of C (0.86) and the key of F (0.81), while the progression a-C implies both the key of C (0.86) and the key of a (0.83). In other words, in the absence of additional key-defining context, the IV-I and vi-I progressions exhibit a considerable degree of tonal ambiguity—according to the model.

Table 1
Key Correlations for Diatonic Two-Chord Progressions Ending with C Major

Progression	Key Correlations								
	C	c	d	e	F	f	G	g	a
C-C	0.89	0.66	-0.06	0.40	0.43	0.25	0.36	0.09	0.55
c-C	0.81	0.86	-0.15	0.29	0.37	0.31	0.31	0.18	0.39
d-C	0.89	0.48	0.58	0.19	0.78	0.27	0.46	0.30	0.66
D-C	0.79	0.45	0.38	0.33	0.40	-0.08	0.61	0.25	0.55
e-C	0.88	0.44	-0.12	0.77	0.24	0.03	0.55	0.10	0.56
E-C	0.74	0.32	-0.17	0.68	0.16	0.13	0.28	-0.19	0.57
F-C	0.86	0.53	0.26	0.13	0.81	0.55	0.14	0.04	0.62
f-C	0.75	0.62	0.02	0.05	0.61	0.70	0.02	-0.07	0.44
G-C	0.90	0.63	0.00	0.59	0.31	0.07	0.74	0.43	0.38
g-C	0.81	0.70	0.09	0.37	0.40	0.10	0.58	0.60	0.28
a-C	0.86	0.45	0.13	0.39	0.52	0.14	0.30	-0.02	0.83
A-C	0.77	0.28	0.11	0.39	0.42	0.00	0.29	-0.06	0.87
b ^o -C	0.89	0.52	0.21	0.48	0.58	0.25	0.58	0.21	0.44
b-C	0.75	0.41	-0.02	0.60	0.18	-0.16	0.66	0.05	0.40
B-C	0.64	0.54	-0.37	0.55	0.09	-0.04	0.40	-0.15	0.36

Note. Each entry in the table is a correlation coefficient (r), obtained by (a) calculating the pitch-class salience profile (according to Parncutt, 1989) for each chord in the progression, (b) multiplying the values for the first chord in each pair by 0.71, (c) adding the values for the second chord, and (d) comparing the result with the key profiles of Krumhansl & Kessler (1982) for the keys indicated at the top of each column. Upper case letters indicate major triads or keys; lower case letters indicate minor triads or keys. Diminished triads are indicated by a circular superscript.

Considering the above tentative and preliminary analyses, the model might appear to capture some of the order effects in tonality perception identified by Butler. Given these initial results, a more formal evaluation of our model seemed warranted.

Evaluation of the Model

To evaluate our model, we compared outputs from the model with listener responses from published tonality experiments by Krumhansl and Kessler (1982) and by Brown (1988). Krumhansl and Kessler's stimuli consist of harmonic chord progressions, whereas Brown's stimuli consist of monophonic pitch sequences.

Structural Tonality: Comparison with Krumhansl and Kessler

In the first instance, 10 chord progressions studied by Krumhansl and Kessler (1982) were given as input to the model (see Krumhansl, 1990, pp. 218-228). In the original experiments, stimuli consisted of a sequence of 9 triads constructed using octave-spaced tones. Each chord had a duration of 0.5 s followed by a 0.125 s silence. In our simulation, each chord was given a nominal duration of 0.625 s. The model was given octave-spaced tones as input.

In Krumhansl and Kessler (1982), changes in perceived tonality over the course of the progressions were traced using the probe-tone technique. Initial trials consisted of just the first chord in the progression followed by 12 probe tones that corresponded to each of the 12 pitch-classes. Listeners rated how well each probe tone fit with the preceding context, thus generating tone profiles. In a second set of trials, this procedure was repeated with probe tones following the first two chords in the progression. This procedure was repeated in subsequent trials until tone profiles were generated following each chord in the progression.

Two sets of chord progressions were investigated by Krumhansl and Kessler. One set consisted of chords deemed consistent with a single key. A second set consisted of chords deemed to modulate from an initial key to some other final key. Krumhansl and Kessler presented their results in the form of graphs showing the changes in correlation between successive tone profiles (collected from their listeners) and a given key profile. For example, Figure 4 shows the results for a chord progression considered consistent with the key of C major. The solid line traces the changes in correlations between the probe tone ratings at each serial chord position and the C major key profile (as reproduced in Figure 9.2 of Krumhansl, 1990). As can be seen, the strongest key correlations coincide with the plagal and authentic cadence points. The dotted line indicates the output from our model for the same input. That is, in our simulation, key correlation values were output by the model following each successive chord in the progression.

In using our model, it is necessary to define a "decay" value representing the rate of echoic memory loss. The values plotted in Figure 4 resulted from a memory simulation in which the half-life was given a value of 0.9 s. The correlation between our model's output for this progression and the key correlation data of Krumhansl and Kessler is +0.90.

Of course, the correlations between the theoretical and experimental results

change according to the memory half-life value. In the case of Krumhansl and Kessler's progression in C minor, for example, the correlation between our calculations and their key data varies between +0.47 to +0.78 as half-life values are varied over the range 0.2 to 4.0 s. As in the case of the C major progression, the optimal fit ($r = +0.78$) was found to occur for a half-life of 0.9 s.

In the case of modulating progressions, Krumhansl and Kessler presented key correlations for both the initial key and final key. As the music modulates away from the initial key, the initial key correlations decrease, and the final key correlations increase. In evaluating our model, we calculated correlations for both the initial and final keys over the course of the modulating progressions.

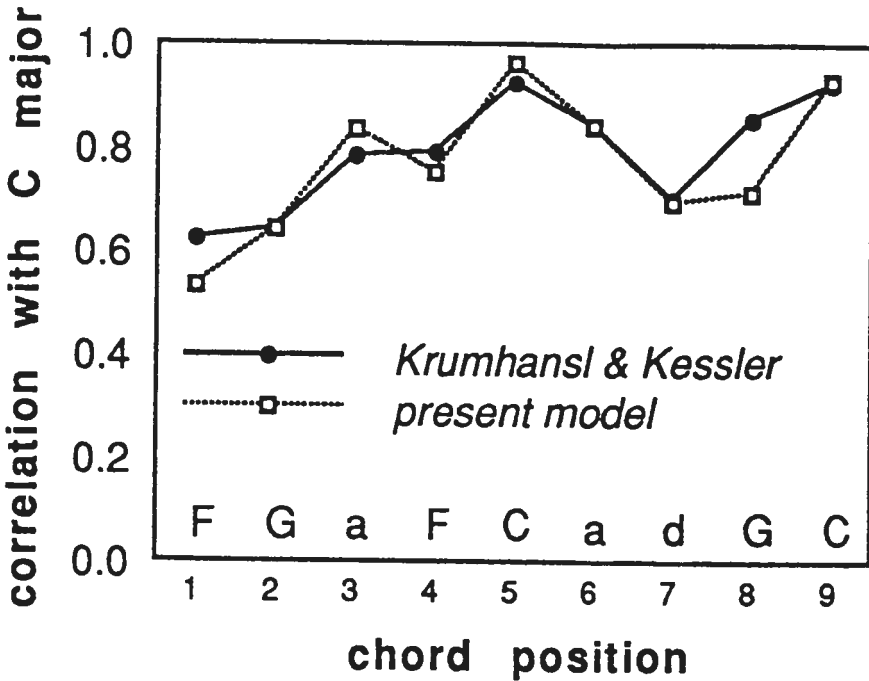


Figure 4. Comparison of experimental data of Krumhansl and Kessler with calculations according to the present model, for a chord progression in C major. Major and minor chords are indicated by upper- and lower-case letters respectively. Chord progression proceeds from left (F) to right (C). The solid line represents correlations between the experimentally determined probe tone ratings at each serial chord position and the C major key profile. The dotted line represents corresponding output from the theoretical model described in the text.

To evaluate the significance of pitch salience and subsidiary pitches in tonality perception, we compared our model with a modification of Krumhansl's algorithm in which echoic memory, but not pitch salience, was included. The half-life value for the echoic memory decay was then systematically varied in order to find the optimal correlation. Table 2 presents the results for both models.

Table 2

Correlation Coefficients (r) Between Results of Krumhansl and Kessler (1982) and Calculations, According to Two Models (With and Without Pitch Salience)

Key Areas of Chord Progression	Pitch Salience Included		Pitch Salience Excluded	
	Optimal half - life (s)	Correlation coefficient (r)	Optimal half - life (s)	Correlation coefficient (r)
C major	0.9	0.90	2.1	0.94
c minor	0.9	0.78	∞	0.53
C→G	0.9	0.82	9.0	0.81
C→a	0.8	0.81	∞	0.66
c→f	2.4	0.82	1.0	0.73
c→C	0.6	0.91	0.4	0.81
c→Ab	1.6	0.85	1.3	0.73
C→d	2.1	0.74	2.2	0.71
C→Bb	0.7	0.93	0.6	0.85
c→c#	0.6	0.94	0.6	0.84
<i>M</i>	1.15	0.85	2.15*	0.76
<i>SD</i>	0.65	0.07	2.85*	0.11

Note. * Calculations exclude the infinity values.

As can be seen in Table 2, the inclusion of subsidiary pitches and pitch salience information produces results that better correlate with listener perceptions than inclusion only of notated pitches (on the average, $r = +0.85$ vs. $r = +0.76$). In addition, the optimal half-life values are more stable in the pitch salience model than in the simple notated pitch-class prevalence model. This stability is evident by comparing the standard deviations of the optimum half-life values for both models: 0.65 s in the case of the pitch salience model vs. 2.85 s in the case of the pitch-class prevalence model (excluding two infinite values). The optimum values in the pitch salience model (left column of Table 2) suggest a typical half-life value for echoic memory of about 1 s.

It is useful to compare this optimum half-life value with experimental measures of the duration of echoic memory. Triesman and Howarth (1959) showed that an uncoded auditory input could be retained for between 0.5 and 1 s. Results from Guttman and Julesz (1963) suggest a duration of roughly 1 s—with a maximum of 2 s. Triesman (1964) found 1.3 s, or less than about 1.5 s. Crowder (1969) estimated decay time at between 1.5 and 2 s. Glucksberg and Cowen (1970) found a duration of less than 5 s. Triesman and Rostron (1972) found a "brief auditory store whose contents are lost in about 1 sec." Their curves were near asymptote at 1.6 s. Darwin, Turvey, and Crowder (1972) reported a value of "something greater than 2 sec but less than 4." Rostron (1974) found that "the decay was mainly finished within a second of the end of the stimulus presentation." On the basis of a rather complex experiment, Kubovy and Howard (1976) concluded that "1 sec. ...represents a lower bound on the average half-life of echoic memory." Massaro and Idson (1977)

studied backward recognition masking, and found that correct identification of the relation between the frequencies of a target tone and a masking tone approached optimal performance as the duration of the silent intertone interval approached 0.5s.

Note that the experimental measures of the duration of echoic memory typically represent the time taken for performance on memory tasks to fall to chance level. This time period considerably exceeds the corresponding half-life value. The experimental literature thus suggests a half-life value for echoic memory of considerably less than 1 s. It is possible that our simulated half-life value of around 1 s incorporates elements of a longer-term form of memory—such as short-term memory as defined by Deutsch (1975), or the psychological present (Fraisse, 1963; Michon, 1978).

Functional Tonality: Comparison with Brown

In addition to testing the model against data collected by Krumhansl and Kessler, we also tested our model against data collected by Brown (1988). Brown carried out a number of experiments in which listeners were asked to sing the tonic after hearing a monophonic sequence of pitches. The pitch-classes used were drawn from brief music excerpts. Duplication of pitch-class was avoided in the stimuli. The main experimental manipulation concerned the ordering of pitches within a sequence. For each pitch sequence, three different orderings were presented, denoted by the letters A, B, and C. The A orderings were designed to evoke a tonal center consistent with that of the original music excerpt from which the pitch-classes were extracted. The B orderings were designed to evoke a tonal center different from that of the original excerpt. The C orderings were designed so as to produce tonally ambiguous perceptions; that is, Brown purposely arranged the order of the pitch-classes so as to elicit the greatest variety of tonic responses from her listeners.

For each pitch sequence, Brown collected data indicating the number of listeners who responded by singing a given pitch-class as their chosen tonic. We tested our model using the nine experimental sequences described in Brown (1988). The nine pitch sequences correspond to three different orderings of three different pitch-class sets. We used complex tone inputs with an echoic memory decay half-life of 1 s and tone durations of 0.5 s. Twenty-four correlation coefficients were calculated for each sequence—each coefficient pertaining to one of 12 major or 12 minor keys. In Brown's work, tonic data were gathered without reference to a major or minor modality. To compare the coefficients from our model with Brown's experimental results, the parallel major and minor coefficients for each pitch-class were summed together (i.e., C major + C minor, C# major + C# minor, etc.)—resulting in twelve "tonic pitch" coefficients. Amalgamating the major and minor modes in this way may be expected to reduce the overall correlation values.

The twelve tonic coefficients from our model then were correlated with Brown's data, which identified the number of times different pitch-classes were chosen as tonics. For example, Table 3 compares model predictions with listener tonic responses for Brown's stimulus sequence "2C." The coefficient of correlation between the responses by Brown's listeners and the tonic coefficients produced by the model is +0.67—much lower than the correlation coefficients found when modeling Krumhansl and Kessler's data.

To test the effectiveness of our model, we compared the model outputs for each of the three pitch orderings (A, B, C) with the listener responses for the same three stimuli. The results for the A, B, and C orderings for Brown's pitch-set No. 2 are given in the correlation matrix shown in Table 4. A successful model would produce the largest correlation values in the diagonal table positions: A-A, B-B, and C-C.

Of the 9 sequences (3 orderings of 3 pitch-set sequences), only 2 correlations showed the predicted maxima in the diagonal positions. In short, our model was entirely unable to distinguish between the listener responses for the different pitch-set orderings.

Tonal melodies often imply an associated harmonic progression. Music experience suggests that the key of a tonal melody is the same as the key of the chord progression (or progressions) that the melody implies. A possible reason for the failure of our model to predict the key of a melody thus may involve the chord progression implied by the melody.

Table 3
Comparison of Model Predictions With Listener Tonic Responses for Brown's "Stimulus 2c"

Tonic	Responses	Model
C	14	+0.63
C#	0	-1.00
D	3	+1.40
D#	2	-0.52
E	0	-0.24
F	3	-0.04
F#	0	-0.57
G	9	+1.37
G#	0	-0.89
A	5	+0.30
A#	1	-0.25
B	2	-0.17

Note. "Responses" are the number of times Brown's listeners reported a given tonic. "Model" is the predicted tonic salience of each tonic candidate according to the model described in the text.

Table 4
Correlation Matrix for Three Pitch-Orderings in Brown's "Stimulus No. 2."

Stimulus	Model		
	2A	2B	2C
2A	0.70	0.27	0.26
2B	0.46	0.23	0.57
2C	0.88	0.55	0.67

Note. Row labels refer to experimental results, and column labels refer to predictions according to the model.

In a crude attempt to account for possible implied chord progressions, we tried a technique that may be dubbed "pitch smearing." Rather than treat the input sequences as simple successions of pitches, we tried overlapping successive sonorities so that implied harmonic relationships between successive pitches might emerge. In one approach, we simulated 0.5 s of tonal "smearing." In effect, pairs of successive pitches were treated as harmonic dyads. The results of these simulations were marginally different from those found without the smearing. Foregoing any subtlety, we tried amalgamating successive tones to form three-pitch sonorities. For example, the pitch sequence: V, W, X, Y, Z, would be given to the model as V, W+V, X+W+V, Y+X+W, and Z+Y+X. Although the key correlations improved slightly, our model was still unable to predict which tonic responses corresponded to a given re-ordering of the pitch-class content. These results suggest either that the implied chord progression hypothesis is false, or that extra sophistication is needed in identifying the specific implied chords and the transition points between successive chords.

In our evaluation of the model against Brown's experiments, one unexpected trend was observed. The results of our model consistently showed higher correlations with the listener responses for Brown's C orderings of the pitch-classes (see, for example, the third row in Table 4). Amalgamating the results for all of Brown's stimuli, the A row correlations showed an average correlation of +0.19, the B row showed an average of +0.36, and the C row showed an average of +0.59. In order to confirm this observed trend, a chi-square analysis was carried out. Chi-square values were calculated by comparing the ratio of C stimuli showing maximum correlations, versus A and B stimuli showing maximum correlations. This analysis confirmed the significant difference of the C stimuli at the $p < 0.05$ confidence level.

Similar results were found using "pitch smearing." Once again, we found consistently higher correlations for the C orderings. Amalgamating the results from smearing three pitches and comparing them with Brown's stimuli, the A row coefficients showed an average correlation of +0.36, the B row showed an average of +0.39, and the C row showed an average of +0.64 (for tones assumed to be 1.0 s in duration). A chi-square analysis again confirmed the C row correlations to be significantly higher ($p < 0.05$).

Recall that Brown's C orderings were created to produce tonally ambiguous perceptions. That is, Brown arranged the order of the pitches to avoid or obscure functional implications. The fact that our model was better able to predict the tonic responses of listeners for these stimuli suggests that the model is able to account only for nonfunctional aspects of tonality perception. The model appears to have little or no utility in accounting for functional aspects of tonality perception.

Discussion

Clearly, our model of tonality perception was better able to account for Krumhansl and Kessler's experimental data than for Brown's data. Initially, this difference in the performance of the model would seem to be due to a difference in texture: Brown's stimuli were unaccompanied melodies whereas Krumhansl and Kessler's stimuli were chord progressions. However, the heightened performance

of our model in Brown's ambiguous pitch sequences suggests that the model is able to capture at least some of the aspects of tonality perception evident in monophonic sequences.

When a modification to a model produces significant improvements on the one hand, while failing to account for other aspects of a phenomenon, there is good reason to suspect that what is presumed to be a single phenomenon may entail two separate aspects or dimensions. As in the case of depth perception in human vision, tonality perception may arise through separate synchronic and diachronic processes. Functional and structural theories of tonality perception may thus be complementary rather than contradictory (as previously supposed). Overall, the results of this paper are consistent with the hypothesis that tonality perception is determined by both structural and functional factors, and that our modeling efforts succeeded in further illuminating only the structural aspects of tonality perception.

Conclusion

In summarizing our investigations, we found that

- (a) Krumhansl's structural approach to key perception can be significantly improved by incorporating a psychoacoustic model of pitch perception. Specifically, pitch salience information enhances the correspondence between the model predictions and listener responses for harmonic stimuli.
- (b) simulating the effect of tonal memory decay is able to account for some temporal phenomena in the perception of tonality.
- (c) the above improvements notwithstanding, the ensuing model is still unable to account for different key implications arising from re-ordering pitch-classes in monophonic pitch sequences.
- (d) in evaluating our model against experimental data collected by Brown (1988), our model was able to predict listener tonic responses only for those stimuli in which pitch-classes were ambiguously ordered. In the case of pitch sequences organized to suggest specific tonal implications, the outputs from our model failed to conform to listeners' responses.

The results are consistent with the view that both "structural" and "functional" factors play a role in the determination of key or tonality perceptions, and that our model predicts only the structural aspects of tonality.

References

- Averbach, E., & Coriell, A. S. (1961). Short-term memory in vision. *Bell Systems Mechanical Journal*, 40, 309-328.
- Brown, H. (1988). The interplay of set content and temporal context in a functional theory of tonality perception. *Music Perception*, 5, 219-250.
- Butler, D. (1983). The initial identification of tonal cues in music. In J. Sloboda & D. Rogers (Eds.), *Acquisition of symbolic skills*. pp. 251-261. New York: Plenum Press.

- Butler, D. (1989). Describing the perception of tonality in music: A critique of the tonal hierarchy theory and a proposal for a theory of intervallic rivalry. *Music Perception*, 6, 219-242.
- Butler, D., & Ward, W. D. (1988). Effacing the memory of musical pitch. *Music Perception*, 5, 251-260.
- Cook, N. (1987). The perception of large-scale tonal closure. *Music Perception*, 5, 197-205.
- Crowder, R. G. (1969). Improved recall for digits with delayed recall cues. *Journal of Experimental Psychology*, 82, 258-262.
- Darwin, C. J., Turvey, M., & Crowder, R. G. (1972). An auditory analogue of the Sperling partial report procedure: Evidence for brief auditory storage. *Cognitive Psychology*, 3, 255-267.
- Deutsch, D. (1972). Effect of repetition of standard and comparison tones in recognition memory for pitch. *Journal of Experimental Psychology*, 93, 156-162.
- Deutsch, D. (1975). The organization of short-term memory for a single acoustic attribute. In D. Deutsch & J. A. Deutsch (Eds.), *Short-term memory*. (pp. 108-151). New York: Academic Press.
- Dewar, K. M., Cuddy, L. L., & Mewhort, D. J. K. (1977). Recognition memory for single tones with and without context. *Journal of Experimental Psychology: Human Learning and Memory*, 3, 60-67.
- Fraisse, P. (1963). *The psychology of time*. New York: Harper and Row.
- Glucksberg, S., & Cowen, G. N., Jr. (1970). Memory for nonattended auditory material. *Cognitive Psychology*, 1, 149-156.
- Guttman, N., & Julesz, B. (1963). Lower limits of auditory periodicity analysis. *Journal of the Acoustical Society of America*, 35, 610.
- Huron, D. (1989). Voice denumerability in polyphonic music of homogeneous timbres. *Music Perception*, 6, 361-382.
- Huron, D., & Fantini, D. A. (1989). The avoidance of inner-voice entries: Perceptual evidence and musical practice. *Music Perception*, 7, 43-48.
- Jones, M. R. (1976). Time, our lost dimension: Toward a new theory of perception, attention, and memory. *Psychological Review*, 83, 323-355.
- Krumhansl, C. L. (1990). *Cognitive foundations of musical pitch*. Oxford: Oxford University Press.
- Krumhansl, C. L., & Kessler, E. J. (1982). Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys. *Psychological Review*, 89, 334-368.
- Kubovy, M., & Howard, F. P. (1976). Persistence of pitch-segregating echoic memory. *Journal of Experimental Psychology: Human Perception & Performance*, 2, 531-537.
- Massaro, D. W. (1970a). Retroactive interference in short-term recognition memory of pitch. *Journal of Experimental Psychology*, 83, 32-39.
- Massaro, D. W. (1970b). Preperceptual auditory images. *Journal of Experimental Psychology*, 85, 411-417.
- Massaro, D. W., & Idson, W. L. (1977). Backward recognition masking in relative pitch judgments. *Perceptual and Motor Skills*, 45, 87-97.
- Michon, J. A. (1978). The making of the present: A tutorial review. In J. Requin (Ed.), *Attention and performance* (Vol. 7). New York: Academic Press.
- Moore, B., & Glasberg, B. (1983). Suggested formulae for calculating auditory-filter bandwidths and excitations patterns. *Journal of the Acoustical Society of America*, 74, 750-753.
- Neisser, U. (1967). *Cognitive Psychology*. New York: Meredith.
- Olsen, R. K., & Hanson, V. (1977). Interference effects in tone memory. *Memory and Cognition*, 5, 32-40.
- 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.

- Plomp, R. (1967). Pitch of complex tones. *Journal of the Acoustical Society of America*, 41, 1526-1533.
- Ritsma, R. J. (1967). Frequencies dominant in the perception of the pitch of complex tones. *Journal of the Acoustical Society of America*, 42, 191-198.
- Rostron, A. B. (1974). Brief auditory storage: Some further observations. *Acta Psychologica*, 38, 471-482.
- Terhardt, E. (1974). Pitch, consonance, and harmony. *Journal of the Acoustical Society of America*, 55, 1061-1069.
- Terhardt, E. (1979). Calculating virtual pitch. *Hearing Research*, 1, 155-182.
- Terhardt, E., Stoll, G., & Seewann, M. (1982a). Pitch of complex signals according to virtual-pitch theory: Test, examples, and predictions. *Journal of the Acoustical Society of America*, 71, 671-678.
- Terhardt, E., Stoll, G., & Seewann, M. (1982b). Algorithm for extraction of pitch and pitch salience from complex tonal signals. *Journal of the Acoustical Society of America*, 71, 679-688.
- Thompson, W. F., & Parncutt, R. (1988). *Using a memory-fade model to track the movement of musical keys*. Contributed paper at the International Congress of Psychology, Sydney.
- Triesman, A. M. (1964). Monitoring and storage of irrelevant messages in selective attention. *Journal of Verbal Learning and Verbal Behavior*, 3, 449-459.
- Triesman, A. M., & Howarth, C. I. (1959). Changes in threshold level produced by a signal preceding or following the threshold stimulus. *Quarterly Journal of Experimental Psychology*, 11, 129-142.
- Triesman, A. M., & Rostron, A. B. (1972). Brief auditory storage: A modification of Sperling's paradigm applied to audition. *Acta Psychologica*, 36, 161-170.
- Wickelgren, W. A. (1966). Consolidation and retroactive interference in short term recognition memory for pitch. *Journal of Experimental Psychology*, 72, 250-259.

Footnote

¹Note that Figure 3a shows greater calculated saliences for the pitches B4 and C5 than is the case for the same tones in Figure 3b. The reason for the difference involves normalization of pitch salience. Parncutt (1989, p. 93, eq. 4.25) adjusted the calculated salience of tone sensations within a sonority in such a way that the sum of the saliences of all tone sensations evoked by a sonority equals the calculated multiplicity of that sonority; that is, the estimated mean number of simultaneously noticed tones. The calculated multiplicity of the F-B and E-C dyads in Figure 3a is greater than that E-B and F-C dyads in Figure 3b, consistent with the idea that a perfect fifth dyad blends more easily into one sound than do dyads such as the tritone and the minor sixth. To account for the difference in blending of the two intervals, the total area of the noteheads in Figure 3a is greater than the total area of the noteheads in Figure 3b.

Author Note

Reprint requests should be sent to Dr. David Huron, Conrad Grebel College, University of Waterloo, Waterloo, Ontario, Canada N2L 3G6.