THE EFFECT OF STRUCTURE AND RATE VARIATION ON KEY-FINDING

Morwaread Farbood, Gary Marcus†, Panayotis Mavromatis*, and David Poeppel†§*

*Music and Audio Research Laboratory (MARL)
†Department of Psychology
§Center for Neural Science
New York University

ABSTRACT

While substantial evidence suggests that key perception depends in part on the statistical distribution of pitch classes in a given piece, it has also been suggested that structural factors may play an important role. To further the understanding of what such structural factors might be, we conducted two experiments using statistically ambiguous yet structurally unambiguous stimuli that consisted of a uniform distribution of diatonic pitches drawn from the union of two neighboring keys, sequenced to clearly imply one particular key. In Experiment 1, subjects with substantial music theory training were asked to listen to melodic sequences and identify the key. The results indicate that purely structural factors can greatly influence key perception when statistical information about overall key profile is highly ambiguous. In Experiment 2, we examined the temporal psychophysics of structurally-based key induction by systematically varying the tempo of a subset of the statistically ambiguous/structurally unambiguous materials presented in Experiment 1 at rates ranging from 7 BPM to 3400 BPM. Twenty-two musically trained subjects were asked to indicate whether each sequence sounded resolved (ending on an implied tonic) or unresolved (ending on an implied dominant). The results agree strongly with those from Experiment 1 and show a preferred range of tempi in which participants provide robust key judgments (30 BPM - 400 BPM). This suggests that structurally-based key-finding can be accurately computed by trained listeners in as little as 150ms per note or 1.2s for an entire sequence of eight notes.

1. INTRODUCTION

Key perception is often studied through the examination of the statistical distribution of pitch classes (Longuet-Higgins & Steedman, 1971; Krumhansl & Kessler, 1982; Vos & Van Geenen, 1996; Yoshino & Abe, 2004; Temperley, 2007; Temperley & Marvin, 2008). The results of these studies have indicated that there is a significant link between statistical key profiles and key identification.

Other studies have indicated that *structural* factors play an important role in key identification (Brown, 1988; Brown, Butler, & Jones, 1994; Butler, 1989; Vos, 1999; Matsunaga & Abe, 2005). However, it remains unclear what precisely these structural features are. In a recent study by Matsunaga and Abe (2009), listeners were asked to identify keys for 450 melodies that consisted of the same pitch set—the intersection of two closely related keys such as G major and C major—in various permutations. The local properties systematically examined included position of particular pitch classes within a sequence, position of a particular interval, and position of intervals separated by one to four interposed tones. The results suggested that except for the final tone, none of the specific local properties examined contributed significantly to key identification. The authors thus concluded that key identification is derived from something other than specific local properties. However, these properties are somewhat arbitrary from a functional perspective and do not explicitly consider harmonic implications and contour. The melodic sequences are completely free of any tritones, rendering a major aspect of key induction unavailable. This might partially explain the weak results.

The goal of this paper is to explore structural features of melodies that contribute to key identification. We take a different methodological approach from previous studies, one that attempts to first define structural features important to tonal induction that are based on music-theoretic principles. The work outlined here describes the first two experiments in an ongoing series of studies.

2. EXPERIMENT 1

The purpose of Experiment 1 was to produce a set of melodic sequences that had *statistically ambiguous key profiles* with *strong structural cues* that implied a particular key in an empirically verified way. Unlike the sequences used in Matsunaga and Abe's experiments (2005, 2009), the melodies for this study contained the *union* of two fifth-related keys rather than the intersection, e.g. [C, D, E, F, F#, G, A, B] for C major and G major. All of the sequences contained the same set of relative pitch classes and the same cardinality, the only difference being the ordering of the pitches.

2.1 Materials

A total of 31 melodic sequences were composed, 10 of them intended to strongly imply Key Type I (e.g. C major for the C major/G major pitch set), 10 of them intended to strongly imply Key Type II (e.g. G major for the C major/G major pitch set), and 11 designed to be ambiguous.

The melodies were composed with the following three structural factors in mind:

1. Strength of harmonic implication.
2. Resolution of tendency tones.
3. Issues of time-decay and interference in working memory.

In addition, certain constraints were imposed on the permutations so that the sequences could be used without modification in a follow-up neuroimaging (MEG) experiment. The constraints were to ensure that at least one note event be fixed in time and pitch. The constraints are described below for a C major/G major sequence, i.e. [C, D, E, F, F#, G, A, B]:

1. All sequences end on the same pitch, the tonic of Key Type II (G).

2. All sequences have a penultimate note that is an interval of either a second (A) or a third (B) above the final note.

The rationale for ending them all on the pitch G was that G could function either as an implied dominant to C major or tonic to G major, thus serving as important pitches for both Key Types.

All sequences consisted of monophonic, isochronous tones rendered in a MIDI grand piano timbre with QuickTime. The interonset interval between note events was 600 ms (100 beats per minute). The 31 sequences were randomly transposed to all 12 chromatic pitch class levels.

2.2 Participants and Task

The participants were six experts with professional-level training in music. One participant was a doctoral candidate in music at New York University, and the other five were faculty members who taught music courses at NYU, three of whom specialized in music theory.

Key label	Confidence value	Numerical equivalent
Type I	4	-4
Type I	3	-3
Type I	2	-2
Type I	1	-1
Ambiguous	Any	0
Type II	1	1
Type II	2	2
Type II	3	3
Type II	4	4

Table 1: Numerical values based on key type indicated and confidence value assigned to subjects' responses to a melodic sequence.

The subjects accessed the study through a website that presented the 31 melodic sequences in random order. All sequences were provided in audio format (MIDI files converted to MP3 format) accompanied by a visual representation in staff notation. Participants were instructed to take as much time as needed and listen to the sequences as many times as they liked. They were then asked to specify the precise letter name of the tonic of the perceived key for each melody; if they felt that the sequence was not in any particular key, they were instructed to label it "ambiguous." In addition, they were asked to rate the confidence of their response on a scale from 1 to 4 (1 = very unsure, 4 = very confident).

2.3 Results

Results were tallied by assigning numerical values to each response based on key labels and confidence values. All of the subjects chose key labels that corresponded to one of the two intended Key Types. The numerical values for each response are shown in Table 1.

Type design	Sequence number	Ending type	Mean	Standard deviation
Type I	16	A-G	-3.00	0.00
Type I	20	B-G	-3.00	0.71
Type I	3	B-G	-2.80	0.45
Type I	7	A-G	-2.60	1.52
Type I	27	B-G	-2.60	0.89
Type I	11	A-G	-2.20	1.30
Type I	30	B-G	-2.00	2.00
Type I	12	A-G	-1.60	1.34
Ambiguous	31	B-G	-1.60	2.51
Type I	22	A-G	-1.60	2.88
Type I	23	B-G	-1.20	1.30
Type I	4	A-G	-1.00	3.74
Ambiguous	26	A-G	-0.80	1.92
Ambiguous	18	A-G	-0.60	1.95
Ambiguous	13	A-G	-0.20	1.64
Type II	15	B-G	-0.20	2.39
Ambiguous	6	B-G	0.20	1.48
Ambiguous	8	B-G	0.20	1.48
Ambiguous	10	B-G	0.40	2.07
Ambiguous	21	A-G	0.40	1.52
Ambiguous	2	B-G	0.60	2.51
Type II	25	A-G	0.60	3.85
Type II	14	A-G	0.80	3.49
Ambiguous	24	A-G	1.00	1.22
Type II	28	A-G	1.00	2.83
Ambiguous	29	A-G	1.20	2.17
Type II	5	B-G	2.00	2.83
Type II	9	B-G	2.00	2.92
Type II	17	A-G	2.60	3.13
Type II	19	B-G	3.00	0.71
Type II	1	B-G	4.00	0.00

Table 2: Results for Experiment 1. The "Ending type" assumes a C major/G major transposition.

After the responses were quantified, they were averaged across subjects for each of the 31 melodies. Table 2 displays the sorted results for all of the sequences. Figure 1 shows the top five sequences in each of the Key Type categories; Figure 2 shows Sequence #13, one of the melodies rated the most ambiguous. A high standard deviation indicates a higher level of disagreement among the subjects. In the case of sequences with average scores close to zero, high standard deviations indicated that the melody in question was not necessarily identified as ambiguous in key but rather leaned toward different keys, depending on the listener's perspective—i.e. the musical analogue of a Necker cube. Low standard deviations, on the other hand, indicated agreement on the lack of a clear key.

In summary, the results suggest that the structural factors used to compose the melodies were successful in influencing key perception even when the statistical/distributional information about overall key profile did not favor any particular key. Furthermore, unlike the case for Matsunaga and Abe's study (2009), the pitch of the final tone did not seem to affect the results. The ending type—whether a descending major second or major third from the penultimate to the final note—did not seem to make a difference.

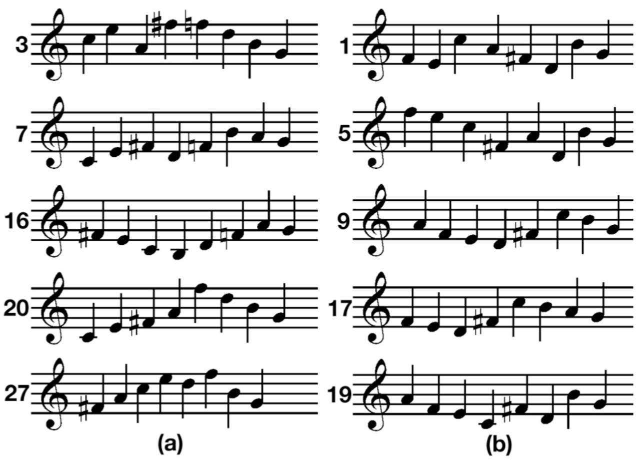

Figure 1: The top five sequences for (a) Key Type I, intended to sound like C major, and (b) Key Type II, intended to sound like G major (assuming transposition pitch set [C, D, E, F, F#, G, A, B]) resulting from Experiment 1. The numbers next to each staff refer to the sequence numbers in Table 2.

Figure 2: One of the ambiguous melodic sequences from Experiment 1 (Sequence #13).

3. EXPERIMENT 2

In Experiment 2, we examined the temporal psychophysics of structurally-based key induction. The five sequences in each Key Type category with the highest agreement ratings from Experiment 1 were used in a study that took into account tempo as well as pitch. Although Experiment 2 was designed for musically trained subjects, we did not want listeners to look at melodies in staff notation or identify a precise key—we simply asked them to identify whether a sequence sounded like a Key Type I or Key Type II melody.

3.1 Materials

Seventeen versions of each of the 10 sequences from Experiment 1 were generated at 17 different tempi ranging from 7 beats per minute to 3400 BPM (7, 15, 30, 45, 60, 75, 95, 120, 200, 400, 600, 800, 1000, 1200, 1600, 2200, and 3400 BPM). As in Experiment 1, stimuli were rendered with a MIDI piano timbre. Although the data very quickly indicated that 2200 BPM was beyond the threshold for which subjects could successfully resolve the notes in the sequence, the fastest tempo, 3400 BPM, was only added after the fifth subject to ensure that the limit was surpassed (thus data for 3400 BPM is available for 17 out of the 22 subjects).

3.2 Participants

The participants were 22 New York University students (mean age 23.8 years; 14 male, 8 female) who were skilled at instrumental performance and had an average of 15.5 years of musical training (SD = 6.4) and had taken at least one music theory course. All were familiar with the terms "dominant" and "tonic" harmony. Four subjects reported having absolute or partial absolute pitch. There were two additional subjects whose data were not included in the analysis because they had difficulty understanding the task, presumably due to lack of sufficient musical training. These subjects self-rated a 2 or lower on an overall musical proficiency scale of 1 (lowest) to 5 (highest).

3.3 Task

Each subject listened to 170 sequences (i.e. 10 sequences played at 17 different tempi) in a pseudorandomized order that took into account tempo, key, and original sequence. Thus the following constraints ensured that the sheer amount of repetition would be less evident:

- No stimulus was preceded by another stimulus generated from the same original sequence or having the same tempo.
- No stimulus was in the same key as the *two* preceding stimuli.
- All stimuli were transposed such that they were at least three sharps/flats (i.e. steps on the circle of fifths) away from the key of the immediately preceding stimulus.

Subjects were asked to indicate whether each sequence sounded resolved (ending on an implied tonic) or unresolved (ending on an implied dominant). They were instructed to ignore aspects such as perceived rhythmic or metric stability when making their decision. The experiment took approximately 40 minutes to complete.

3.4 Results

The psychophysical data show a preferred range of tempi in which participants provide robust key judgments (30 BPM - 400 BPM). Above that modulation frequency, listeners were not able to reliably use structural information to guide key determination, suggesting that structurally-based key-finding requires temporal integration and analysis that is three to four times slower than necessary for the detection of pitch, but still fast enough to be accurately computed by trained listeners in no more than 150 ms per note (1200 ms for an entire sequence of eight notes).

Figure 3 shows the mean percent correct responses for each tempo across all sequences and all subjects; Figure 4 shows the mean of d' values across all subjects for each tempo, calculated using a maximum hit rate of .9 and a minimum false alarm rate of 0.1 for a perfect set of responses. Note that judgment consistency decreases for tempi below 30 BPM (0.5 Hz) and above 400 BPM (10 Hz), with a fairly steep decline occurring above 400 BPM. There appears to be no signal at 2200 BPM (36.7Hz).

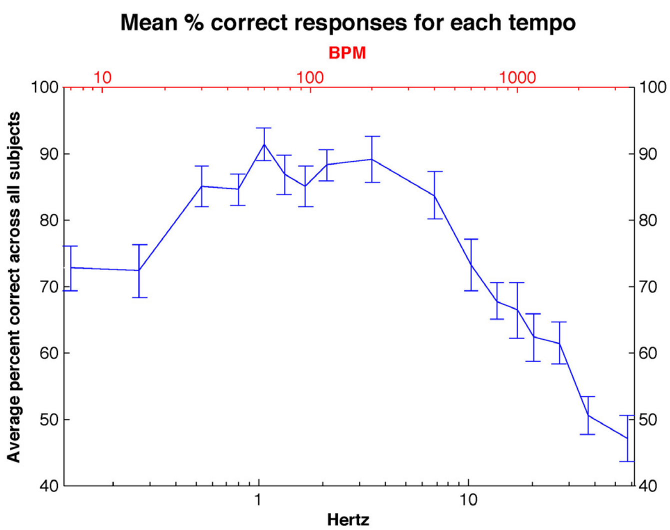

Figure 3: Average percent correct responses for each tempo. The error bars indicate estimated standard error.

4. DISCUSSION

The results of Experiment 1 and 2 show that structural features are sufficient to determine the key of melodies even when the statistical information inherent in key profiles is ambiguous and the stimulus durations extremely brief. Listeners can make robust judgments at both very slow and very fast rates ranging from 30 BPM (16s for an entire sequence of eight notes) to 400 BPM (1.2s for an entire sequence). Even at rates above and below these thresholds, many subjects were highly consistent in their judgments, and these judgments conformed to the intended key for each sequence. Furthermore, the pitch of the final note in the sequence made no discernable difference in the results. This does not necessarily mean that the pitch of the last tone is unimportant, but rather that it has less influence in the context of strong structural cues. When such cues are absent, as in the case of Matsunaga and Abe's study, the final tone might have more direct influence.

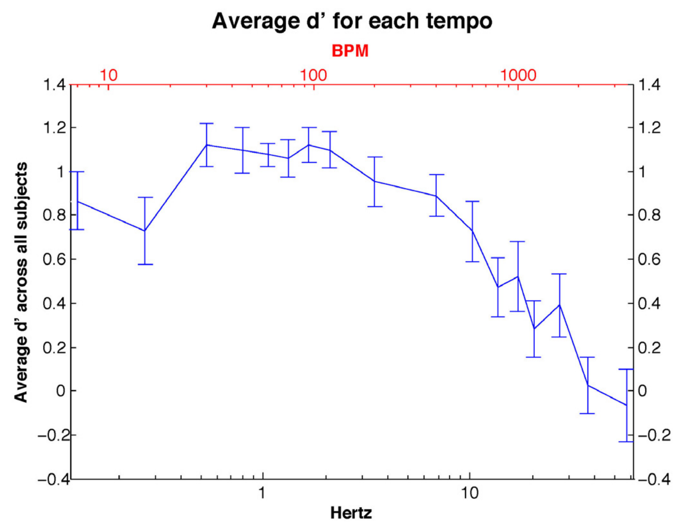

Figure 4: Average of d' values for each tempo. The error bars indicate estimated standard error.

What can these data tell us? First of all, it is surprising how quickly listeners can—robustly—determine a key for a sequence: 1.2s is a relatively short period of time in a musical context. This can only be the case if the ordering of the pitches plays a crucial role in determining key. One might argue that local structural features might be more important at a slow tempo since it would be more difficult to apply a statistical approach to determining key given the sparseness of pitch input. However, it appears that this is not the case.

The next step is to explore the structural features that were used to construct the stimuli in this study in more detail. In particular, "unusual" intervals such as tritones, as well as contour features that can contribute to the salience of chordal patterns, can influence the strength of harmonic implications to great degree. While time decay seems to be a minimal issue in sequences of very short length, the inhibition of certain pitches due to the immediate succession of other pitches close in frequency (for example, and F-natural followed by an F#) can also play a role in pitch memory, even on a short timescale. Pitch memory studies by Deutsch (1970, 1972) have indicated that there is indeed an inhibitory influence that is more significant when a tone is immediately followed by a tone close in pitch.

The sequences for which there was the greatest listener agreement in key judgment all had clearly delineated dominant seventh chords near the end of the sequence, not merely a well-placed tritone. In the case of two closely related keys (with tonics a fifth apart), specifically C major and G major, the F-natural, is an important indicator of C major—where is it positioned, and in what context, can be crucial both in terms of inhibitory pitch mechanisms and harmonic cues. Likewise, the F# is the leading tone for G major and its implied function is crucial in determining whether a sequence sounds like it's in G major or C major. If the F# is merely perceived as a chromatic lower neighbor tone to a G that has been clearly establish as the dominant in C major, its effect in implying a G major key area is considerably diminished.

If this study is any indication, the explanatory power of a purely statistical model may be fairly weak compared to one that also incorporates structural factors. Indeed, the present results even raise a more extreme possibility, which is that the success of statistical models could be an epiphenomenon, in which the relevant statistics actually derive from psychologically-represented structural factors. Future work will be needed to decide between this more radical possibility and the possibility that *both* statistical *and* structural factors make important, and in some cases independent, contributions to the identification of musical key.

5. REFERENCES

Brown, H. (1988). The interplay of set content and temporal context in a functional theory of tonality perception. *Music Perception, 11*, 371-407.

Brown, H., Butler, D., & Jones, M. R. (1994). Musical and temporal influences on key discovery. *Music Perception, 11,* 371-407.

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.

Deutsch, D. (1970). Dislocation of tones in a musical sequence: A memory illusion. *Nature, 226,* 286.

Deutsch, D. (1972). Effect of repetition of standard and comparison tones on recognition memory for pitch. *Journal of Experimental Psychology, 93,* 156-162.

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.

Longuet-Higgins, H. C., & Steedman, M. J. (1971). On interpreting Bach. *Machine Intelligence, 6*, 221-241.

Matsunaga, A., & Abe, J. (2005). Cues for key perception of a melody: Pitch set alone? *Music Perception, 23*, 153-164.

Matsunaga, A., & Abe, J. (2009). Do local properties function as cues for musical key perception? *Japanese Psychological Research, 51*, 85-95.

Temperley, D. (2007). *Music and Probability*. Cambridge, MA: MIT Press.

Temperley, D. & Marvin, E. W. (2008). Pitch-class Distribution and the Identification of Key. *Music Perception, 25*, 193-212.

Vos, P. G. (1999). Key implications of ascending fourth and descending fifth openings. *Psychology of Music, 27*, 4-18.

Vos, P. G., & Van Geenen, E. W. (1996). A parallel processing key-finding model. *Music Perception, 14,* 185-224.

Yoshino, I., & Abe, J. (2004). Cognitive modeling of key interpretation in melody perception. *Japanese Psychological Research, 46*, 283-297.