Title
Incremental Process of Musical Key Identification

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 29(29)

ISSN
1069-7977

Authors
Matsunaga, Rie
Abe, Jun-ichi

Publication Date
2007

Peer reviewed

Incremental Process of Musical Key Identification

Rie Matsunaga (matunaga@psych.let.hokudai.ac.jp)
Japan Society for the Promotion of Science

Jun-ichi Abe (abe@psych.let.hokudai.ac.jp)
Department of Psychology, Hokkaido University, N10 W7, Kita-ku
Sapporo, 060-0810, JAPAN

Abstract

Pitch set is a primary cue for key perception. However, pitch set alone cannot account for the phenomenon that listeners perceive different keys for melodies that consist of the same pitch set but exhibit different temporal arrangements. Contrary to previous results, a recent study demonstrated that additional cues based on sequence properties (e.g., the augmented fourth, a pitch class in the final position of a melody, etc.) did not contribute to key perception. To explain this phenomenon, we traced how listeners developed a sense of key as melodies (with the same pitch set differing in pitch sequence) unfolded over time. In each melody, listeners identified the key following the presentation of a segment of pitches where the number of tones within a segment increased with successive presentations. Results suggested that listeners gradually established the sense of key. Throughout the progress of melodies, the listeners' key responses were governed by a set of pitches within a segment provided a given point in time. These findings suggest that key identification is derived from the incremental changes of the pitch set with unfolding of a melody.

Keywords: Music perception; Tonal organization; Musical key identification

Introduction

Musical key identification plays a fundamental role in music perception. Empirical evidence suggests that key identification results from the organization of a tone sequence into a hierarchical system of tonality according to a listener's internal schema (e.g., Abe & Hoshino, 1990; Krumhansl, 1990). What kind of a property in an arbitrary melody functions as a cue for key identification?

There are three approaches to this issue. The first approach stresses "pitch set (the collection of pitch classes in a melody, regardless of their order)", which is a global property in a melody. Many studies have shown that pitch set indeed functions as a primary cue for key identification (e.g., Abe & Hoshino, 1990; Balzano, 1982; Longuet-Higgins, 1987). More specifically, these studies suggest that listeners, who are familiar with Western music, perceive a melody to be in a given key when the constituent pitches of this melody are all interpretable as scale tones of a certain diatonic scale.

A second approach emphasizes "the distribution of pitch classes" which is also a global property in a melody (e.g., Krumhansl, 1990; Oram & Cuddy, 1995; Temperley, 2004). The underlying idea in this approach posits that listeners interpret the most commonly occurring pitch class within a melody as its tonal center. However, evidence favoring this approach is often correlational, meaning that inferences about the causal relationship between the frequency of a pitch class and the perception of it as a tonal center remain tentative. In other words, pitch distributional properties (e.g., relative frequency) may not provide cues for key identification.

The third approach stresses that "a local property", not a global property, in a melody. For example, Butler and Brown proposed rare intervals (i.e., the augmented fourth or the diminished fifth) as cues (e.g., Butler & Brown, 1994). Others have suggested that different local property cues (other than rare intervals) function in this capacity; thus, proposed cues include the inclusion of a perfect fifth and either the major third or the minor third (e.g., Huovinen, 2002), an ascending fourth or a descending fifth in the opening (e.g., Vos, 1999), the pitch class of the final tone (e.g., Creel & Newport, 2002), or the opening pitch class in conjunction with the final pitch class (e.g., Cuddy, Cohen & Mewhort, 1981). Finally, however, on this question there appears to be little current consensus, across these various studies.

It is clear that, in spite of research on these three proposals, the particular properties that contribute to reliable key identification remain open to question. Consider the two melodies shown in Figure 1. Although both are composed of the same set of six pitches, they differ in the sequential arrangement of these pitches. If pitch set alone serves as a cue to key, then listeners should identify these two melodies to have the same key. But they do not. In fact, listeners generally interpret Melody 1 as being in C major and Melody 2 as being in G major.

Matsunaga and Abe (2005) explored possible determinants of this phenomenon. They required musically trained listeners and untrained listeners to identify a key (or

Figure 1: Melodies consisting of the same pitch set [C D, E, G, A, B] differing in pitch sequence.

a tonal center) for melodies consisting of the same pitch set but differing in temporal arrangements of pitches. They found that, regardless of their musical training, listeners' key identifications were governed not only by the pitch set but also by certain other melodic properties.

In pursuing the latter, Matsunaga and Abe (2007) investigated whether certain local sequence properties might be responsible for different key identifications. The experimental design of Matsunaga and Abe differed from those of previous studies in that they examined what local property have more influence than any other local property on key identification. Using linear discriminant analyses, they evaluated relative contributions of as many different local sequence properties as possible (e.g., the augmented fourth, the perfect fifth, a pitch class in the final position, etc.). Their listeners were asked to provide key identifications for a variety of different intact melodies that comprised the same pitch set but differed in temporal arrangements of pitches. The results led to the conclusion that none of the local sequence properties examined contributed significantly to key identification. Such findings appear to undermine hypotheses about key identification that rest on specific local sequence properties, although they do not rule out the possibility that as yet unexamined local properties function as critical cues to key. In sum, the research of Matsunaga and Abe (2005, 2007) has shown that the previous approaches to key identification fall short in explaining how listeners arrive at different keys of melodies

The purpose of the present study is to examine certain other melodic properties that might lead listeners to perceive different keys for melodies that comprise the same pitch set but differ in temporal arrangements of pitches. These melodies share a set of all constituent pitches in a whole melody; that is, *pitch full-set* is common to the melodies. On the other hand, the melodies do not share a set of pitches within each of the melody segments which start at the first pitch and increase by one pitch. That is, *pitch subset* is uncommon to the melodies. For example, Melodies 1 and 2 (Figure 1) share the pitch full-set [C, D, E, G, A, B], while they have different pitch subsets within a segment that consists of the first two pitches (i.e., [C, G] in Melody 1 and [D, B] in Melody 2). Likewise, they differ in pitch subsets within a segment that consists of first three pitches (i.e., [C, E, G] in Melody 1 and [C, D, B] in Melody 2), first four pitches (i.e., [C, E, G, A] and [C, D, G, B]), or first five pitches (i.e., [C, D, E, G, A] and [C, D, E, G, B]). Such differences in pitch subsets among melodies may, in turn, contribute to corresponding differences in key identification. Accordingly, our hypothesis is that unfolding pitch subsets (i.e., as presented at successive points in time) will govern the time course of key identification

To assess this hypothesis, we enlisted the experimental tracking procedure outlined in Figure 2. Here, although each stimulus tone sequences consisted of six pitches, initially listeners heard the first two tones of each sequence (Stage 2) and then identified keys. Following this, the listeners heard

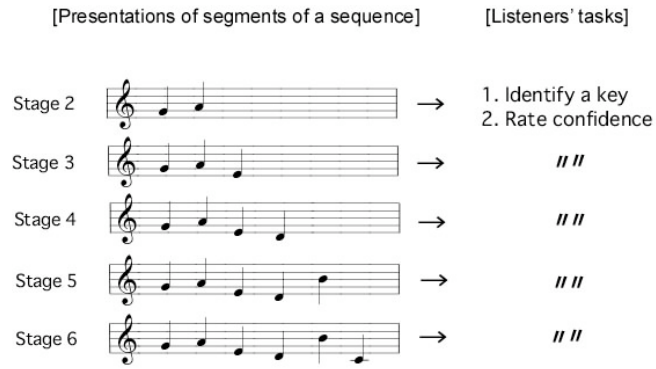

Figure 2: The experimental procedure of the present study.

the first three tones of each sequence (State 3) and identified keys. This procedure was continued until key response distributions were obtained for sequences of Stages 2 through 6.

Our task required listeners to directly name the key of a presented sequence. Although this is simplest and most direct means of measuring key identification, it does require that listeners are capable of key naming. For this reason we employed musically trained listeners (referred to as musicians) with absolute pitch as participants; musically untrained listeners (i.e., non-musicians) cannot perform this task. In addition, musicians are typically less influenced by unexpected factors and bias. Note, however, that these differences between musicians and non-musicians does not imply that only musicians can respond to specialized perceptual cues. Many previous studies indicate that musicians and non-musicians make similar distinctions between melodic and jumbled sequences (e.g., Hoshino & Abe, 1981), are similarly sensitive to tonal hierarchies (e.g., Hèbert, Peretz, & Gagnon, 1995), and basically identify the same tonal center (e.g., Matsunaga & Abe, 2005). Such findings suggest that perceptual cues for key identification are common to all listeners that share a common cultural exposure to music, regardless of their musical training and absolute pitch abilities.

Method

Participants

The participants were 15 undergraduate students (range = 18-22 years), who were familiar with Western music. All reported that they possessed absolute pitch. They had an average of 15.3 (range = 12-17) years of musical training. Musical instruments that they routinely played were the piano for 10 participants, and the electronic organ for five participants.

Materials and apparatus

Thirty-nine sequences of six tones were used as musical stimuli. All comprised the same pitch full-set, but they differed in the temporal arrangement of the constituent pitches. The 39 tone sequences were chosen from the stimulus sequences used in the Matsunaga and Abe (2007). The pitch full-set employed was [C, D, E, G, A, B]. All constituent tones of this pitch full-set can be interpreted as scale tones of the following four keys: C major, G major, E minor, and A minor. To create as many kinds of intervals as possible within the pitch full-set, we generated two pitch full-sets: [C_4, D_4, E_4, G_4, A_4, B_4] (Pitch Full-Set I) and [D_4, E_4, G_4, A_4, B_4, C_5] (Pitch Full-Set II). There were 20 possible intervals between two pitches in the two pitch full-sets: (±1)[1], (±2), (±3), (±4), (±5), (±7), (±8), (±9), (±10), and (±11) in semitone. Of the 39 tone sequences, 13 originated from Pitch Full-Set I and 26 came from II.

All the tone sequences were monophonic isochronous melodies whose tones were contiguous and did not overlap. All sequences were presented at the same tempo; the duration of each tone was equal (i.e., 0.6 s), for a total of 3.6 s per tone sequence. The timbre of each pitch was that of an acoustic grand piano. The tone sequences were created as MIDI files using sequencing software (Roland "Cakewalk" software) installed on a Windows PC.

Procedure

Each listener (i.e., participant) was seated in front of two speakers. Each listener was given a response sheet listing 12 major and 12 minor key categories plus an atonal category. Each sequence was played in the following fashion: First, only the opening two pitches of the sequence (Stage 2) were presented; next, the first three pitches (Stage 3) were presented, and so on until the presentation ended with the whole tone sequence (Stage 6). After each presentation, listeners were asked to identify the most plausible key and to rate their subjective confidence in their key identification on a 7-point scale (7 = *full confidence* to 1 = *poor confidence*). The presentations were self-paced; participants indicated their readiness for the next presentation by conveying verbally. After three practice trials, the 39 experimental trials were presented in randomized order across the listeners.

Results and Discussion

Transitions of key responses and confidence ratings

Distributions of key responses for the two pitch full-sets were highly similar. Therefore, the data were pooled across the pitch full-sets.

To examine how the listeners developed the sense of a key while tone sequences unfolded over time, we analyzed the data for key identification responses and confidence ratings separately. The first analysis examined percentages of key switch responses, namely those responses that reflected a change in key from one presentation (e.g., Stage n) to the next (e.g., Stage n+1). After Stage 3-4, the average percentage of key switch responses decreased as tone sequences progressed (Stages 2-3 = 48%, Stages 3-4 = 51%, Stages 4-5 = 44%, and Stage 5-6 = 43%). A one-way ANOVA revealed a significant main effect of the stages, $F(3, 42) = 2.97$, $p = .043$. No comparisons between the stages were significant (all ps > .10). Next, we examined confidence ratings. The confidence ratings provided by each listener were averaged for each of the five stages (Stages 2-6). A one-way ANOVA revealed that confidence increased significantly over stage (M = 3.7, 3.8, 3.9, 4.0, 4.6 for Stages 2-6, respectively) with $F(4, 56) = 10.72$, $p < .001$. However, the only significant pair-wise differences were between Stage 6 and each of the remaining four stages (all ps < .01), suggesting that the main effect was responsible for confidence ratings in Stage 6. It may not be all that surprising that the confidence of key identification was the highest at the end of each sequence, because the listeners knew that it was the end of each sequence.

These results showed that the listeners vacillated among a few key categories in their key identifications but that these vacillations attenuated and confidence increased as a sequence progressed. Consistent with the results of previous studies (Krumhansl & Kessler, 1982), our findings suggest that listeners' key identifications are unlikely to be stabilized abruptly in a particular location; but rather listeners gradually establish the sense of key.

Contributions of pitch subsets to key responses

Results of key identification responses appear in Table 1. In each of the stages, from Stage 2 through 5, participants tended to limit their key responses to C major, G major, and A minor. Responses to remaining keys were relatively infrequent. These three keys contain all constituent tones of all the pitch subsets as diatonic scale tones. Some of the tone sequences in each stage elicited response agreement among the majority of the listeners for C major, or G major, or A minor. This suggested that the listeners systematically selected a key from these three keys.

In order to determine whether pitch subsets led the listeners to distinguish among C major, G major, and A minor responses, we performed Multiple Discriminant Analyses (MDA) with dummy variables separately for each of the four stages. If MDA shows that pitch subsets satisfied the following two criteria, then we may infer that such

Table 1

Percent of each of the three key responses and the number of tone sequences that elicited response agreement among majority of the participants (eight or more of the15 participants) in each of the three keys.

	Percent of key responses (%)			Number of tone sequences		
	C major	G major	A minor	C major	G major	A minor
Stage 2	39.7	24.6	12.5	12	5	3
Stage 3	35.9	25.6	14.2	9	5	3
Stage 4	31.5	33.0	16.2	5	7	2
Stage 5	32.1	36.6	15.7	6	6	1
Stage 6	40.5	38.1	11.6	15	11	0

[1] In this paper, intervals were denoted by positive integers for ascending intervals and by negative integers for descending intervals (one unit = a semitone). For example, the ascending major third and the descending major third were denoted as (+4) and (-4) respectively.

subsets made significant contributions to key responses: (1) a pitch subset was associated with one key response (e.g., C major) but it was not associated with another key response (e.g., G major); and (2) pitch subsets associated with each of the key responses were transpositionally equivalent -- for different absolute pitches with the same tonal functions (e.g., tonic, dominant).

It is necessary to distinguish between the major and minor modes because the two modes differ in the sequence of intervals (in semitones) between adjacent tones. Due to this difference, tonal functions of intervals in major keys are not always equivalent to those in minor keys even though intervals of major keys and minor keys are the same. Here, we performed MDAs using the C major, G major, and A minor groups as dependent variable groups, while we decided to focus mainly on listeners' ability to differentiate the two major keys (C and G majors).

Figures 3-6 present the results of MDAs for key identification responses in Stage 2-5 respectively. We begin with the results of MDA for Stage 2, which involve first-pitch subsets comprising two tones (Figure 3). The dependent variable groups of the MDA was three-group key identification (C major, G major, and A minor groups), while independent variables were 14 types of the two-pitch subsets (e.g., [C, D], [C, E], etc). The sample observations were 449 responses, which consisted 232 of responses of C major, 144 of G major, 73 of A minor. On the basis of locations of group centroids in a space defined by the

A

	Discriminant function	
	1	2
C major group	-0.30	0.58
G major group	-0.71	-0.80
A minor group	2.34	-0.25

B

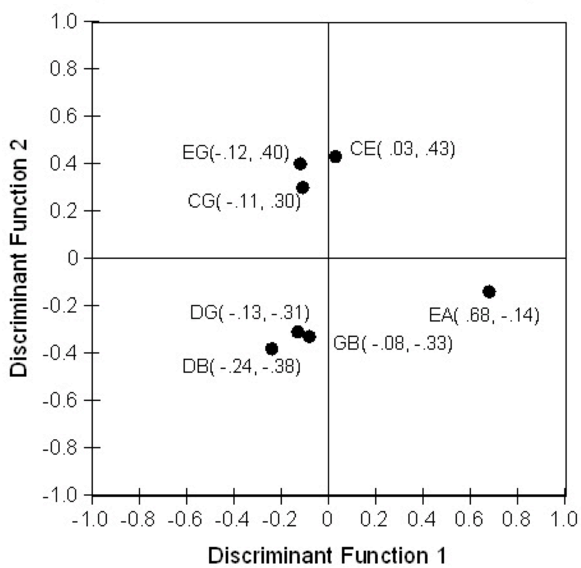

Figure 3: Results of MDA on Stage 2. 3A represents group centroids for C major, G major, and A minor groups. 3B represents pitch subsets in a space defined by the two structure coefficients. Pitch subsets having ±.20 or higher on either discriminant function 1 or function 2 are represented in the 3B.

significant discriminant functions, we assigned the label "C major-like" and "G major-like" to the positive and negative directions of Function 2 respectively, and "A minor-like" to the positive direction of Function 1 (Figure 3A). If the structure coefficient near ±0.30 or higher of independent variable has the same sign in the discriminant function as the group centroid of a group, this indicates that the independent variable contributes positively to defining this group (cf. Hair, Anderson, Tatham, & Black, 1998). Visual inspection of Figure 3B shows that [C, E], [E, G], and [C, G] are associated with "C major-like", while [D, B], [G, B], and [D, G] are associated with the "G major–like".

In the analysis presented in Figure 3, the pitch subsets of [C, E], [E, G], [C, G] may be interpreted as [tonic, mediant], [mediant, dominant], and [tonic, dominant] in C major, respectively. On the other hand, [D, B], [G, B], and [D, G] are interpretable as [dominant, mediant], [tonic, mediant], and [dominant, tonic] in G major, respectively. These relationships between pitch subsets of C major and those of G major indicated that the three separate pairs of the pitch subsets for each key differed in pitch classes but they nonetheless shared comparable tonal functions. The [C, E] and [G, B] include (±4) or (±8); [E, G] and [D, B] include (±3) or (±9); [C, G] and [D, G] include (±7) or (±5). The results suggest that pitch subsets with (±4) or (±8), those with (±3) or (±9), and those with (±7) or (±5) made significant contributions to distinction of the major keys. Finally, this MDA also reveals listeners' sensitivity to key information associated with A minor -- A minor was associated with [E, A], which is interpretable as [dominant, tonic] in A minor.

The results for MDA in Stage 3, which involve pitch subsets that include the first three tones, appear in Figure 4. Dependent variables were a three-group key identification, while independent variables were 17 types of the three-pitch subsets (e.g., [C, D, E], [C, E, G], etc.). The sample observations were 442 responses. The correspondences of [C, E, G] and [D, E, G] with positive direction of Function 1 reflect their associations with "C major-like", while the correspondence of [D, G, B] and [G, A, B] with the negative direction of Function 2 reflect their associations with " G major–like." The pitch sets of [C, E, G] and [D, E, G] are interpretable as [tonic, mediant, dominant] and [supertonic, mediant, dominant] in C major, respectively. [D, G, B] and [G, A, B] are interpretable as [dominant, tonic, mediant] and [tonic, supertonic, mediant] in G major, respectively. The [C, E, G] of C major and [D, G, B] of G major differed in pitch classes but shared tonal functions. The [C, E, G] and [D, G, B] subsets include intervals [(±4), (±3), (±7)], [(±8), (±3), (±5)], or [(±9), (±4), (±5)], because there were two kinds of C (i.e., C_4 and C_5) in the used tone sequence. This result of the major keys was consistent with that of A minor -- A minor was associated with [C, E, A], which is interpretable as [mediant, dominant, tonic] in A minor.

The results for MDA in Stage 4, which involves the first pitch subset containing four tones, is shown in Figure 5. Independent variables were 12 types of the four-pitch

A

	Discriminant function	
	1	2
C major group	0.44	0.31
G major group	-0.02	-0.63
A minor group	-1.07	0.35

B

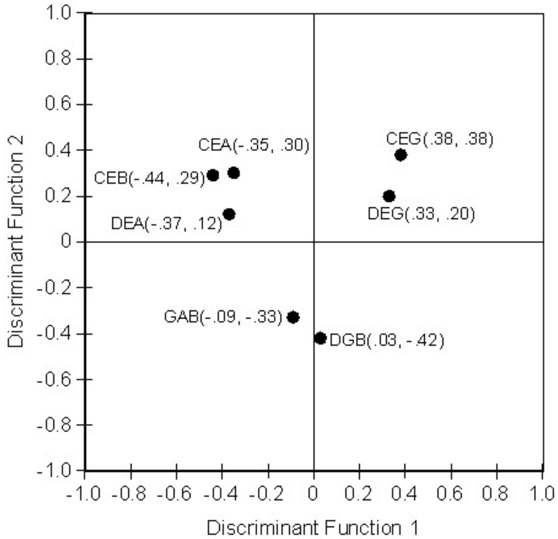

Figure 4: Results of MDA on Stage 3. The representations of the table and the figure are the same as those in Figure 3.

subsets (e.g., [C, D, E, G] etc.). The sample observations were 472 responses. The correspondences of [C, E, G, B] and [C, D, E, G] with positive direction of Function 1 reflect their associations with "C major-like", while the correspondence of [D, G, A, B] with the negative direction of Function 2 reflects its association with " G major–like." The [C, E, G, B] and [C, D, E, G] subsets are interpretable as [tonic, mediant, dominant, leading-tone] and [tonic, supertonic, mediant, dominant] in C major respectively. The [D, G, A, B] subset is interpretable as [dominant, tonic, supertonic, mediant] in G major. Thus, [C, D, E, G] subset of C major and [D, G, A, B] of G major differed in pitch classes but shared tonal functions, and these pitch subsets include a set of intervals [(±2), (±3), (±4), (±5), (±7)], [(±2), (±3), (±5), (±8), (±10)], or [(±2), (±4), (±5), (±7), (±9)]. Again, this analysis reveals listeners' sensitivity to A minor -- A minor was associated with [C, E, A, B], which is interpretable as [mediant, dominant, tonic, supertonic] in A minor.

Finally, MDA in Stage 5, which involves the first-five-pitch subsets, showed that none of the pitch subsets made significant contributions to key identification. The result might reflect the fact that combinations of five pitches (of six pitches) were highly similar.

In summary, the results of MDAs revealed that the pitch subsets that figured in Stages 2-4 made reliable contributions to key identification responses. Across these three stages, the significant pitch subsets shared (±7) and its inversion (±5), (±4) and its inversion (±8), and (±3) and its inversion (±9). In other words, these pitch intervals

A

	Discriminant function	
	1	2
C major group	0.43	-0.27
G major group	-0.04	0.42
A minor group	-0.74	-0.35

B

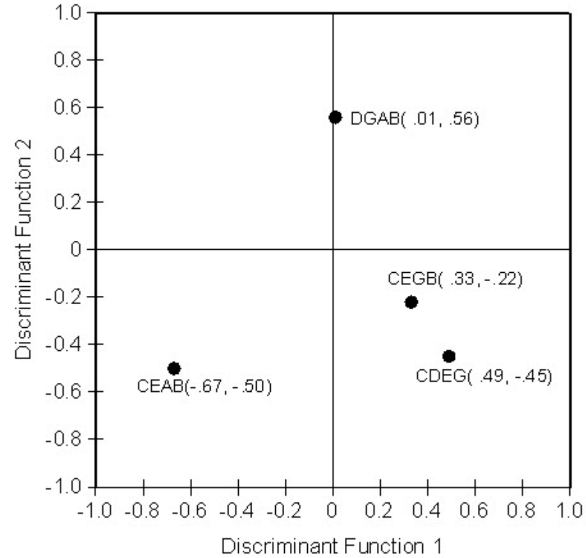

Figure 5: Results of MDA on Stage 4. The representations of the table and the figure are the same as those in Figure 3.

A

	Discriminant function	
	1	2
C major group	-0.36	-0.13
G major group	0.01	0.18
A minor group	0.73	-0.14

B

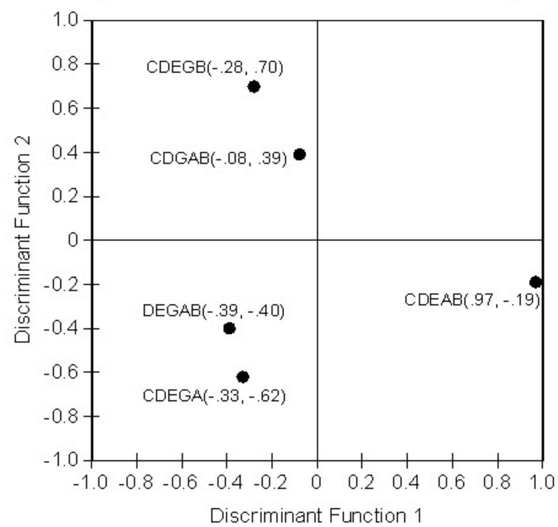

Figure 6: Results of MDA on Stage 5. The representations of the table and the figure are the same as those in Figure 3.

correspond to intervals that consist of the "tonic triad" of a diatonic scale. Thus, in each stage, the choice of a key response is likely to be based on those pitch subsets that include constituent intervals of the tonic triad.

General Discussion

The present experiment indicates that the listener gradually established the sense of key as melodies unfolded. The MDAs, across the first successive few stages, revealed that only a small number of pitch subsets provided in each stage made significant contributions to the key identification responses. Specifically, the results suggested that in a given stage, key identifications were determined from pitch subsets that offered constituent intervals of a tonic triad of a diatonic scale. This is consistent with other evidence that the tonic triad of a diatonic scale facilitates key identification for listeners familiarized with Western music (e.g., Abe, 1987; Krumhansl & Kessler, 1982). The present results suggest that, throughout progress of a tone sequence, the sense of key is governed by pitch subsets provided at each point in time. In other words, key identification emerged gradually on a pitch subset by pitch subset basis.

As mentioned in Introduction, our previous studies (Matsunaga & Abe, 2005, 2007) showed that neither the original pitch set approach (e.g., Longuet-Higgins, 1987) nor the local property approach (e.g., Butler & Brown, 1994; Huovinen, 2002; Vos, 1999) explain the phenomenon that listeners identified different keys for melodies that comprised the same pitch set but differing in the temporal arrangement of pitches. By tracing listeners' development of a sense of key, the present study found that identification of different keys for the melodies is due to pitch subsets' contributions that are dynamically accumulated throughout the course of musical passages. This implies that tonal organization is based on incremental changes of pitch set as a melody unfolds over time.

References

Abe, J. (1987). How is a melody processed? In G. Hatano (Ed.), *Music and cognition* (pp. 41-68). Tokyo: Tokyo University Press. (In Japanese.)

Abe, J., & Hoshino, E. (1990). Schema driven properties in melody cognition: Experiments on final tone extrapolation by music experts. *Psychomusicology, 9*, 161-172.

Balzano, J. G. (1982). The pitch set as a level of description for studying musical pitch perception. In M. Clynes, (Ed.), *Music, Mind, and Brain*. New York: Plenum Press.

Butler, D., & Brown, H. (1994). Describing the mental representation of tonality in music. In R. Aiello (Ed.), *Musical perceptions* (pp. 191-212). New York: Oxford University Press.

Creel, S. C., & Newport, E. J. (2002). Tonal profiles of artificial scales: Implications for music learning. In C. Stevens, D. Burnham, G. McPherson, E. Schubert, & J. Renwick (Eds.), *Proceedings of the 7the International Conference on Music Perception and Cognition*, (pp. 281-284). Adelaide, Australia: Causal Productions.

Cuddy, L. L., Cohen, A. J., & Mewhort, D. J. K. (1981). Perception of structure in short melodic sequences. *Journal of Experimental Psychology: Human Perception and Performance, 7,* 869-883.

Hair, J. F., Anderson, E. R., Tatham R., & Black, W. C. (1998). *Multivariate data analysis* (5th ed.). NJ: Prentice Hall.

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

Hoshino, E., & Abe, J. (1981). *Tonality and the coherence in melody cognition*. (Hokkaido Behavioral Science Report Series P No. 23). Sapporo, Japan: Hokkaido University, Department of Psychology. (In Japanese.)

Huovinen, E. (2002). *Pitch class constellations: Studies in the perception of tonal centricity*. Turku, Finland: Suomen Musiikkitieteellinen Seura.

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

Krumhansl, C. L., & Kessler, E. J. (1982). Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys. *Psychological Review, 89*, 334-368.

Longuet-Higgins, H. C. (1987). *Mental processes: studies in cognitive science*. London: MIT Press.

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

Matsunaga, R., & Abe, J. (2007). *What kinds of local sequence properties function as cues for musical key perception?* Manuscript submitted for publication.

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

Temperley, D. (2004). *The cognition of basic musical structures*. Cambridge, MA: MIT Press.

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