KEER2010, PARIS | MARCH 2-4 2010

INTERNATIONAL CONFERENCE ON KANSEI ENGINEERING AND EMOTION RESEARCH 2010

EXTRACTING THE MINIMUM STRUCTURES OF MUSICAL SCHEMAS FROM TRADITIONAL JAPANESE AND CHINESE FOLK SONGS

Akihiro KAWASE * [a] **and Akifumi TOKOSUMI**[a]

[a] *Department of Value and Decision Science, Tokyo Institute of Technology, Japan*

ABSTRACT

In this study, we extract the pitch transition patterns from both traditional Japanese and Chinese folk songs and examine the characteristics of their respective schemas. Specifically, we sample 1,794 works from Nihon Min-yo Taikan (Anthology of Japanese Folk Songs, 1944-1993) for Japanese folk songs and 2,040 folk songs from a website providing virtual musical scores for Chinese folk songs, and probabilistically create a tree structure in modeling a variable-length Markov chain to compare minimum transition patterns occurring with high probabilities in terms of pitch intervals. A variable-length Markov chain, also known as a FSMX model or a finite-memory source, is a Markovian process having a sparse memory structure with states that closely cohere. The structure can be characterized by a parsimonious number of transition probabilities for stationary categorical time series. The results indicate that (1) the minimal structures of Japanese folk songs tend to create a longer schema than Chinese folk songs, and vertical transitions are sung within a small range; to be exact, below intervals within a perfect fourth pitch. On the other hand, the minimal structures of Chinese folk songs tend to create a shorter schema than Japanese folk songs, and their vertical transitions extend beyond the interval of a perfect fourth pitch, and (2) the formations of perfect fourth pitches and perfect octave characterize the respective musical schema.

Keywords: Music cognition, Folk songs, Schema

*Corresponding author:** Tokyo Institute of Technology, 2-12-1 Ookayama, Meguro-ku, Tokyo 152-8552, Japan. kawase@valdes.titech.ac.jp

1. INTRODUCTION

1.1. Purpose of This Study

Aspects of musical structure, such as meter, phrase structure, contrapuntal structure, pitch spelling, harmony, and key, are well known and understood by many music students, and, thus, are frequently taken for granted as musical facts. However, one question that has yet to be answered is what process underlies the inference of such structures [1]. Existing studies, such as Meyer [2] and Lerdahl and Jackendoff [3] to mention just a couple, have been mainly concerned with the art music of the common practice periods (identified as Baroque, Classical and Romantic in Western Europe). We believe that the fundamental reason for that focus is the consequence of the musical score and notation, which rests on the following two points. First, the vast size of the musical corpus that is available, having been accumulated over several hundreds of years, allows for large-scale analysis. Second, systemized composition theories, based on the musical score, have offered great promise for the extraction of sophisticated structures. Thus, in both a quantitative as well as a qualitative sense, there is no doubt that Western art music based on musical notation is convenient and suitable for computational analysis.

However, there is also an undeniable sense that many aspects of musical structure may play important roles in non-Western music as well. In particular, the characteristics of melodies in Japanese folk songs, which were created by anonymous nonprofessional musicians and orally transmitted down, represents a tonal system that has been derived from and formed by antiphonal singing within a primitive society. Therefore, there is far more chance of extracting musical patterns of primitive conditions from Japanese folk songs than from Western art music. There have been many case studies about Japanese folk songs from a musicological point of view, such as the tetrachord theory [4], but those studies were conducted with a small amount of data, and have rarely been reaffirmed by objective analyses utilizing large amounts of data for computational analysis.

The goal of this study is to extract the pitch transition patterns from both traditional Japanese and Chinese folk songs and to examine the characteristics of their respective structures. This kind of structure is generally called a schema in psychology. The reason why we have chosen Chinese folk songs as a target for comparison to Japanese folk songs is that despite the considerable influence of Chinese culture on Japan (a brief history of that influence is described in the next section), the musical characteristics of both countries differ in many aspects.

The analysis method used to achieve the present study’s goals employs a musical corpus that is almost 25 times larger than the probabilistic approach used by Kawase and Tokosumi [5].

1.2. Brief History of Japanese and Chinese Musical Scales

During the Zhou Dynasty from 1122 to 256 BC, Chinese musicians divided the octave into twelve pitches that were roughly equivalent to the twelve-tone equal temperament of modern Western music. For a long time within Western music theory, the twelve perfect fifths and seven octaves have been regarded and treated as being the same interval. By skipping a circle of perfect fifths, one can eventually reach a pitch that is approximately seven whole octaves above one’s starting pitch. The degree of discrepancy is well known as the Pythagorean comma. According to Huainanzi, Chinese mathematicians were aware of the Pythagorean comma as early as 122 BC. With such materials, Chinese musicians constructed a scale of seven pitches and provided a twelve-tone chromatic scale.

In the Heian period from 794 to 1185, various forms of music and musical instruments were introduced to Japan, together with Buddhism and Confucianism, and foreign cultures began to develop. However, after 1639 when the feudal government of the time entered an era of isolation by breaking off relations with foreign countries and prohibited foreign travel, Japan musicians revised and established their own musical schema differing from the Chinese one.

2. OVERVIEW OF THE DATA

We sampled the five largest song genres within the music corpora included in the Nihon Min-yo Taikan (Anthology of Japanese Folk Songs, 1944-1993) for Japanese folk songs, and three song genres included in The Essen Folksong Collection from KernScores [6], a website providing virtual musical scores, for Chinese folk songs. In total, there were 202,246 tones in the sample of 1,794 Japanese folk song pieces, and 124,677 tones in the sample of 2,040 Chinese folk song pieces.

While the Chinese folk songs were copied down with absolute pitch (Figure 1), the scores for the Japanese folk songs were taken down with relative pitch and transposed with either three-flat key signatures (C minor key or E-flat major key) or no sharps/flats (A minor key or C major key). These are based on two basic types of pentatonic scales called the in scale (Figure 2) and the yo scale (Figure 3).

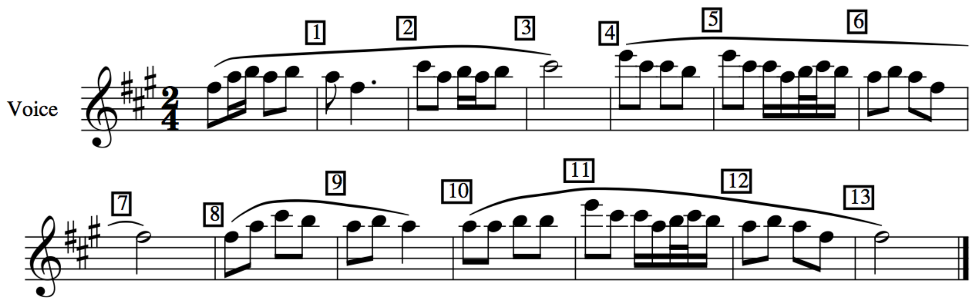

Figure 1: Example of a score of Chinese folk song (Xiu he bao)

3. PITCH REPRESENTATION

In order to digitize the song pieces, we generated a sequence of notes by adopting the interval representation to each melody, not only to handle song pieces represented in absolute pitch and relative pitch on an equal basis, but also because the interval representation is more compatible with human perception compared to other melody representation, as people employ interval infor-

Figure 2: Example of a score written in *in* scale

Figure 3: Example of a score written in *yo* scale

mation when they memorize, distinguish, or sing a melody [7]. The details of the procedures are as follows.

1. For a given Standard MIDI File (SMF), extract pitch information from note events and sort $S = (s_1, s_2, s_3, \cdots, s_k, \cdots, s_e)$ in ascending order, where element $s_k (k \in \mathbf{Z}^+)$ is the pitch value for number k, and e is the number of elements in sequence S.

2. Digitize each note in terms of its relative pitch using the MIDI Tuning Standard, where the pitch normally associated with A-5 gives the pitch 69, and then generate $\Sigma = (\sigma_1, \sigma_2, \sigma_3, \cdots, \sigma_k, \cdots, \sigma_e)$ where σ_k is the value for number k.

3. By subtracting the next value, generate a pitch sequence $X = (x_1, x_2, x_3, \cdots, x_t, \cdots, x_{e-1})$ that carries information about the pitch interval to the next note, where x_t is the interval for number t. Thus, an entire score is first converted into a sequence where elements are symbolized as tones in ascending order (e.g., E-5, G-5, A-5, G-5, $\cdots\cdots$), then each integer is semitonically assigned to the notes (e.g., 64, 67, 69, 67, $\cdots\cdots$), and then at last to intervals (e.g., +3, +2, -2, $\cdots\cdots$).

The corresponding musical intervals are listed in Table1. We treat this sequence X as a categorical time series, and fit variable-length Markov chains to construct a tree model. Note that at the stage of constructing a tree model, we excluded the 0 intervals from the sequence X in order to highlight the vertical transitions in pitch.

Table 1: Corresponding pitch intervals

$\lvert x \rvert$	**Pitch intervals**
1	Augmented unison / minor second
2	Major second
3	Augmented second / minor third
4	Major third / diminished fourth
5	Perfect fourth
6	Augmented fourth / diminished fifth
7	Perfect fifth
12	Perfect octave

4. VARIABLE-LENGTH MARKOV CHAIN MODEL

4.1. Variable-Length Markov Chains

Our main method of extracting peculiar transition patterns from Japanese and Chinese folk songs is to fit variable-length Markov chains (VLMCs) from the musical data. VLMC has its origin in data compression within information theory. The art of VLMC modeling has been successfully applied to the classification and identification of DNA sequences [8] in bioinformatics and coding data compression [9] in information theory.

A variable-length Markov chain, also known as a FSMX model or a finite-memory source, is a Markovian process having a sparse memory structure with some states that closely cohere. The structure can be characterized by a parsimonious number of transition probabilities for stationary categorical time series, and can be used to construct a probabilistic suffix tree-structured model. We probabilistically created a suffix tree in modeling a variable-length Markov chain to compare minimum transition patterns occurring with high probabilities.

4.2. Context Algorithm

In order to construct the models for both Japanese and Chinese folk songs, we used the context algorithm. The context algorithm is a method to divide a sequence (stationary categorical time series) X into wu, a concatenation of two contexts w and u, by measuring the differences between $P(\cdot|wu)$ and $P(\cdot|w)$ with the Kullback-Leibler divergence [10]. In principle, the smaller the Kullback-Leibler divergence, the closer the two probabilities. Therefore, our aim is to filter out an optimal model that minimizes the Kullback-Leibler divergence, and to represent the estimated minimal space.

4.3. Transition Probability

After pattern extraction, we look at the most frequent combinations for the features of each element by calculating the estimated transition probability $\hat{P}(x)$ with the chain rules of

$$P(x) = P(x_1) \prod_{i=2}^{t} P(x_i|\hat{c}(x_{i-1}, x_{i-2}, x_{i-3}, \cdots\cdots, x_1))$$

where $\hat{c}(x_{i-1}, x_{i-2}, x_{i-3}, \cdots\cdots, x_1)$ is the estimated pattern from sequence X using the context algorithm.

5. RESULTS

5.1. Frequency of the First Transition

A cumulative relative frequency diagram of the first transitions $|x_t|$ for both types of folk songs is shown in Figure 4. The profile shows that the frequencies are extremely high in the 0 and ± 2 intervals, and form a symmetric pattern taking the 0 interval as an axis in both types of folk songs. This implies that pitch transitions occur almost equally in both descending and ascending directions.

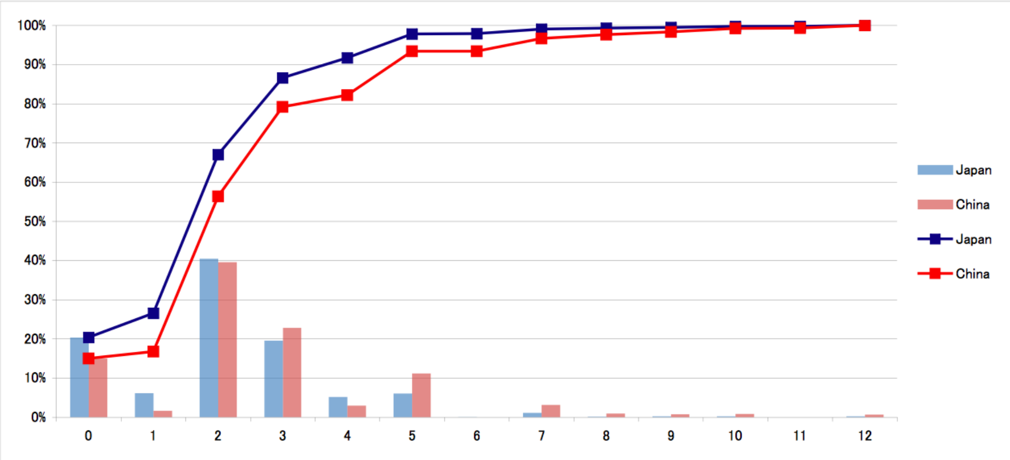

Figure 4: Cumulative relative frequency diagram

In order to compare the differences in the first transitions for Japanese and Chinese folk songs in more detail, we prepared an expected contingency table with a null hypothesis and applied the chi-square test to twelve semitones that represent approximately 99.8% of the frequency. The results are shown in Table 2.

Table 2: Contingency table for twelve semitones

	0	**± 1**	**± 2**	**± 3**	**± 4**	**± 5**	**± 6**
Japanese folk songs	41,130	12,431	81,789	39,453	10,477	12,253	218
	▲ 20.37%	▲ 6.16%	▲ 40.50%	▽ 19.54%	▲ 5.19%	▽ 6.07%	▲ 0.11%
Chinese folk songs	18,664	2,133	49,297	28,443	3,684	13,977	21
	▽ 15.00%	▽ 1.71%	▽ 39.63%	▲ 22.86%	▽ 2.96%	▲ 11.24%	▽ 0.02%

	± 7	**± 8**	**± 9**	**± 10**	**± 11**	**± 12**	**Total**
Japanese folk songs	2,317	381	478	455	47	511	201,940
	▽ 1.15%	▽ 0.19%	▽ 0.24%	▽ 0.23%	▲ 0.02%	▽ 0.25%	100.0%
Chinese folk songs	3,980	1,206	944	1,113	15	922	124,399
	▲ 3.20%	▲ 0.97%	▲ 0.76%	▲ 0.89%	▽ 0.01%	▲ 0.74%	100.0%

As the value of the chi-square statistic exceeds the criteria for the $p = 0.01$ level (d.f. = 12), we may reject the null hypothesis. In the table, the signs ▲ and ▽ indicate significantly high and low patterns, respectively. From the results of the contingency table, compared to Japanese folk songs, Chinese folk songs have a greater tendency to reach the intervals of minor thirds (± 3), perfect fourths (± 5) and above perfect fifths (± 7). On the other hand, Japanese folk songs tend to take intervals below the perfect fourth.

5.2. Overall Summary of the VLMCs

For Japanese folk songs, the maximal and the mean Markov chain depths for the data were 8 and 3.443, respectively. In the same way, for Chinese folk songs, the maximal and the mean Markov chain depths along the data were 7 and 2.785, respectively.

Here, we look at a confusion matrix from the fitted model and calculate its accuracy as the proportion of all predictions that were correct. A confusion matrix, a visualization tool typically used in machine learning, contains information about predicted and actual classifications [11]. The accuracy results for the models for Japanese and Chinese folk songs are 51.97% and 45.24%, respectively. These indicate that between approximately 45 to 50% of the musical data obey the Markov property; however, the values were smaller than we had expected. This remains as a matter for further research and other diagnostics need to be considered.

5.3. Estimated Structures

In this section, due to space limitations, we refer only to transitions with probabilities in excess of 1.00% within the order-4 Markov chains rather than drawing the entire tree-structure for both Japanese (Figure 5) and Chinese (Figure 6) folk songs.

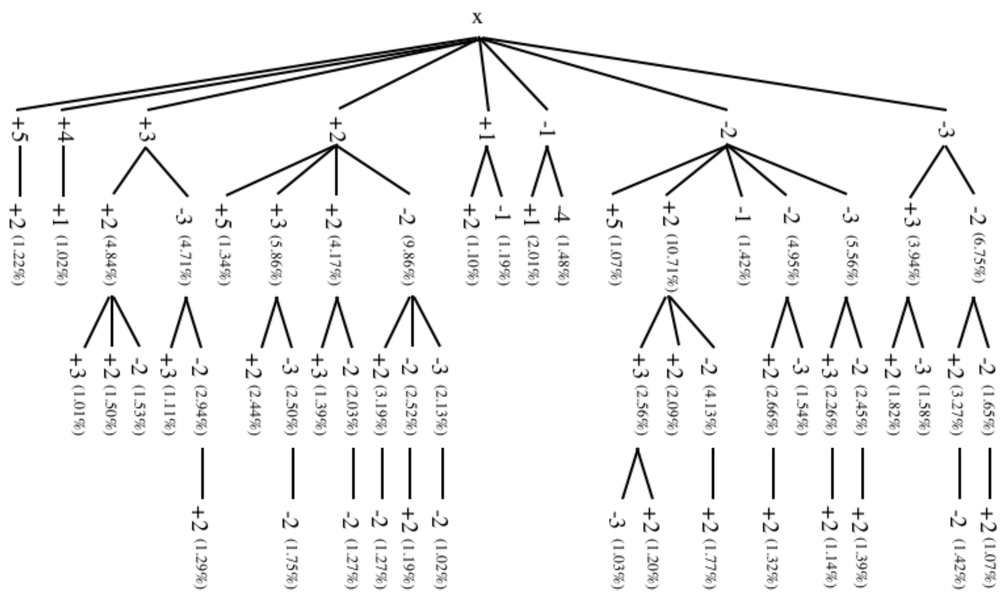

Figure 5: Tree structure for Japanese folk songs

5.3.1. Mean Markov Chain Depths

As we compare the mean Markov chain depths considering the frequency results for the first transitions, the minimal structures for Japanese folk songs tend to create longer schemas than for the Chinese folk songs, and vertical transitions are sung within a small range; to be exact, below intervals within the perfect fourth pitch. On the other hand, the minimal structures for Chinese folk songs tend to create shorter schemas than Japanese folk songs, and vertical transitions extend

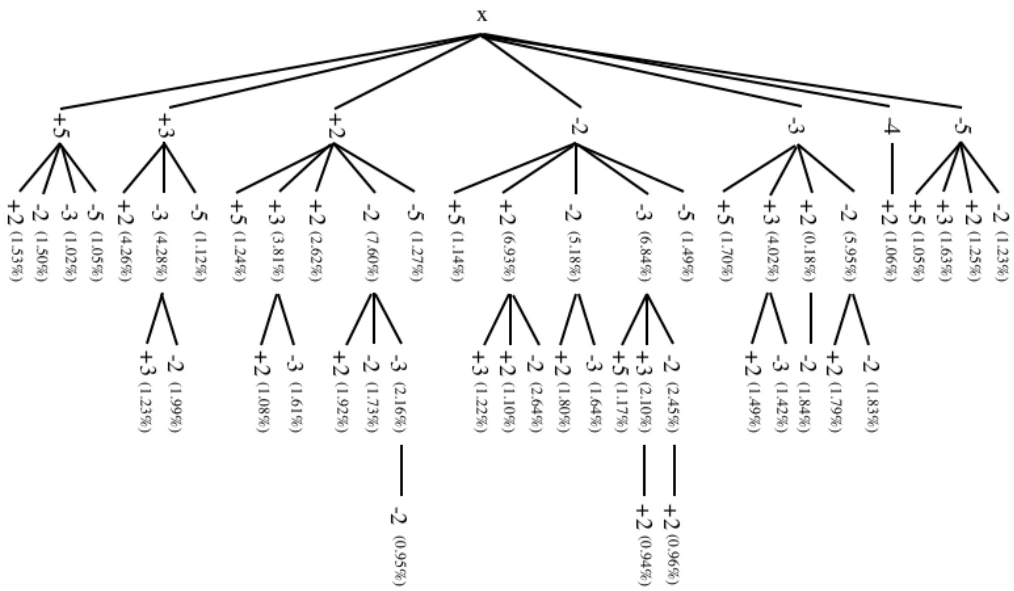

Figure 6: Tree structure for Chinese folk songs

beyond the interval of the perfect fourth pitch.

5.3.2. Patterns summing to zero

Irrespective of the country of origin, we find that most values of the estimated chains sum to 0. Although 0 intervals were excluded from the sequence X before constructing the trees, this finding corresponds to the schema that melodic leaps are followed by progressions back to the first note in Western music [2].

5.3.3. Patterns summing to five

We find that the values of the estimated chains also sum up to ± 5, which is the interval of the perfect fourth. For reasons of expediency, we bring in Koizumi 's tetrachord theory [4]. Tetrachord is a unit consisting of two stable outlining tones that fill the interval of a perfect fourth, called nuclear tones, and one unstable intermediate tone located between the nuclear tones.

In our previous study [5], we estimated from minimal data that transition patterns for tetrachords are salient characteristics in the melodies of Japanese folk songs. In this study, as expected, tetrachords are a major factor in Japanese folk songs. While there were more terachords located internally in the estimated structures of Japanese folk songs, the overall number of structures forming tetrachords was lower than for Chinese folk songs.

5.3.4. Patterns summing to twelve

While the Japanese folk songs had a transition probability of 0.08% for structures with overall intervals summing up to ± 12, Chinese folk songs accounted for 2.05%. As mentioned at the beginning of the paper, the Japanese people have experienced strong Chinese influences. However, this result implies the possibility of Chinese folk songs having had a strong influence, precisely in the sound sensation of the octave, for a fairly long time (meanwhile Japan entered an era of isolation from foreign countries), from countries, such as India and Europe, where a strong perceptual similarity exists between tones consisting of the perfect octave interval [12].

REFERENCES

[1] Temperley, D. *The Cognition of Basic Musical Structures*. MIT Press, Cambridge, Massachusetts, 2001.

[2] Meyer, L. B. *Explaining Music*. University of California Press, Berkeley, California, 1973.

[3] Lerdahl, F and Jackendoff, R. *A Generative Theory of Tonal Music*. MIT Press, Cambridge, Massachusetts, 1983.

[4] Koizumi, F. *Nihon dento ongaku no kenkyu 1 (Studies on Traditional Japanese Music 1)*. Ongaku no tomosha, Tokyo, 1958.

[5] Kawase, A and Tokosumi, A. Estimating the structures of Japanese folk songs uing a VLMC model. In *Proceedings of the 6th International Conference of Cognitive Science 2008*, Seoul, Korea, 2008.

[6] CCARH. KernScores. http://kern.humdrum.net/, 2009. [Accessed 2009 September 5].

[7] Chai, W and Vercoe, B. Folk music classification using hidden Markov models. In *Proceedings of the International Conference on Artificial Intelligence 2001*, Las Vegas, Nevada, 2001.

[8] Shmilovici, A and Ben-Gal, I. Using a VOM model for reconstructing potential coding regions in EST sequences. *Computational Statistics*, 22(1):49–69, 2007.

[9] Begleiter, R, El-Yaniv, R, and Yona, G. On prediction using variable order Markov models. *Journal of Artificial Intelligence Research*, 22:385–421, 2004.

[10] Bühlmann, P and Wyner, A. J. Variable length Markov chains. *Annals of Statistics*, 27:480–513, 1999.

[11] Provost, F and Kohavi, R. On applied research in machine learning. *Machine Learning*, 30:127–132, 1998.

[12] Deutsch, D. *The Psychology of Music. 2nd Edition*. Academic Press, San Diego, California, 1999.