Music Generation System Based on Human Instinctive Creativity

Hidefumi Ohmura[1], Takuro Shibayama[2], Keiji Hirata[3], and Satoshi Tojo[3] ⋆

[1] Tokyo University of Science
[2] Tokyo Denki University
[3] Future University Hakodate
[4] Japan Advanced Institute of Science and Technology
hidefumi.ohmura@gmail.com

Abstract. Most human beings, not only composers, have musical creativity; they can hum and whistle without musical education. We focused on the creativity of creating simple melodies such as humming, and developed a system that generated three melodies based on the physical relationship of notes and probability density functions. We confirmed that the system could create simple melodies like humming and whistling. Moreover, we confirmed that the output melodies of the system included various musical elements such as mode, scale, and rhythm.

Keywords: pitch, mode, scale, note value, rhythm

1 Introduction

Music is written by composers in various cultures, and their musical creativity is exemplary.

However, most human beings, not only composers, have musical creativity. The reason is that we can hum or whistle simple melodies unconsciously [1]. Humming is considered a beneficial human ability for surviving in society. Interestingly, most hummers and whistlers do not receive musical education. Though some of them cannot read musical scores, they can hum and whistle. Therefore, it is an indisputable fact that most human beings have musical creativity. In this paper, we focus on this creativity.

Humming melodies are improvised, except those that are remembered. Instead of composing consciously, people compose melodies unconsciously and intuitively. Thus, they select a matching note instinctively from moment to moment. We regarded this instinctive matching between notes as primitive physical relationships of notes [2]. From this idea, in this study we developed a musical generation system for three simple melody lines (like humming melodies).

First, we review the instinctive matching between notes, that is, the primitive physical relationships of notes. Next, we propose a computational model that

⋆ This work was supported by JSPS KAKENHI Grant Numbers JP17K12808, JP16H01744, JP17K02347, JP18K00210.

creates melodies on the basis of the relationships of notes. Finally, we describe the system using a model and discuss the system outputs.

2 Relationships of Musical Notes

Controlling musical notes quantitatively based on physical relationships, we can create improvised simple melodies like humming. In this section, by defining the relationships of notes, we show a method to treat both rhythm and pitch. Then, to use the relationships quantitatively, we introduce theories of musical emotion.

2.1 Relationships Between Pulses (Note Values)

In music, an iteration is an important primitive pattern, and a pulse is the most primitive element in a rhythm. Cognitions of patterns from cyclical relationships between one pulse and another help in understanding rhythms.

Essentially, relationships between pulses are defined using ratios of prime numbers, and the simplest relationship is 1:1. The relationship of 1:1 means that the cycle of one pulse coincides with that of another, so the listener may perceive just a single pulse. When a listener hears two pulses (whose relationship is 1:2), they may feel a duple meter. When a relationship is 1:3, a listener may perceive a triple meter. Moreover, when a relationship is 1:5, a listener may perceive a quintuple meter. However, one generally perceives a quintuple meter as a $2 + 3$ meter, because the quintuple meter is relatively difficult for the human ear.

Next, let us consider 2:3 and 3:2 ratios combining duple and triple meters. These relationships provide the listener with a polyrhythm. A listener perceives one basic pulse as an upbeat, and another as a downbeat. In the cases of 3:4 and 4:3, a listener perceives a polyrhythm consisting of triple and quadruple meters.

Actual music consists of many pulses. A listener perceives the strongest or most-frequent pulse as the meter of the music, and less-frequent pulses as weak beats and up beats. Monophony, however, lacks beats, so that a listener at times may not perceive any meter. For example, some pieces of Gregorian chant provide some rhythmic interpretations. This is also true in melodies of humming.

2.2 Relationships Between Pitches (Intervals)

In music, patterns consisting of pitches are also important. These patterns are explained in musical theories of temperament (how to determine the frequency of each note) and mode (which notes to use). The value of a pitch depends on the vibrational frequency of air. Real sound consists of multiple frequencies, so that we perceive the lowest frequency as the pitch, which is also called the fundamental frequency. As with rhythm, the patterns of two pitches are called musical intervals, and are defined by the relationship between the frequencies.

A relationship of 1:1 between two pitches creates an interval of a perfect prime, called unison. A relationship of 1:2 creates an interval of a perfect octave.

A relationship of 1:3 creates an interval of a perfect fifteenth, which is two octaves. These patterns provide more consonant intervals to a listener.

Next, let us consider 2:3 and 3:2, combining duple and triple frequencies (see Fig. 1). Regarding a basis of 660 Hz, 990 Hz is a pitch of 3/2 times the base. In letter notations, this interval indicates the relationship between E5 and B5. However, depending on temperament the interval is imprecise. Similarly, 440 Hz is a pitch of 2/3 times the base. In letter notations, this interval denotes the relationship between E5 and A4. This pattern is called a perfect fifth, which is a consonant interval following an octave.

Let us consider 3:4 and 4:3, combining triple and quadruple (see Figure 1). Regarding a basis of 660 Hz, 880 Hz is a pitch of 4/3 of the base. In letter notations, this interval denotes the relationship between E5 and A5. Depending on temperament, this interval is also imprecise. Here, 495 Hz is a pitch of 3/4 of the base. In letter notations, it refers to the relationship between E5 and B4. This pattern is called a perfect fourth, which is a consonant interval following a perfect fifth. The Pythagorean scale (and principle of the Sanfen Sunyi) only use relationships of duples and triples.

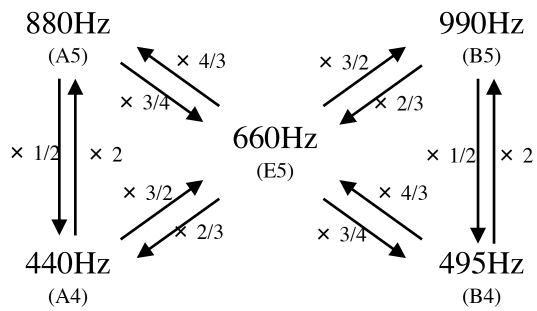

Fig. 1. Relationships between pitch (interval)

In the case of intervals, unlike that of rhythm, a quintuple is important. In particular, in a relationship between three pitches, a 3:4:5 creates a consonance code, called a major triad. A relationship consisting of a double, triple, and quintuple creates intonation.

2.3 Theories of Musical Expectations

In this study, we tried to control musical expectations on the basis of theories of musical expectations [11]. Here, we introduce these theories.

Meyer demonstrated that the deviations of expectations arouse emotions when listening to music [3]. This concept is based on Dewey's theory, whereby conflict causes emotions [4]. The deviation from the listeners' expectation when listening to music arises from a partial or complete disregard of rules that were accepted in advance. This indicates an increase in contingency because of augmented uncertainties. These uncertainties have a commonality with complexity in the optimal complexity model [9], which illustrates the relationship between complexity and hedonic value (Figure 2). This commonality suggests the existence of a relationship between uncertainty and emotion. Comparing sounds in our everyday life and their relationships reveals interesting viewpoints. Let us survey the points in Figure 2. At Position ①, the complexity is quite low. A listener can easily predict the features of the sound, and a prediction event is likely to arise in the proposed system. For example, the pure tick-tock beat of a clock not only sounds boring, but also causes displeasure in listeners. Indeed, some clock users cannot sleep with this sound. Position ② has a higher complexity than Position ①, causing pleasure in the listener. The listener can approximately predict the next sound, and can recognize both realizations and deviations from expectation at that position. Listeners may regard sounds as musical, because the sounds have rules as well as deviations from the rules at this position. There are different levels of complexity in each musical genre. Rhythms, children's songs, and folk songs have lower complexities than pop music, for example. Therefore, a wide range of sounds fall under this position. Position ③ has an appreciably high complexity. Here, a listener cannot predict the next sound and can only recognize deviations from expectations. Free jazz and contemporary music are comparatively new genres that can be considered in this context. These styles of music are unpleasant for some listeners. Non-musical sounds such as the noise from a crowd at a sports field may belong to Position ③. According to Berlyne, the relationship between complexity and emotion adapts to the listener's experiences (the dotted line in Figure 2). That is, music at Position ③ elicits the pleasure of highly experienced listeners. They may feel displeasure from the music at Position ②. A listener who has experienced a large amount of music may discover the complexity in music.

As we regard complexities as uncertainty, we can adapt entropy as presented in Shannon's information theory [10] to the computational model. In the next section, we attempt to formulate human humming melodies.

3 A Model of Human Humming Melodies

We conducted a study to propose a model of musical rhythm and interval patterns [2]. In the model, we can control both pitches and note values with an identical idea. Here, we show the model.

3.1 Lattice Space with Duple and Triple Relationships

First, we provide a lattice space for intervals, which consists of ratios of 1:2 and 1:3 (see Figure 3). The x-axis indicates triple relationships, and the y-axis duple relationships. Each point of intersection is the frequency of a note. In this figure,

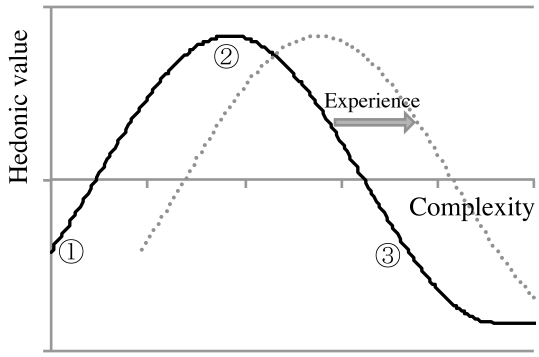

Fig. 2. Optimal complexity model (modified from [9]). When a listener has many musical experiences, his or her function moves to more/less complex.

440 Hz is a basic frequency, such as a root note. Each frequency value is rounded. Though each point of intersection shows a letter notation with a frequency, the farther the point is from the base frequency (in this case 440 Hz), the larger the error based on the Pythagorean comma becomes.

In the lattice space of Figure 3, the distance between one frequency and another of close value is far. Thus, the lattice space of Figure 3 is slanted as Figure 4. This will be useful in setting parameters.

Second, we provide a lattice space for rhythm, which also consists of ratios of 1:2 and 1:3 (see Figure 5). The unit in this lattice space is beats per minute (bpm). In this figure, 72 bpm is the basic frequency of pulse. The x-axis indicates triple relationships, and the y-axis duple relationships. Each point of intersection is the frequency of a pulse. In this figure, there are symbols of musical notes; a quarter note is 72 bpm.

3.2 Probability Distribution in Lattice Spaces

The proposed model provides probabilities at points of intersection using a probability density function. In this study, we employ a two-dimensional Gaussian function for the probability density function. The function is represented as follows:

$$
\begin{aligned}
f(x, y) = & \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \\
& \times exp\left(-\frac{1}{2(1-\rho^2)}\left(\frac{(x-\mu_x)^2}{\sigma_x^2} - 2\rho\frac{(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y} + \frac{(y-\mu_y)^2}{\sigma_y^2}\right)\right)
\end{aligned}
$$

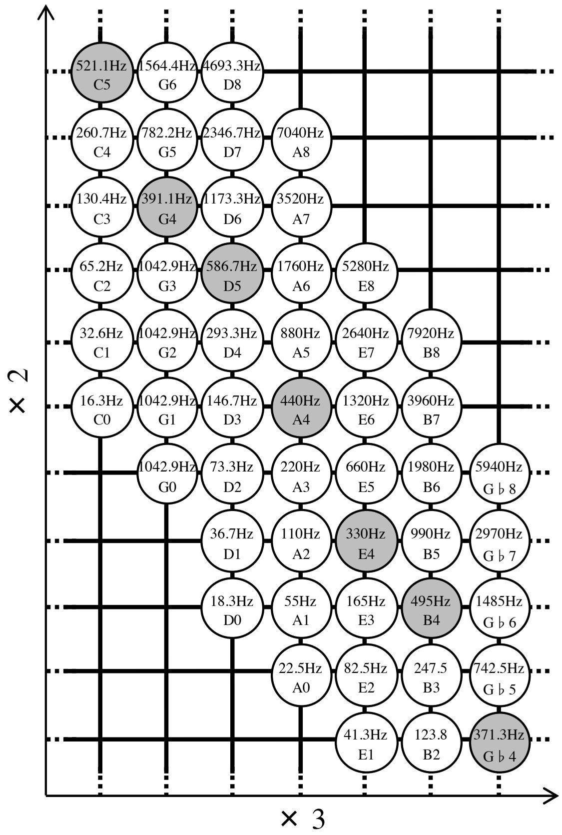

Fig. 3. Lattice space for pitches with duple and triple relationships

This equation describes a two-dimensional normal distribution with mean μ and variance σ^2 in each axis. These variances function like the parameters of the complexity or entropy. ρ is a coefficient of correlation between values on the x- and y-axis. By assigning values to x and y, the probabilities are calculated from μ_x, μ_y, σ_x, σ_y, and ρ.

In the lattice space of pitch, when we apply a mixture distribution to a probability density function, the model expresses musical modes such as major and minor [11].

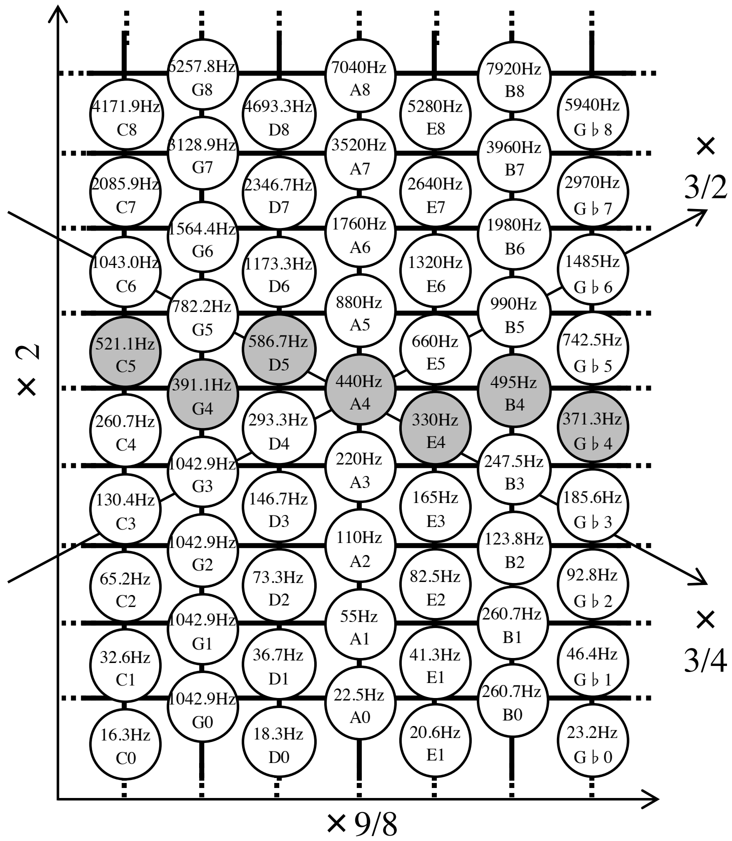

Fig. 4. Lattice space of slanted Figure 3

4 Music Generation System

4.1 Implementation

We implemented the proposed model using HTML and JavaScript for a music generation system[5]. We confirmed the operation of the system in Google Chrome. The system outputs up to three melody lines. Users control parameters of the probability density functions for pitches and rhythm.

The system includes the lattice space for pitches like that of Figure 4. The center of the lattice space is 440 Hz. The system also includes the lattice space for musical values like that of Figure . The center of the lattice space is 72 bpm.

[5] `https://sites.google.com/site/hidefumiohmura/home/program/csmc2018`

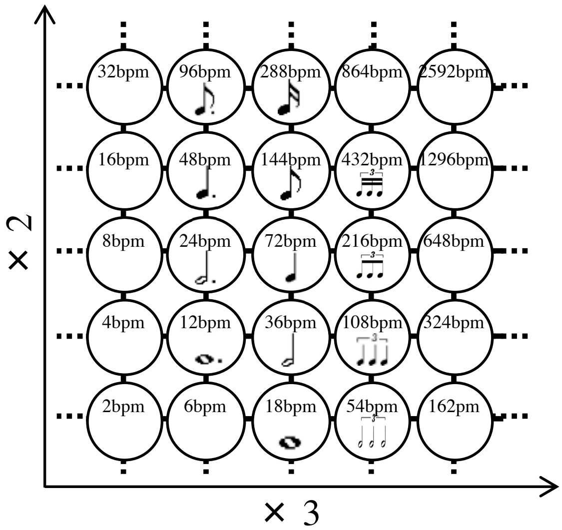

Fig. 5. Lattice space for musical values with duple and triple relationships

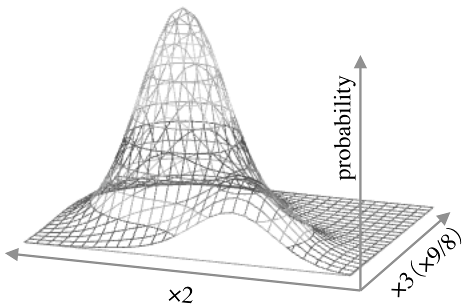

Fig. 6. Probability of notes in lattice spaces

There are probability density functions on these lattice spaces. Controllable parameters of the functions are mean (μ_x, μ_y), variance (σ_x, σ_y), and correlation (ρ). Furthermore, users can also control parameters of sub-functions. The parameters of a sub-function includes a weight (w), which is the ratio of a sub-function to a main function. When a weight (w) is 0, a sub-function is ignored.

We explain the flow of execution through the program. When users push a play button, the program executes iterative processing as follows:

1. Select a pulse from the lattice of value notes according to the probability density function.
2. Is the timing of the pulse hitting a note?
   yes: Select a pitch from the lattice of pitches according to the probability density function and output it.
   no: Do nothing.
3. Go to 1 as a next step.

The iterative processing executes at intervals according to a parameter of tempo whose initial value is 2592 bpm.

4.2 System Operating Instructions

The operation screen consists of three panels: Sound Control Panel, Pitch Control Panel, and Note Value Control Panel. Here, we explain step-by-step how to use them.

Sound Control Panel At [Sound Control], users can control play/stop, volume, tempo, duration, waveform, and melody lines of the outputs. The header of the operation screen also includes a play/stop button. The values of volume, tempo, and duration are controlled by sliders. The value of tempo indicates the program cycle time in bpm. The value of duration is the length of time of each note. By controlling this value, melodies show articulations as staccato and tenuto. With the waveform selector, users can select from "Sin," "Square," "SawTooth," and "Triangle." Moreover, users can select "Bongo" and "Piano" as actual sound source samples. Though each pitch is calculated by duple and triple, the pitches of Bongo are defined by the 12 equal temperaments. Moreover, [Sound Control] includes a preset selector that provides each setting for discussion.

Pitch Control Panel At [Pitch Control], users can control parameters of each probability density function for pitches of melody lines using sliders. Each value of the probability density function is shown in the upper right [Pitch Cells]. The values of the melody lines are shown in different colors: The first line is cyan, the second line is magenta, and the third line is yellow. A darker color indicates a higher value. The initial values of means are set for 440 Hz (A). By means of the buttons at the bottom in [Pitch Cells], each probability density function is set as visible or invisible. The operations of the melody lines are independent. Using the upper left buttons, users select an operating melody line. The parameters of the main function are controlled by sliders at [Main-function Settings]. The parameters of the sub-function are controlled by sliders at [Sub-function Settings]. During system execution, selected pitches are shown at the bottom right [Circle of Fifth]. Therefore, users can confirm the output pitches in real time.

Note Value Control Panel At [Note Value Control], users can control parameters of each probability density function for note values of melody lines using sliders. Each value of the probability density function is shown in the upper right [Note Value Cells]. As is the case with [Pitch Control], the values of the melody lines are shown in different colors: the first line is cyan, the second line is magenta, and the third line is yellow. A darker color indicates a higher value. By means of the buttons at the bottom in [Note Value Cells], each probability density function is set as visible or invisible. The operations of the melody lines are independent, as in the case with [Pitch Control]. During system execution, selected note values are shown at the bottom right [Pulses]. Therefore, users can confirm the output pulses of note values in real time. The pulses can be zoomed by using buttons, and displayed on a log scale by using a toggle button.

5 Discussion

We prepared example settings of the parameters as presets. We discuss them here.

First, [Humming (a melody line)] provides only one melody line like a humming melody. The variance of the function in the lattice space of pitches is a comparatively low value, and the mean is set near 440 Hz. Depending the settings, the system outputs A, E, D, B, and G frequently. The y-axis variance of the function in the lattice space of note values is comparatively high, so that the system outputs quarter notes, half notes, and eighth notes (regarding a pulse of 72 bpm as a quarter note). Outputs sound like melodies of the pentatonic scale, which are used in folk songs of Scotland, East Asia, and other areas. Rhythms of the outputs sound like duples or quadruples. Therefore, the outputs sound like simple melodies such as humming.

Second, [Humming (three melody line)] provides three melody lines. The variances of the functions in the lattice space of pitches are higher than those of the previous preset. The mean of the function for the first line is the same value as that of the previous preset. To distinguish melody lines, the mean for the second line is higher than that of the first line, and the mean for the third line is lower than that of the first line. In the lattice of note values, the mean of the first line is the same value as that of the previous preset, the y-axis mean of the second line is lower than that of the first line, and the x-axis mean of the line is higher and the y-axis mean is lower than those of the first line. The outputs sound like a melody in six-eight time.

The previous two presets use only the main function in the lattice space of pitch. Therefore, the outputs sound like melodies of Dorian mode. Here, we discuss presets using two functions in the lattice space of pitch. In [positive mode], the variance of the main function is a comparatively low value, and the variance of the sub-function is higher than that of the main function. Moreover, the mean of the sub-function is higher than that of the main function. The outputs sound like melodies of Ionian or Lydian mode. On the other hand, in [negative mode], the mean of the sub-function is set opposite the values of [positive mode]. The other parameters of [negative mode] are same as those of [positive mode]. The outputs sound like melodies of Aeolian or Phrygian mode.

We now discuss the preset of [Polyrhythm]. In this preset, each variance of the function in the lattice space of note values is set to the minimum value because the system selects only one pulse. The output of the first line is a pulse of 72 bpm, the output of the second line is a pulse of 54 bpm, and the output of the third line is a pulse of 108 bpm. The relationship between the first line and the second line is 4:3, and the relationship between the first line and the third line is 2:3. Therefore, outputs sound like melodies of polyrhythm.

The presets of [Okinawa] and [Miyako-bushi] are controlled as each mode in the lattice space of pitches. Unfortunately, the settings of the parameters are not good. The outputs include notes outside the modes.

As can be heard from the presets, the system outputs not only simple melodies (such as humming and whistling), but also melodies with musical elements such as scale, mode, rhythm, and metrical structure. These outputs reveal that musical mode and scale are not discrete but continuous, and that there are rhythm structures without sequences such as musical scores.

It may be premature to conclude that the system creates music because the outputs do not include a musical theme or chorus, and are far from real music. However, it is important that the system could create melodies like humming and whistling.

As future work, we will investigate various genres of music, not only western music, but also that of various other cultures. Moreover, we will compare the outputs with some animal songs, such as bird song.

6 Conclusion

In this study, we focused on the creativity of creating simple melodies such as humming, and developed a system to generate three melodies. We confirmed that the system could create melodies like humming and whistling. Moreover, we confirmed that the outputs of the system included various musical elements such as mode, scale, and rhythm.

References

1. Jordania, J.: Music and emotions: humming in human prehistory, in *Proceedings of the International Symposium on Traditional Polyphony (Tbilisi),* pp. 41–49, (2010)

2. Ohmura, H., Hirata, K., Tojo, S., and Shibayama, T.: Investigation of a computational unit model for mode and rhythm based on deviations and realizations from musical expectations, in *Proceedings of the 13th International Symposium on Computer Music Multidisciplinary Research (CMMR2017),* pp. 439–449, (2017)
3. Meyer, L. B.: *Emotion and meaning in music*, University of Chicago Press (1956)
4. Dewey, J.: The theory of emotion: I: Emotional attitude, *Psychological Review*, Vol. 1, No. 6, pp. 553–569 (1894)
5. Narmour, E.: *The analysis and cognition of basic melodic structures*, University of Chicago Press, (1990)
6. Huron, D.: *Sweet anticipation: Music and the psychology of expectation*, MIT Press (2006)
7. Lerdahl, F. and Jackendoff, R.: *A generative theory of tonal music*, MIT Press (1983)
8. Forte, A.: *The structure of atonal music*, Yale University Press (1973)
9. Berlyne, D. E.: *Aesthetics and psychobiology*, Appleton Century Crofts (1971)
10. Shannon, C. E.: *The mathematical theory of communication*, University of Illinois Press (1949)
11. Ohmura, H., Shibayama, T., and Hamano, T.: Generative music system with quantitative controllers based on expectation for pitch and rhythm structure, in *Proceedings of the Eighth International Conference on Knowledge and Systems Engineering (KSE2016),* (2016)