Music Synthesis based on Impression and Emotion of Input Narratives

Saya Kanno
Ochanomizu University
saya@itolab.is.ocha.ac.jp

Takayuki Itoh
Ochanomizu University
itot@is.ocha.ac.jp

Hiroya Takamura
Tokyo Institute of Technology
takamura@pi.titech.ac.jp

ABSTRACT

This paper presents a technique to synthesize the music based on the impression and emotion of the input narratives. The technique prepares a dictionary which records the sensibility polarity values of arbitrary words. The technique also supposes that users listen to the sample chords and rhythms, and input the fitness values to the pre-defined impression word pairs, to learn the relations between features of chords/rhythms and these impression. After these processes, the technique interactively synthesizes the music for input narratives. It estimates the fitness values of an input narrative to the impression word pairs using the dictionary, and then selects the chord and rhythm progressions those impressions and emotions are the closest to the narrative. Finally, the technique synthesizes the output tune by combining the chord and rhythm. We suppose this technique encourages to express impression and emotion of the input narratives by generating music.

1. INTRODUCTION

We may want to express the impression and emotion of narratives by creating another media. We sometimes write our sentiments as review documents, or illustrate the scenes of the novels. Here, we may provide too much information to the readers while writing reviews even though if they do not want to know the contents and details before reading the narratives. Or, we may provide unnecessary or inadequate impression from the illustrations. Our motivation of this study is to express the impression and emotion of narratives applying music.

This paper presents a technique to express the impression and emotion of narratives by synthesizing music. Music has effects to derive imagination, behavior, and emotion to the listeners. These are effective to every people, whether the listeners are deeply interested in the music or not [15]. We expect such effects are helpful to feel the impression and emotion of input narratives, without knowing too much information. Moreover, we have been already familiar with multi-modal arts which combines music and other expression, such as Opera and Ballet. Therefore, we expect that people may feel new types of sensation and inspiration by listening to the music created for the narratives while the people are reading them.

We had a questionnaire "Would you like to listen to the music based on the impression and emotion of the narratives after you read that narratives?" performed on the Web, and received answers from 57 participants. As a result, 24 participants answered "Strongly want to listen", and 26 of others answered "Rather want to listen". This result suggests that 88% of the participants are interested in listening to the music based on the impression of the narratives. We expect the music generated based on the impression of the input narratives is effective to stimulate interest of users. We can imagine the abstract impression of the narrative before reading by listening to the music generated by the presented system. Or, we can play the background music that matches the impression of the narratives which we are reading them. Moreover, we expect this technique is useful to publish and intercommunicate the impression of literature by uploading the music. We also expect the uploaded tunes bring interest and imagination to the narratives.

2. RELATED WORK

There has been a long history of studies on automatic music generation with different media including images and narratives. Many of the studies generate music based on high-level semantics such as structures of stories or scenes, while many other studies are based on abstract impression.

Background music generation for slideshows, movies, computer animations, or real spaces is a recent active research issue. Some of the studies are based on high-level semantics [3] [11] while some of others are based on abstract impression [1] [12] [14]. The study presented in this paper is closer to the latter type of studies.

There have been smaller number of studies on music synthesis adapting to narratives applying natural language processing techniques. Endo et al. [4] presented a technique to generate music based on arguments and grammatical structures of input narratives. This technique is not based on the impression and emotion of the narratives. Kitahara et al. [10] presented a technique to automatically compose music by a note sequence generation from the impression and emotion of the input narratives. We subjectively suppose this approach is not always suitable to generate user-preferable music, because it just automatically generates sequences of notes without learning the impression of listeners and utilizing the user-preferred music patterns. Cruz et al. [2] also presented a technique to generate music based on the emotion of input narratives. Again, this technique does not adopt preferences or impression of users to the generation of music. On the other hand, our technique is based on the mash up of user-prepared chord and rhythm progressions, and learning of the impression and emotion of the chords and rhythms. This approach can take into account the impression of the listeners to the music, and utilize the user-prepared musical patterns.

Ishizuka et al. [9] presented a theme music arrangement system based on impression of story scenes. Its architecture is close to our study since it arranges input theme tunes based on the numeric impression values. However, it does not deeply discuss how to calculate impression values and learn users' own impressions. Also, this technique does not apply user-prepared musical patterns.

3. PRESENTED TECHNIQUE

This section describes the processing flow and implementation detail of the presented technique. The technique consists of the following three technical components:

1. **Preliminary data construction:** Selection of impression words and musical features, and dictionary construction applying a semantic orientation calculation technique [16]. These steps are applied once by the system developer.
2. **Learning:** Calculation of coherency between the musical features and these impressions. This step is applied once for each user as a preprocessing.
3. **Interactive process:** Music synthesis for input narratives.

Here, we prepared the chords and rhythms by reference to the commonly used patterns in pops, rock, jazz, and classical music. Our current implementation supposes BPM (Beat per Minute) and number of measures of all the chords and rhythms are equal.

Our current implementation generates music by just synthesizing chords and rhythms. It is our on-going work to implement the procedure to synthesize melodies to generate the music, since the presented mechanism to select chords and rhythms can be similarly applied to melodies.

The below subsections describe the processing flow and implementation detail of the each technical component.

3.1 Preliminary data construction

Our preliminary data construction phase includes the following two processes: 1) selection of impression word pairs and musical features, and 2) dictionary construction. Our current implementation constructs Japanese dictionary; however, the presented mechanism is not limited to specific natural languages.

3.1.1 Selection of impression word pairs and musical features

This step firstly selects impression word pairs used for semantic orientation calculation, and musical features used for the selection of chords and rhythms. We listed the impression word pairs and musical features shown in Table 1 as the candidates. We learned impression word pairs from Ikezoe et al. [8], and musical features from Hasegawa et al. [6]. Also, we subjectively added frequency of several chords, ratio of inharmonic tones, and frequency of tones for each of drums, because they are often effective to express the particular mood or emotion.

Table 1. Candidates of impression word pairs and musical features.

Impression word pairs for chord progression
Bright - Dark
Light - Heavy
Enjoyable - Wistful
Brassy - Simple
Tripping - Quiet
Energetic - Calm
Musical features for chord progression
Average of tones
Distribution of tones
Number of simultaneous tones
Ratio of inharmonic tones
Frequency of major, minor, seventh, major seventh, and minor seventh chords
Impression word pairs for rhythm progression
Fast - Slow
Light - Heavy
Quiet - Loud
Brassy - Simple
Energetic - Calm
Musical features for rhythm progression
Frequency of tones for each of drums
Total number of tones
Frequency of 16-, 8-, and 4-beat notes
Frequency of triplets

Then, we conducted a questionnaire for the selection of impression word pairs and musical features. We asked participants to listen to the sample chord and rhythm progressions, and answer the subjective fitness between the samples and impression word pairs in the five-point Likert scale. We calculated the correlativity between each of the fitness values and each of the musical feature values. If a fitness value was not well correlated with any of the musical feature values, we removed the impression word pair. At the same time, we removed a musical feature value from the candidate, if it was not well correlated with any of the fitness values. As a result, we listed the impression word pairs and musical features shown in Table 2 in our implementation.

3.1.2 Dictionary construction

We constructed a Japanese dictionary containing noun, verb, adjective, and adverb, with normalized values representing the fitness to all the impression word pairs. Here, we defined the fitness value corresponding to one of the impression words as "1", and the value corresponding to the other word as "-1". We calculated the fitness values of each word in the dictionary for each of the impression word pairs, applying a semantic orientation calculation technique [16]. As a result, the j-th word in the dictionary has a M_w di-

Table 2. Finally selected impression word pairs and musical features.

Impression word pairs for chord progression
Bright - Dark Enjoyable - Wistful Tripping - Quiet Energetic - Calm
Musical features for chord progression
Average of tones Distribution of tones Number of simultaneous tones Ratio of inharmonic tones Frequency of major, major seventh, and minor seventh chords
Impression word pairs for rhythm progression
Fast - Slow Light - Heavy Brassy - Simple
Musical features for rhythm progression
Frequency of tones for Toms, Snare drums, Bass drums, cymbals, and High-hats Total number of tones Frequency of 16-beat notes Frequency of triplets

mensional vector $g_j = \{g_{j1}, g_{j2}, ..., g_{jM_w}\}$, where M_w is the number of impression word pairs, and g_{ji} is the fitness value of the j-th word to the i-th impression word pair.

3.2 Learning of impression and emotion

This step learns the relationships between the musical features and these impressions. This step is applied once for each user as a preprocessing. Our current implementation supposes to ask a user to listen to the sample chord and rhythm progressions, and answer their fitness to the impression word pairs. This process then calculates the relationships between the musical features of the sample chords/rhythms and the fitness values answered by the user. Here, our current implementation applies a linear multi-regression analysis to solve the relationships, because we suppose the relationships can be approximated as linear functions. This study applies the following equation to express the relationships between the musical features and the fitness values:

$$f_i = \sum_{j=1}^{M} a_{ij} m_j \tag{1}$$

Here, f_i is the fitness value for the i-th impression word pair, M is the number of musical features, m_j is the j-th musical feature, and a_{ij} is the coefficient for the i-th impression word pair and the j-th musical feature. The linear multi-regression analysis solves the values of the coefficients, given the set of values of f_i and m_j. Our implementation applies this process independently to chords and rhythms. We can estimate the impression of later provided chord and rhythm progressions by calculating the fitness values using the above equation, after solving the coefficients of the equation.

3.3 Interactive process

This step synthesizes music adopting to the impression and emotion of input narratives. Our implementation interactively provides the synthesized music when a narrative is given. This process is divided into the following components: document analysis, selection of chord and rhythm, and music synthesis.

3.3.1 Document analysis

This step firstly applies a morphological analysis to the input narratives. Our current implementation applies an open source Japanese morphological analysis software MeCab [17] to the narratives. It then extracts noun, verb, adjective, and adverb from the result, and then calculates the average of fitness values recorded in the dictionary generated by the preliminary data construction process, for each of the impression word pairs. The technique treats the average values as the estimated impression and emotion of the narrative.

Here, long narratives or novels contain progression and variation of impression and emotion. Therefore, it is not always adequate to calculate the average impression/emotion of the whole narrative. Our implementation calculates the averages of the fitness values scene-by-scene. Currently we suppose that input narratives are manually divided to multiple scenes. We would like to apply automatic scene recognition techniques to divide the input narratives as a future work.

As a result, we express the impression of a scene of the input narrative as a M_w dimensional vector $h = \{h_1, h_2, ..., h_{M_w}\}$. We calculate the fitness value to the i-th impression word pair as $h_i = \frac{1}{M_s} \sum_j^{M_s} g_{ji}$, where M_s is the total number of words appeared in the scene. g_{ji} is the fitness value of the j-th word to the i-th impression word pair, extracted from the preliminary constructed dictionary.

3.3.2 Selection of chord and rhythm

Next, this step selects chord and rhythm progressions. The technique calculates the musical features m_j for all the prepared chords and rhythms, and then estimates the fitness values f_i for the chords and rhythms by using the equation (1). Our current implementation uses the four-dimensional fitness values for the following impression word pairs, [Bright - Dark], [Enjoyable - Wistful], [Tripping - Quiet], and [Energetic - Calm] for the chord progressions. For the rhythm progression, it uses the three-dimensional fitness values for the following impression word pairs, [Fast - Slow], [Light - Heavy], and [Brassy - Simple].

Our technique generates trajectories of the fitness values consisting of N_{scene} vertices and $N_{scene} - 1$ segments, where N_{scene} is the number of scenes. It then selects the chords and rhythms so that the trajectories of the fitness values of the selected chords or rhythms looks similar to the trajectory generated from the input narrative. Our current implementation selects sets of chords or rhythms of the scenes which minimizes the following formula:

$$min \left(\alpha \sum_{i}^{N_{scene}} \|\mathbf{f}_i - \mathbf{g}_i\| + (1-\alpha) \sum_{i}^{N_{scene}-1} \left(1 - \frac{\mathbf{f'}_{i,i+1} \cdot \mathbf{g'}_{i,i+1}}{|\mathbf{f'}_{i,i+1}||\mathbf{g'}_{i,i+1}|} \right) \right) \tag{2}$$

Here, α is a user-defined constant real value satisfying $0 \leq \alpha \leq 1$, $\mathbf{f}_i$ is a fitness value vector of a chord or rhythm for the i-th scene, $\mathbf{g}_i$ is a fitness value vector of the i-th scene calculated from the input narrative. $\mathbf{f'}_{i,i+1}$ and $\mathbf{g'}_{i,i+1}$ are defined as follows:

$$\mathbf{f'}_{i,i+1} = \mathbf{f}_{i+1} - \mathbf{f}_i, \mathbf{g'}_{i,i+1} = \mathbf{g}_{i+1} - \mathbf{g}_i. \tag{3}$$

The first term of equation (2) attempts to minimize the sum of distances between the fitness value vectors of the chords/rhythms and narratives, while the second term attempts to minimize the geometric differences of the trajectory between the chords/rhythms and narratives. We expect this definition realizes the selection of chords and rhythms taking into account the total variation of the generated music in addition to the similarity of impressions.

3.3.3 Music synthesis

Finally, this step synthesizes the selected chord and rhythm. Our current implementation supposes that the chords and rhythms are stored as independent MIDI files which contain a single track for the chord or rhythm. It simply copies the tracks of selected chord and rhythm into another MIDI file consisting of the two tracks. This implementation will be extended to incorporate melodies as a future work.

4. EXPERIMENT AND DISCUSSION

This section introduces our experiments with the presented technique. We had the following steps for the experiments:

Step 1: We asked participants to listen to the sample chords and rhythms. Then, we asked them to answer the fitness of the impression word pairs in six-point Likert scale, to learn their sensibility for the music.

Step 2: We generated tunes for input narratives applying the learning results of each of the participants. Then, we asked them to evaluate the degree of coincidence of the impression between the narrative and tune in five-point Likert scale. We also asked them to freely comment their impressions for the output tunes.

Here, all chords recorded in the sample MIDI files were played as half notes, and their timbre specified by the program number was piano, in this experiments. Lengths of all sample MIDI files of chords and rhythms were eight measures, and their tempo was 120 BPM (Beat Per Minute).

Table 3. The contents of the input narrative.

Scene 1	your profile of when you sing with smile was very beautiful.
Scene 2	I fell in love with you by looking that profile.
Scene 3	But, I was disappointed in love.

Table 4. The fitness for impression word pairs in each scene of the input narrative.

	Scene 1	Scene 2	Scene 3
Bright - Dark	-0.034480	0.013741	-0.176004
Enjoyable - Wistful	-0.115318	0.171532	0.122126
Tripping - Quiet	-0.149857	0.617872	0.580159
Energetic - Calm	0.325041	0.528052	0.531251
Fast - Slow	-0.122990	0.236133	0.220244
Light - Heavy	0.564965	0.436106	-0.091217
Brassy - Simple	0.521458	0.468449	-0.241370

Table 5. Results and evaluation of participants A and B.

Participant A			
	Chord	Rhythm	Evaluation
Scene 1	chord 12	rhythm 2	4
Scene 2	chord 9	rhythm 3	5
Scene 3	chord16	rhythm 11	1
Total evaluation	2		
Participant B			
	Chord	Rhythm	Evaluation
Scene 1	chord 9	rhythm 19	2
Scene 2	chord 11	rhythm 19	3
Scene 3	chord 16	rhythm 16	4
Total evaluation	4		

In the Step 1, we provided the same sample chords and rhythms for both the participants. After asking participants to listen to the generated tunes in the Step 2, we showed participants the contents of input narratives and explained how the scenes are split. We did not explain how the narrative was split in order, because we wanted participants to be unconscious of contents of the input narrative while they were evaluating the output music. We prepared a dramatic short narrative divided to three scenes in this experiment. Participants evaluated three tunes generated for the scenes of the narrative. Table 3 describes the scenes, and Table 4 shows the fitness of impression word pairs.

The participants of this experiment were two female students in the master's course, who had experiences and expert skills of musical instruments and vocals, and a certain level of musical knowledge. The length of the MIDI files were unified in 16 seconds in this experiment. As a result, lengths of all the tunes synthesized by the proposed technique were 48 seconds. We prepared 23 pieces of chords and rhythms for this experiment.

Table 5 shows the chord/rhythm selection results for two participants, and the evaluations by the participants. The result denotes the synthesized tunes are different despite completely same narrative was provided to the two participants. This suggests our technique can generate different tunes according to the sensibility of each user.

Let us discuss on the results of participant A. Table 6 and Table 7 show the fitness values for chords and rhythms respectively.

Participant A mentioned that "I had an impression that wistful event was happened but I thought it was not a serious scene" while listening to the tune for Scene 1. The fit-

Table 6. The fitness values of participant A for the chords in each scene.

	Scene 1	Scene 2	Scene 3
Bright - Dark	-0.166850	0.888388	0.014798
Enjoyable - Wistful	-0.270462	0.165336	-0.190081
Tripping - Quiet	0.090049	0.443500	0.382450
Energetic - Calm	0.042111	0.653096	-0.068819

Table 7. The fitness values of participant A for the rhythms in each scene.

	Scene 1	Scene 2	Scene 3
Fast - Slow	0.075839	0.496695	0.033937
Light - Heavy	0.145658	0.437712	0.752391
Brassy - Simple	0.075839	0.996108	0.485212

ness value for [Enjoyable - Wistful] of the chord for Scene 1 was negative, close to “Wistful”. The above comment for Scene 1 is consistent to the fitness value for the chord. Participant A also mentioned “This scene seems happy, and more exciting than the previous scene” for Scene 2. Here, all fitness values for the chord and rhythm in Scene 2 are higher than the fitness values in Scene 1, as shown in Table 6 and Table 7. This variation of fitness values conforms to the transition of the fitness values of input narrative shown in Table 4. Again, these results denote the above comment for Scene 2 is consistent to the variation of fitness values. Actually, evaluations of the participant A for Scenes 1 and 2 were 4 and 5, relatively high, as shown in Table 5. On the other hand, participant A mentioned that “I had an impression that a good event was happened and end up from music, however, the narrative of this scene was sad.” for Scene 3. This comment denotes impression of participant A for music and input document was really opposite. Many of the fitness values of Scene 3 in the input narrative are positive as shown in Table 4, which suggest nimble and energetic impression. These fitness values are inconsistent to the comment that participant A felt sad impression after reading Scene 3. To solve this problem, we would like to extend our implementation of document analysis component customizable to users’ sensibility.

Next, let us discuss on the results of participant B. Table 8 and Table 9 show the fitness values for chords and rhythms respectively.

Participant B mentioned that “Although I received a cheerful impression like exercising from the music, it was actually a quiet scene with no movements” for Scene 1. However, the fitness value for [Tripping - Quiet] and [Energetic - Calm] of the chord for Scene 1 were negative, as shown in Table 8, where these values denote the quiet impression. This result suggests our experiment might not successfully learn the sensibility of participant B. In addition, participant B mentioned that “I felt that there was no change between Scene 1 and 2, because the rhythm of Scenes 1 and 2 were the same”. This comment suggests rhythm was an important musical factor for participant B, and we may need to analyze which musical factor participants remark. Participant B mentioned that “I felt the sudden change to have a serious feeling” while listening to the tune for Scene 3. Actually, many fitness values of chords and rhythms for Scene 3 got smaller than those for Scene 2, as shown in Table 8 and Table 9. This result demonstrates our technique could represent the significant changes of the impression between the scenes. Participant B mentioned that “Entire flow of music substantially coincides with the flow of story I imagined from the music”, and finally rated the total evaluation as 4.

Table 8. The fitness values of participant B for the chords B in each scene.

	Scene 1	Scene 2	Scene 3
Bright - Dark	0.059933	0.896135	0.043383
Enjoyable - Wistful	-0.141256	0.108170	0.216329
Tripping - Quiet	-0.235864	0.969589	-0.123636
Energetic - Calm	-0.041046	0.015811	-0.111593

Table 9. The fitness values of participant B for the rhythms in each scene.

	Scene 1	Scene 2	Scene 3
Fast - Slow	0.507989	0.507989	0.322039
Light - Heavy	0.18577	0.18577	0.363906
Brassy - Simple	0.642532	0.642532	-0.111264

Our experiment assigned the same chord progression for Scene 3 for participants A and B, as shown in Table 6 and Table 8. However, participant A mentioned “I got the impression like happy ending”, while participant B answered “It is a serious scene”. These comments are actually opposite. This result suggests that the same chord and rhythm progression may give different impression to different users.

The results introduced in this section suggest that the impression of input narratives was close to the impression of generated music in many cases. We would like to customize the document analysis based on users’ sensibility so that we can improve the degree of coincidence of the impression between the music and narratives.

5. CONCLUSION

We proposed a technique to synthesize music based on emotion and impression of the input narratives. This technique selects the chords and rhythms to the scenes of the narratives using by estimating the fitness values for the impression word pairs from the musical features. In the preliminary data construction step, we selected sets of impression word pairs as a result of survey on the correlations between musical feature values and pairs of sensibility words. This paper introduced experiments which generated different impression of music in response to the sensibility of the participants. The result suggests that we could generate tunes which roughly match to the fitness values specified by the sensibility polarity dictionary, while we still have issues on document analysis.

Our future issues include the following:

- Review of the association of impression word pairs and musical features.

- Re-design of interface to input the sensibility of users.
- Implementation of automated recognition of scene breaks in the narratives.
- Association of temporal deployment of music and narrative.
- Embedding melodies while the synthesis of music.

The presented study listed candidates of music features of chords and rhythms excluding several important ones such as velocity of the notes. We suppose participants might have bias in representation of impressions and feelings because such important features were constant in our sample chords and rhythms. We would like to review the music features again, and then extend our implementation by adding important features.

We received several comments from the participants of the presented experiment regarding the methodology of the learning step. One of the typical comments was that "I could not keep the constant criteria in mind, while listening to the sample chords or rhythms, and answering the fitness values." Reflecting these comments, we would like to test other interfaces to input the sensibility of users. Hevner presented a music evaluation method [7] which asks participants to choose one of the word groups which matches to the listened tunes. Several studies on music evaluation applied tournament methods which ask participants to comparatively select one of the tunes, and finally specify the best tune. We would like to apply such methods to appropriately learn the preferences of participants.

Our current implementation learns users' sensibility just for automatic selection of chords and rhythms. We expect our technique can be improved by re-defining fitness values of characteristic words in the preliminary constructed dictionary according to the semantics of the narratives or users' sensibility. We would like to implement a mechanism to customize the dictionary based on this discussion.

Finally, we would like to develop the melody selection process, in addition to larger database construction for chords and rhythms, to realize the generation of more impressive and emotional music. Our on-going work applies an automatic music composition method featuring a genetic algorithm [13] and maximum likelihood estimation [5].

6. REFERENCES

[1] P. Casella, A. Paiva, MAgentA: An Architecture for Real Time Automatic Composition of Background Music *International Workshop on Intelligent Virtual Agents*, pp. 224-232, 2001.

[2] R. Cruz, A. Brisson, A. Paiva, E. Lopes, I-Sounds - Emotion-Based Music Generation for Virtual Environments, *Affective Computing and Intelligent Interaction*, pp. 769-770, 2007.

[3] P. Dunker, P. Popp, R. Cook, Content-Aware Auto-Soundtracks for Personal Photo Music Slideshows, *IEEE International Conference on Multimedia and Expo*, pp. 1-5, 2011.

[4] J. Endo, T. Kitadate, T. Ogata, A Music Generation/Expression Mechanism in the Narrative Generation System: A Consideration from the Application Systems, *The 29th Annual Meeting of the Japanese Congnitive Science Society*, pp. 3-29, 2012.

[5] S. Fukayama, K. Nakatsuma, S. Sako, Y. Yonebayashi, T.-H. Kim, S.-W. Qin, T. Nakano, T. Nishimoto, S. Sagayama, Orpheus: Automatic Composition System Considering Prosody of Japanese Lyrics, *Entertainment Computing - ICEC 2009*, Springer Berlin Heidelberg, pp. 309-310, 2009.

[6] T. Hasegawa, T. Nishimoto, N. Ono, S. Sagayama, Composer Identification from MIDI Data by Combination Features of Pitch and Duration based on Musical Knowledge, *IPSJ SIG Technical Reports[MUS] Vol. 2009-13*, pp. 47-52, 2009.

[7] K. Hevner, Experimental studies of the elements of expression in music, *American Journal of Psychology*, Vol. 48, pp. 246-268, 1936.

[8] T. Ikezoe, Y. Kajikawa, Y. Nomura, Music Database Retrieval System with Sensitivity Words Using Music Sensitivity Space, *Journal of Information Processing Society in Japan*, Vol. 42, No. 12, pp. 3201-3212, 2001.

[9] K. Ishizuka, T. Onisawa, Generation of Variations on Theme Music Based on Impressions of Story Scenes Considering Human's Feeling of Music and Stories, *International Journal of Computer Games Technology*, 2008.

[10] K. Kitahara, H. Watanabe, D. Ando, Method for Automatic Generation of Music Reflecting the Web Activity, *Japanese Society for Sonic Arts*, Vol. 1, No. 1, pp. 8-11, 2012.

[11] H.-C. Lee, I.-K. Lee, Automatic Synchronization of Background Music and Motion in Computer Animation, *Computer Graphics Forum*, Vol. 24, No. 3, pp. 353-361, 2005.

[12] C.-T. Li, M.-K. Shan, Emotion-based Impressionism Slideshow with Automatic Music Accompaniment, *ACM International Conference on Multimedia*, pp. 839-842, 2007.

[13] Y. Maeda, Y. Kajihara, Automatic Generation of Musical Tone Row and Rhythm Based on the Twelve-Tone Technique Using Genetic Algorithm, *Journal of Advanced Computational Intelligence and Intelligent Informatics*, Vol. 14, No. 3, pp. 288-299, 2010.

[14] J. Nakamura, T. Kaku, K. Hyun, T. Noma, S. Yoshida, Automatic Background Music Generation based on Actors Mood and Motion, *The Journal of Visualization and Computer Animation*, Vol. 5, No. 4, pp. 247-264, 1994.

[15] H. Saito, *Kokoro wo ugokasu oto no shinnrigaku (Psychology of sound that moves the sensibility)*, YAMAHA Music Media, 2011.

[16] H. Takamura, T. Inui, M. Okumura, Extracting Semantic Orientations Using Spin Model, *Annual Meeting on Association for Computational Linguistics*, pp. 133-140, 2005.

[17] Joint research unit project of Kyoto University Graduate School of Informatics and NTT Communication Science Laboratories, *MeCab*. http://mecab.googlecode.com/svn/trunk/mecab/doc/index.html, 2015.2.24