A Tambourine Support System to Improve the Atmosphere of Karaoke

Takuya Kurihara, Naohiro Kinoshita, Ryunosuke Yamaguchi, and Tetsuro Kitahara
Nihon University, Japan
{kurihara,kinoshita,yamaguchi,kitahara}@kthrlab.jp

ABSTRACT

Karaoke is a popular amusement, but people do not necessarily enjoy karaoke when they are not singing. It is better that non-singing people engage in karaoke to enliven it, but this is not always easy, especially if they do not know the song. Here, we focus on the tambourine, which is provided in most karaoke spaces in Japan but are rarely used. We propose a system that instructs how a non-singing person plays the tambourine. Once the singer choose a song, the tambourine part for this song is automatically generated based on the standard MIDI file. During the playback, the tambourine part is displayed in a common music game style with the usual karaoke-style lyrics. The correctness of the tambourine beat is fed to the display. The results showed that our system motivated non-singing people to play the tambourine with a game-like instruction even for songs that they did not know.

1. INTRODUCTION

Karaoke is an amusement quite familiar to most people. Many people enjoy singing karaoke. However, they do not necessarily enjoy karaoke when they are not singing. In fact, our survey of 30 students shows that 27 of the 30 students sometimes feel bored when they are not singing. This is mainly because they do not know what to do and is especially common when they do not know the song being sung. In most karaoke spaces in Japan, a percussive instrument, such as tambourine, is provided and customers can freely play it with the song. If someone can appropriately play such an instrument, it enhances the karaoke experience for everyone. In reality however, this instrument is rarely used because people do not know how to play or are too shy to play it. Here we hypothesize computing technologies can facilitate tambourine play by showing how to play it and by reducing shyness. This would be an effective too to enliven karaoke.

There have been many attempts to support and/or enhance karaoke. Cano et al. developed a system that modifies the user's singing voice by morphing it into a prerecorded voice of the same melody sung by another person [1]. Liu et al. developed a system that evaluates singing voice based on pitch and rhythm [2]. Daido et al. developed a system for evaluating singing enthusiasm using three acoustic features: A-weighted power, fall-down, and vibrato extent [3]. Tsai et al. proposed an automated singing evaluation method close to the human rating that exploits various acoustic features, including pitch, volume, and rhythm [4]. Thus, there are no attempts to enhance karaoke with a percussive instrument.

In this paper, we propose a tambourine support system, called Karatan to enhance karaoke. This system automatically generates a tambourine part (instructs how to play the tambourine) from a MIDI file and shows it with a usual karaoke-style lyrics display. A non-singing person can play tambourine according to the system's instruction without thinking of how to play it. Because he/she can enjoy playing the tambourine like common music games, the player will likely not feel self-conscious about it.

The remainder of the paper is organized as follows: In Section 2, we describe the concept and details of our system. In Section 3, we report experimental results conducted to confirm the effectiveness of our system. Finally, we conclude the paper in Section 4.

2. KARATAN: TAMBOURINE SUPPORT SYSTEM FOR KARAOKE

This system is designed to improve the engagement of non-singing people in karaoke. This happens because they do not know what to do when no singing. In general, non-singing people simply listen to the song being sung. However, just listening to the song does not necessarily enliven the experience. If they know the song being sung, they can sing it together. However, even songs known to all people are not always sung by all.

Here, we focus on the tambourine that is provided in most karaoke boxes in Japan, but rarely used. There are two possibilities why the tambourine is not used. The first is that they do not know how to play the tambourine effectively for unknown songs. The second possibility is that no one is motivated to play the tambourine. They think they might be embarrassed by inappropriately playing the tambourine rather than making karaoke more exciting by playing the tambourine.

Our system facilitates use of the tambourine through the following functions:

- Automatic tambourine part generation
  Once a song is chosen, the tambourine part is automatically generated from the standard MIDI file (SMF). We use SMFs because most karaoke machines are performed with music data equivalent to SMFs.
- Real-time tambourine performance feedback
  The tambourine part is displayed on the karaoke video screen and the performance of the player (whether he/she correctly played) is displayed in a common music game style. With this game-style display, people can easily try the tambourine performance.

Figure 1. Wii Tambourine

2.1 System Overview

This system is used with the Wii Tambourine, which we developed by building the Wii Remote into a tambourine (Figure 1). While the user plays the tambourine, the performance is analyzed by observing the accelerations of the Wii Remote. The procedure in this system consists of the following three steps:

1. Practice
   Once this system is launched, the system tries to establish a connection with the Wii Tambourine. After it succeeds, the practice mode starts. The user learns to play the tambourine in the parctice mode. This mode is needed not only for the user to learn this system's tambourine part display but also for the system to acquire the data for performance identification.
2. Tambourine part generation
   After the practice mode completes, the user chooses a song. Then, the system generates a tambourine part from the SMF of the chosen song. The tambourine part here includes all aspects of instructions for playing the tambourine (Figure 2). It includes intensities and body motions as well as the timing of the beat.
3. Real-time tambourine performance feedback
   Once the tambourine part is generated, the system starts to playback the SMF. During the play back, the tambourine part is displayed in a music game style along with the usual karaoke-style lyrics display (Figure 3). The user plays the tambourine while the singer sings along with this display. The tambourine performance is analyzed in real time and, if correctly played, is fed back to the display.

In the rest of this section, we describe these three steps in more details.

2.2 Practice mode

First, the user practices how to play the tambourine according to the instructions on the display (Figure 4). The contents of the practice are (1) beating freely, (2) beating strongly, (3) shaking, (4) beating at the upper position, and (5) beating at the lower position. In this order, the user freely practices the tambourine until he/she is satisfied. The system acquires the following data needed for performance identification:

- The mean value of the temporal differentials of accelerations in beating freely (α_1 in Equation (1)).
- The mean value of the number of direction changes within 20 frames of shaking (α_2 in Equation (3)).

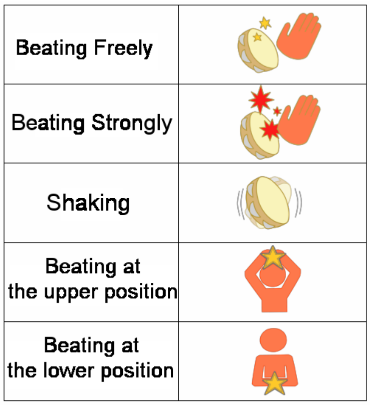

Figure 2. Instructions in tambourine performance

Figure 3. Karaoke screen

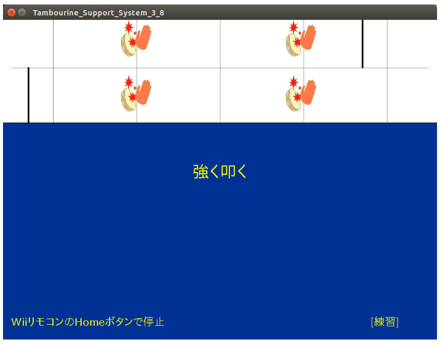

Figure 4. Practice mode screen

- The mean value of the temporal differentials of accelerations in strong beating (α_3 in Section 2.4.3).
- The mean value of temporal differentials and directions of rotation etc. used in body motion classification (Section 2.4.4).

2.3 Tambourine part generation

The difficulty in playing the tambourine in karaoke lies in a dilemma in that the tambourine should enhance the experience but not disturb the singing. In general, the singing voice in the chorus section (i.e., the high point of a song) in the verse-chorus form tends to be excited, so loud tambourine sound and a large body motion would not disturb the singing voice. Also in instrumental solo sections, the tambourine can be relatively freely played because there is no singing voice during these parts. We therefore adopt the following basic policies:

- The tambourine is basically played in the same way

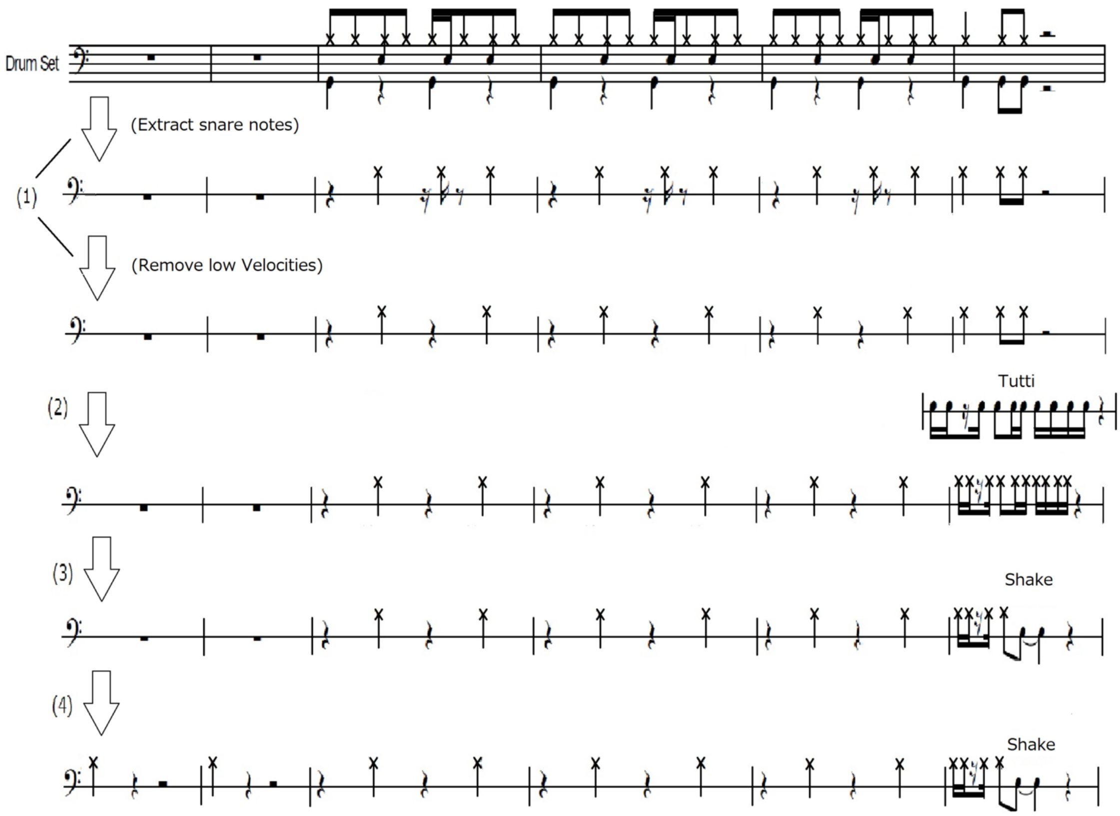

Figure 5. Example of tambourine part generation

as the snare drum.

- The tambourine is sometimes shaken instead of beaten.
- The tambourine is played quietly for quiet songs.
- Body motions are made for chorus and instrumental solo sections.

Based on these policies, our system generates the tambourine part according to the following algorithm:

2.3.1 *Tambourine part generation*

(1) First, a note sequence for the snare drum is extracted at the eighth-note level from the given SMF and regarded as tentative tambourine part. The snare drum part sometimes contains notes with low velocities that are not easy for non-musicians to play with the tambourine. The notes with velocities lower than a defined threshold are therefore removed. The threshold is determined as a velocity at which the histogram of the velocities of all notes in the given SMF's snare part has the lowest valley (Figure 5 (1)).

(2) Next, this tentative tambourine's part is corrected according to tuttis. When seven or more instruments play three or more notes in the completely same rhythm within one measure, these notes are considered a tutti. The tambourine is also played in the same rhythm whether the snare is played or not (Figure 5 (2)).

(3) Successive short notes could be difficult for non-musicians to play. When three or more eighth notes are played successively, the user is allowed to shake the tambourine instead (Figure 5 (3)).

(4) According to this algorithm, the tambourine is not played in the sections that have no snare notes. From the viewpoint of enlivening up karaoke, no tambourine sections should be avoided. For every measure without snare notes, we insert a tambourine note to the first beat (Figure 5 (4)).

2.3.2 *Instruction of intensities*

After the tambourine part is generated, the intensity is determined for each note of the tambourine part. Here, we adopt two types of instructions: "no particular instruction (i.e., with your favorite intensity)" and "beat strongly." We call these a normal note and a strong note, respectively. These instructions should be linked to the mood of the song such that the ratio of strong notes are controlled according to the overall dynamics of the song.

The overall dynamics of the song is defined by the mean value m of the velocities of all snare drum notes. If this value is higher, the ratio of strong notes should be higher. We therefore define the ratio of strong notes by

$$R = \begin{cases} \frac{1}{3}(1 + \frac{m-M}{127-M}) & (m > M) \\ \frac{m}{3M} & (otherwise) \end{cases}$$

where M is the mean value of m obtained from a variety of songs.

2.3.3 *Instruction of body motion*

Moving the body would be more effective than just playing the tambourine, but doing it from the beginning to the end would be burdensome. For the chorus and instrumental solo sections, the system gives instructions on body motions to the user. Specifically, the system gives an instruc-

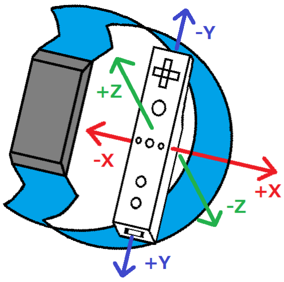

Figure 6. Wii Tambourine

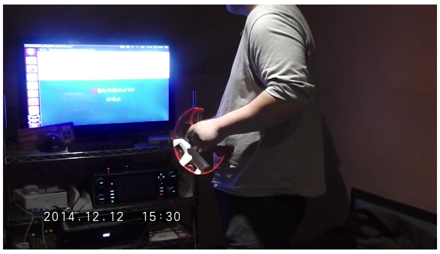

Figure 7. Photo of the experiment

tions on "at the upper position" or "at the lower position" alternately for every measure in the chorus and instrumental solo. The chorus sections are detected on the basis of Takada's method [6], which is based on Goto's idea [7] that the chorus section is included in the longest one of the repeated melodies, but is simplified because the target is a MIDI signal, not an audio signal. The instrumental solo sections can be easily found because the SMF we used includes an annotation of the instrumental solo sections as lyrics data.

2.4 Real-time tambourine performance feedback

While the song is being played, the generated tambourine part is displayed in a common music game style together with a usual karaoke-style lyrics (Figure 3). At the same time, the user's tambourine performance is analyzed through the acceleration data obtained from the Wii Remote. The accuracy of the performance is fed back through the display. Because the Wii Remote is built into the tambourine like Figure 6, it moves along the x-axis when it is beaten or shaken.

2.4.1 Beat detection

Every 16 ms (so the frame rate is 60 fps), the three-dimensional acceleration is observed. When the temporal differential of the x-axis acceleration is higher than a threshold, that is

$$|a_t - a_{t-1}| > 0.7\alpha_1, \quad (1)$$

where a_t is an acceleration at time t, the tambourine is considered to be beaten at time t. The threshold α_1 is the mean value of the temporal differentials of accelerations when tambourine is beaten in the practice mode.

2.4.2 Shake detection

Every 16 ms, the system counts how many times the direction of the tambourine's motion is changed for the last 20 frames. The change in the direction is detected from the zero-crossing of the acceleration. When

$$a_t a_{t-1} < 0, \quad (2)$$

the direction of the tambourine's motion is considered to be changed at time t. A shake is defined as the number c of direction changes is higher than a threshold, that is,

$$c \geq 0.7\alpha_2. \quad (3)$$

The threshold α_2 is the mean value of the numbers of direction changes within 20 frames in the practice mode.

2.4.3 Strong beat detection

Whether the tambourine is beated strongly is identified. This is performed in the same way as Section 2.4.1 except that a different threshold α_3 rather than as instead of α_1. The threshold α_3 is the mean value of the temporal differentials of accelerations when the tambourine is strongly beaten in the practice mode.

2.4.4 Body motion (hand's position) classification

The body motion in playing the tambourine ("at the upper" or "at the lower") is identified based on the method of Sawada et al [8]. This is performed by calculating the summation of the dissimilarities from the template, obtained in the practice mode, in the xy-, yz-, and zx-planes. Below, the definition of the dissimilarity in the xy-plane is described (the dissimilarities for the yz- and zx-planes are defined in the same way).

Every 16ms, the accelerations in the x- and y-axes are observed. Then, some features such as the mean values of temporal differential and direction of rotation for the last S frames are extracted ($S = 20$ in the current implementation). The dissimilarity in the xy-plane is defined as the mean square deviation of these features from those obtained in the practice mode. The dissimilarities in the yz- and zx-planes are defined in the same ways. We can then solve for the motion ("upper" or "lower") that minimizes the summation of the dissimilarities in the xy-, yz- and zx-planes.

3. EXPERIMENTS

We conducted experiments to assess the effects of our system. The participants were divided into groups, each of which consisted of three persons that played the roles of a singer, a tambourine player, and a listener. The three participants in each group knew each other because people usually go to karaoke with friends or acquaintances. We conducted two experiments. In Experiment 1, we used real Japanese popular songs that were included in a karaoke ranking and were known to the participants. In Experiment 2, we used songs unknown to the tambourine players to compare the effects of our system for tambourine players on known songs and unknown songs.

3.1 Experiment 1

3.1.1 Experimental conditions

We asked the participants to play karaoke with our system and the baseline system. The baseline system was implemented by removing the tambourine part generation and display as well as the performance feedback function from our system. This was equivalent to standard karaoke machines. The participants were 12 university students and were divided into four groups. The role (a singer, a tambourine player, or a listener) of each participant was determined considering the familiarity to the target songs described below. The most familiar person sings, and the least familiar person listens. The procedure is as follows:

For Groups 1 and 2:

1. Practice the Wii Tambourine
2. Sing the singer's favorite song with a real karaoke machine
3. Play karaoke with our system (an easy song)
4. Play karaoke with our system (a hard song)
5. Play karaoke with the baseline system (an easy song)
6. Play karaoke with the baseline system (a hard song)

For Groups 3 and 4:

1. Practice the Wii Tambourine
2. Sing the singer's favorite song with a real karaoke machine
3. Play karaoke with the baseline system (an easy song)
4. Play karaoke with the baseline system (a hard song)
5. Play karaoke with our system (an easy song)
6. Play karaoke with our system (a hard song)

We adopted such a crossover style in order to avoid the order effect. As easy songs, we used *Zankoku Na Tenshi No These* (Yoko Takahashi) for our system and *Memeshikute* (Golden Bomber) for the baseline system. As hard songs, we used *Senbonzakura* (WhiteFlame feat. Miku Hatsune) for our system and *Kimi No Shiranai Monogatari* (Supercell) for the baseline system. These songs were taken from the karaoke ranking of JOYSOUND 2013 [9]. We used different songs for our system and the baseline system because we aimed to avoid the effect of experience—if a participant uses the baseline system for a song after using our system for the same song, he/she may play the tambourine better with the baseline system because he/she has the completely same experience. We therefore carefully chose different songs so that their tambourine performance difficulties (e.g., the numbers of beatings, rhythmic complexities, etc.) are similar. The goal of Step 2 (sing the singer's favorite song) is to let the participants relax.

After using our system or the baseline system, we asked the participants the following questions:

For singers:

Q-S1 Do you think the listener enjoyed your song?

Q-S2 Could you sing comfortably?

Q-S3 Did the tambourine performance disturb your song?

Q-S4 Were tambourine's timings appropriate?

Q-S5 Do you want to try the tambourine performance?

For tambourine players:

Q-T1 Was it easy to play the tambourine?

Q-T2 Could you play the tambourine along to the music?

Q-T3 Was your performance monotonous?

Q-T4 Could you play the tambourine to the rhythm?

Q-T5 Could you enliven karaoke through the tambourine?

For listeners:

Table 1. Results of Experiment 1

(a) Singers

	Baseline system		Our system	
	Easy song	Hard song	Easy song	Hard song
Q-S1	5.25	4.75	4.25	4.75
Q-S2	5.50	5.25	4.75	5.25
Q-S3	6.00	5.25	4.75	5.00
Q-S4	5.75	3.75	4.75	2.75
Q-S5	5.00	4.00	4.00	2.75

(b) Tambourine players

	Baseline system		Our system	
	Easy song	Hard song	Easy song	Hard song
Q-T1	4.75	3.50	4.50	4.75
Q-T2	5.00	4.00	5.00	3.50
Q-T3	3.25	1.00	6.00	4.75
Q-T4	4.25	2.00	5.00	4.75
Q-T5	3.75	3.00	3.75	4.50

(c) Listeners

	Baseline system		Our system	
	Easy song	Hard song	Easy song	Hard song
Q-L1	4.50	5.50	3.75	4.75
Q-L2	4.25	4.75	3.50	4.50
Q-L3	5.25	5.25	3.25	4.00
Q-L4	4.75	5.00	3.50	4.75
Q-L5	3.50	3.50	4.50	4.25
Q-L6	3.25	2.25	4.50	4.25

Q-L1 Did you feel bored?

Q-L2 Did you have an interest in the song?

Q-L3 Was the tambourine noisy?

Q-L4 Did the tambourine correspond to the music?

Q-L5 Do you want to sing along with the tambourine?

Q-L6 Do you want to try playing the tambourine?

The answers to these questionnaires were on a scale from one to six. A positive answer has a high value. For example, "6" for Q-L1 means "definitely no" while "6" for Q-L2 means "definitely yes".

3.1.2 Experimental results

The results are listed in Table 1.

Tambourine players rated our system highly. In particular, it was good that our system did not make the performance monotonous. However, a participant answered that the generated tambourine score is complicated in part. In particular, it was difficult to play the tambourine while moving the body. It would be better to introduce a mechanism for controlling the difficulty according to the player's musical skill.

For listeners, the answers to Q-L5 and Q-L6 were particularly high. This means that our system successfully gave listeners high motivations to play the tambourine. Participants gave us an opinion that this system increases the enjoyment as a music game as well as a tambourine support system.

Singers did not evaluate our system highly. This is probably because the tambourine player's mistakes were noticeable because the system sometimes required complex performances over the player's skill. This also suggests that difficulty control should be introduced.

Table 2. Results of Experiment 2

(a) Tambourine players

	Baseline system		Our system	
	Known song	Unknown song	Known song	Unknown song
Q-T1	4.50	3.25	5.00	5.00
Q-T2	5.50	3.00	4.75	5.50
Q-T3	1.75	1.25	4.25	1.75
Q-T4	5.00	2.50	4.75	5.50
Q-T5	2.75	2.00	4.50	4.00

(b) Listeners

	Baseline system		Our system	
	Known song	Unknown song	Known song	Unknown song
Q-L1	5.50	3.25	4.75	4.25
Q-L2	5.50	3.75	4.75	3.75
Q-L3	4.50	3.75	4.00	5.00
Q-L4	4.00	2.50	4.00	3.75
Q-L5	5.00	1.75	4.25	5.00

3.2 Experiment 2

3.2.1 Experimental conditions

When the tambourine player did not know the song being sung, our system was expected to support the tambourine player. To confirm this, we conducted an experiment with a song unknown to the tambourine player (we call it an unknown song). We took an unknown song from RWC Music Database (Popular Music) because this database consists of J-pop-style songs originally composed for research purposes, which are certainly unknown to the participants. We used *REAL Na Gofun* (RWC Music Database) as an unknown song and *Memeshikute* (Goldenbomber) as a known song for our system, and *Replica* (RWC Music Database) as an unknown song and *Zankoku Na Tenshi No These* (Yoko Takahashi) as a known song for the baseline system.

The procedure of this experiment and the questionnaires were the same as Experiment 1 except that the unknown songs were used in Steps 4 and 6 instead of the difficult songs. Because this experiment focused on the tambourine player's behavior, we played the role of singer. Hence, there are no questions for the singer. The participants were the same as those in Experiment 1 except for the singers.

3.2.2 Experimental result

The questionnaire results are listed in Table 2. Our system was highly evaluated in almost all conditions for unknown songs. We found that without our system, it is difficult to enhance karaoke with tambourine play because such play is very conservative. With our system, the tambourine players successfully played the tambourine for unknown songs. In turn, listeners also enjoyed the songs.

4. CONCLUSION

Tambourines are provided in most karaoke boxes in Japan and are expected to play a significant role in enhancing karaoke, but they are rarely used in practice. To give an opportunity to play the tambourine, we developed a tambourine support system that generates a tambourine part. Through this system, the user can play the tambourine like a common music game even if he/she does not know the song being played. The results showed that our system motivated tambourine players to some extent but singers were confused when the tambourine player made a mistake due to complex tambourine instructions over the player's skill. Future issues will include difficulty adjustment of the tambourine part, support of more various body motions, and improvement of tambourine performance identification.

Acknowledgments

This work was supported by JSPS KAKENHI Grant Number 26240025.

5. REFERENCES

[1] Pedro Cano, Alex Loscos, Jordi Bonada, Maarten de Boer and Xavier Serra, "Voice Morphing System for Impersonating in Karaoke Applications" in *ICMC*, Germany, 2000.

[2] Liu Yuxiang, Jin Zeyu, Jia Jia and Cai Lianhong, "An Automatic Singing Evaluation System" in *Applied Mechanics and Materials*, Switzerland, 2012, Vol 128-129, pp. 504–509.

[3] Ryunosuke Daido, Seong-Jun Hahm, Masashi Ito, Shozo Makino and Akinori Ito, "A System for Evaluating Singing Enthusiasm for Karaoke" in *ISMIR*, Miami, 2011.

[4] Wei-Ho Tsai and Hsin-Chieh Lee, "An Automated Singing Evaluation Method for Karaoke System" in *IEEE Transactions on Audio, Speech and Language Processing*, 2011, pp. 2428-2431.

[5] http://www.nintendo.co.jp/wii/controllers/index.html

[6] Tomonori Takada and Hiroki Hashiguchi, "Detection of refrain from melody track in MIDI" in *Saitama University Bulletin*, Japan, 2006-7, pp. 107-109.

[7] Masataka Goto, "A Chorus Section Detection Method for Musical Audi Signals and Its Application to a Music Listening Station" in *IEEE Transactions on Audio, Speech and Language Processing*, 2006-9, Vol 14, No.5, pp.1783–1794.

[8] Hideyuki Sawada and Shuji Hashimoto, "Gesture Recognition Using Acceleration Sensor and Its Application to Musical Performance Control" in *Electronics and Communication*, Japan, 1997, Vol 80, No 5, pp. 9–17.

[9] http://www.joysound.com/st/2013yearranking

[10] Masataka Goto, Hiroki Hashiguchi, Takuichi Nishimura and Ryuichi Oka, "RWC Music Database:Music Genre Database and Musical Instrument Sound Database" in *ISMIR*, 2003-10, pp. 229–230.