A HUMANOID ROBOT THAT CAN SING AND DANCE TO MUSIC BY RECOGNIZING BEATS AND CHORDS IN REAL TIME

Miyoko Tsumaki[1*] **Masanobu Yamato**[2*] **Keigo Wasa**[2*] **Kenya Ikeda**[3*]
Yoshiaki Bando[4] **Misato Ohkita**[4] **Katsutoshi Itoyama**[4] **Kazuyoshi Yoshii**[4]
[1]Doshisha Girls' Senior High School, Japan [3]Nada Senior High School, Japan
[2]Kinki University Wakayama Senior High School, Japan [4]Kyoto University, Japan
{yoshiaki,ohkita,itoyama,yoshii}@kuis.kyoto-u.ac.jp

ABSTRACT

This paper presents a humanoid robot capable of singing and dancing to a song in an improvisational manner while recognizing the beats and chords of the song in real time. Among various kinds of entertainment robots that are expected to live with humans in the future, music robots such as robot dancers and singers are considered as one of the most attractive applications of music analysis techniques. Our robot mainly consists of listening, dancing, and singing functions. The listening function captures music audio signals and recognizes the beats and chords in real time. The dancing function switches dancing movements according to the types and root notes of the estimated chords. The singing function, on the other hand, generates singing voices whose pitches change according to the root notes of the chords. The information on beats and chords are exchanged between the three functions. The preliminary experiment showed the great potential of the proposed dancing robot. We plan to improve the response of dancing and singing functions by predicting next chords.

1. INTRODUCTION

Development of entertainment robots that can interact with humans through music is an important research direction in the field of music information retrieval (MIR). Since various kinds of robots are expected to get into our daily lives in the future, not only task-oriented robots but also entertainment robots that people feel familiarity with have been developed, *e.g.*, a violinist robot [3] and a flutist robot that can play the flute in synchronization with a melody played by a human [5]. Since dancing is a universal form of expression seen in many cultures, in this paper we focus on music robots that can dance interactively with humans. Although many researchers have tackled music signal analysis for content-based music selection (retrieval and recommendation), we aim to enhance a user's experience of listening to selected musical pieces by leveraging the analyzed contents for music robots [4].

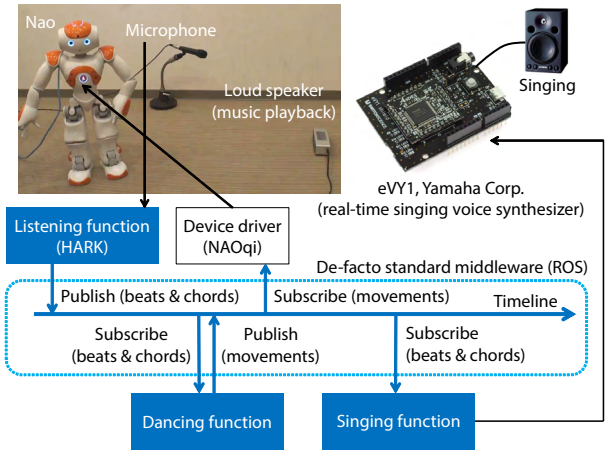

Figure 1. System architecture of a singing robot dancer.

A robot that can dance synchronously with music needs to adaptively control its movements while recognizing the content of music. Several robot dancers have already been developed. Murata *et al.* [4], for example, enabled a bipedal humanoid robot to step and sing in synchronization with musical beats, Kosuge *et al.* [2] devised a dance partner robot that can predict the next step intended by a human dancer, and Kaneko *et al.* [1] developed a humanoid robot that can generate natural dancing movements by using a complicated human-like dynamical system. In contrast to these robots, our robot is capable of singing and dancing to *any* musical pieces in an improvisational manner by recognizing the beats and chords in real time instead of using musical score information.

2. PROPOSED SYSTEM

This section explains the internal architecture of the proposed singing robot dancer (Figure 1). Our system mainly consists of listening, dancing, and singing functions that are communicated with each other in an asynchronous manner through data streams managed by the Robot Operating System (ROS). The listening function, which is implemented on an open-source robot audition software called

*The four high school students contributed equally. This study was supported by the Global Science Campus project, JST, Japan.

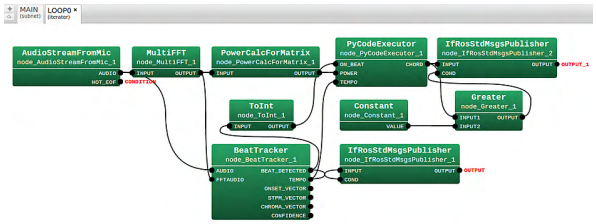

Figure 2. Visual programming interface of HARK.

HARK, takes music audio signals captured by a microphone and recognizes the beats and chords of those signals in real time. The dancing function then receives the recognition results and then determines dancing movements. The singing function also receives the recognition results, determines vocal pitches and onsets, and synthesize singing voices by using a singing-voice synthesizer called eVY1, Yamaha Corp. (MIDI device).

2.1 Listening function

The listening function mainly consists of two modules: beat tracking and chord estimation, which are performed sequentially on the dataflow-type visual programming interface of HARK (Figure 2).

Beat tracking This module, which is included in HARK, is based on an efficient beat tracking method called spectro-temporal pattern matching (STPM) [4]. This method extracts "edges" (spectral components with rapid power increase) from a music spectrogram by using Sobel filters. This is a standard technique of edge extraction in image processing. The tempo and beat times are estimated by calculating the autocorrelation of those components.

Chord estimation This module classifies 12-dimensional beat-synchronous chroma vectors extracted from the music spectrogram into 24 kinds of chords (12 root notes $\times$ 2 types (major/minor)) by using a template matching method based on the cosine distance.

2.2 Dancing function

The dancing function concatenate dancing movements according to the chord progression of the target musical piece. We defined 24 different dancing movements corresponding to the 24 kinds of chords (Figure 3). In fact, a proprietary device driver called NAOqi should be linked to the ROS for controlling the movements of the robot.

2.3 Singing function

The singing function controls the eVY1 device for generating beat-synchronous singing voices whose pitches match the root notes of the estimated chords. eVY1 can be controlled in real time as a standard MIDI device.

3. EXPERIMENT

We conducted an experiment using a sequence of simple chords (toy data) and a Japanese popular song (real data) in a reverberant room. Each of the audio signals were played back from a loudspeaker. The audio signals were captured by using a microphone behind the robot. The distance between the loudspeaker and the microphone was about 1 m.

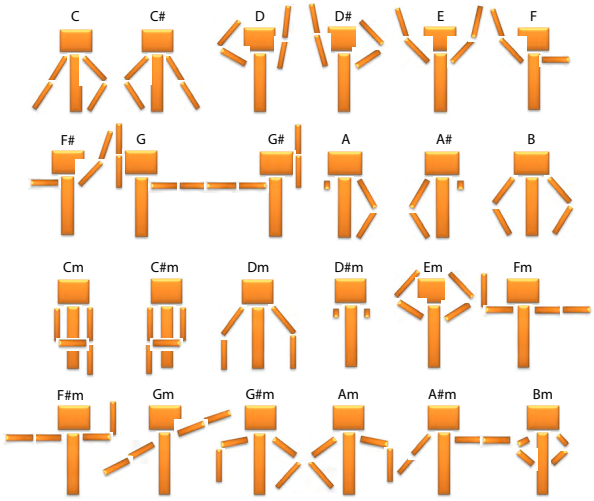

Figure 3. Predefined dancing movements.

The experimental results showed that our robot has great potential as an entertainment robot. It could recognize the chord progressions of both data to some extent and generate chord-aware beat-synchronous dancing movements. The response of dancing and singing, however, was delayed for two beats after new chords began because the robot has no function of chord prediction. Demos are at http://winnie.kuis.kyoto-u.ac.jp/members/yoshii/gsc2015/

4. CONCLUSION

This paper presented a singing robot dancer that can recognize the beats and chords of music audio signals in real time. Our humanoid robot was developed by combining an open-source robot audition software called HARK, signing-voice synthesis hardware called eVY1, and a robot-motion controller in an asynchronous manner through a standard middleware called ROS. Although the experimental results showed the potential of the proposed system, we found that the response of singing and dancing should be improved by implementing real-time prediction of next chords. To enable the robot to use its own ears embedded in the body to capture audio signals, we plan to improve the robustness of the listening function by using dereverberation and ego-noise cancellation techniques.

5. REFERENCES

[1] K. Kaneko *et al.* Cybernetic Human HRP-4C. *Humanoids*, 2009.

[2] K. Kosuge *et al.* Partner Ballroom Dance Robot — PBDR—. *SICE Journal of Control, Measurement, and System Integration*, 1(1):74–80, 2008.

[3] Y. Kusuda. Toyota's Violin-playing Robot. *Industrial Robot*, 35(6):504–506, 2008.

[4] K. Murata *et al.* A Beat-Tracking Robot for Human-Robot Interaction and Its Evaluation. *Humanoids*, 2008.

[5] K. Petersen *et al.* Development of a Aural Real-Time Rhythmical and Harmonic Tracking to Enable the Musical Interaction with the Waseda Flutist Robot. *IROS*, 2009.