The 2009 IEEE/RSJ International Conference on
Intelligent Robots and Systems
October 11-15, 2009 St. Louis, USA

Development of a Aural Real-Time Rhythmical and Harmonic Tracking to Enable the Musical Interaction with the Waseda Flutist Robot

Klaus Petersen, Jorge Solis, Member IEEE, and Atsuo Takanishi, Member IEEE

Abstract— **The Waseda Flutist Robot is able to play the flute at the level of an intermediate human player. This ability opens a wide field of possibilities to research human-robot musical interaction. This research is focused on enabling the flutist robot to interact more naturally with musical partners in the context of a Jazz band. For this purpose a Musical-Based Interaction System (MbIS) has been proposed to enable the robot to process both visual and aural cues coming throughout the interaction with musicians. In a previous publication, we have concentrated on the implementation of visual communication techniques. We created an interaction interface that enabled the robot to detect instrument gestures of partner musicians during a musical performance. Two computer vision approaches were implemented to create a two-skill-level interface for visual human-robot interaction in a musical context. In this paper we focus on the aural perception system of the robot. The method introduced here enables the robot to, a suitable environment provided, detect the tempo and harmony of a partner musician's play, with a specific focus on improvisation. We achieve this by examining the rhythmical and harmonic characteristics of the recorded sound. We apply the same approach to amplitude and frequency spectrum, thus, in the former case tracking amplitude transients. In the latter case, as we focus on communication with monophonic woodwind instruments, we follow the most prominent peak in the frequency spectrum. We specifically use a similar technique for the audio analysis as we did for our previous research on motion tracking. From the experimental results, we have shown that after implementing our algorithm the robot is able to correctly recognize a number of rhythms and harmonies. It is able to engage in a simple form of *stimuli and reaction* play with a human musician.**

I. INTRODUCTION

The research on the Waseda Flutist Robot; since 1990, has been carried out as an approach to understand the human motor control from an engineering point of view as well as introducing novel ways of musical teaching [1]. In particular, we have been focused on improving the mechanical design of the lungs, vocal cord, mouth, etc. as well as the implementation of advanced control strategies [2]. Moreover, some of the perceptual capabilities have been implemented such as automatic melody recognition [3], human face tracking [4], etc. As a result of our research, the latest version of the flutist robot, the Waseda Flutist Robot No.4 Refined IV (WF-4RIV) is able of playing the flute nearly similar to the performance of an intermediate flutist.

With the mechanical capabilities of the robot having reached a satisfactory level we have in the past two years concentrated on developing an interface that allows for the interaction of the flutist robot with human musicians. We present the results of various experiments we performed to confirm the functionality of our method. Besides analyzing the technical performance of our algorithms, we also scrutinize the resulting sound output and try to determine how well our robot dynamically changes musical parameters while interacting with a human player. As we will see in the experimental results section, the physical constraints of the robot play an important role here. Although they are first of all a limitation of the capabilities of the robot, they on the other hand make the interaction experience feeling more natural, human-like for the user.

We observed that in general the two principal ways of a human musicians to interact with each other during a performance are communication through the acoustic and visual channel. Although aural exchange of information seems predominant in a musical band setup there is also a large amount of communication taking place through visual interaction. Authors in previous publications ([5], [6]) have examined methods to visually track the movements of the instrument of a musician performing together with the robot. We based this research on the scenario of improvisation in a Jazz band. During solo play, in most cases, one player at a time takes the lead and the other players provide accompaniment. Upon finishing a solo, through movements with his instrument one player directs the lead to the next person. Experiments to evaluate the functionality of our tracking methods were performed, allowing us to scrutinize the quality of the interaction. Two different levels of interaction were proposed, one to suit the requirements of a beginner musician and one for an advanced inter-actor. The first level used so called virtual faders and buttons as human interface devices for controlling the robot‘s performance. In the second level a particle filter-based tracking algorithm was used to allow the robot to recognize changes in instrument orientation.

II. MUSICAL INTERACTION SYSTEM

The robot is equipped with sensors that allow it to acquire information about its environment. As the robot is a humanoid we emulate two of the human's most important perceptual organs: the eyes and the ears. We integrated two miniature video cameras in the head mechanism of the robot.

Jorge Solis is with Waseda University, Department of Mechanical Engineering and researcher at the Humanoid Robotics Institute (HRI), Waseda University, 3-4-1 Ookubo, Shinjuku-ku, 169-8555. Tokyo, Japan (phone: +81-3-5286-3257; fax:+81-3-5273-2209; e-mail: solis@kurenai.waseda.jp).

Klaus Petersen is with Waseda University, Graduate School of Advanced Science and Engineering, Waseda University, 3-4-1 Ookubo, Shinjuku-ku, 169-8555. Tokyo, Japan (phone: +81-3-5286-3257; fax:+81-3-5273-2209; e-mail: klaus@moegi.waseda.jp).

Atsuo Takanishi is with the Department of Mechanical Engineering, Waseda University and one of the core members of the Humanoid Robotics Institute, Waseda University (e-mail: takanisi@waseda.jp).

978-1-4244-3804-4/09/$25.00 ©2009 IEEE

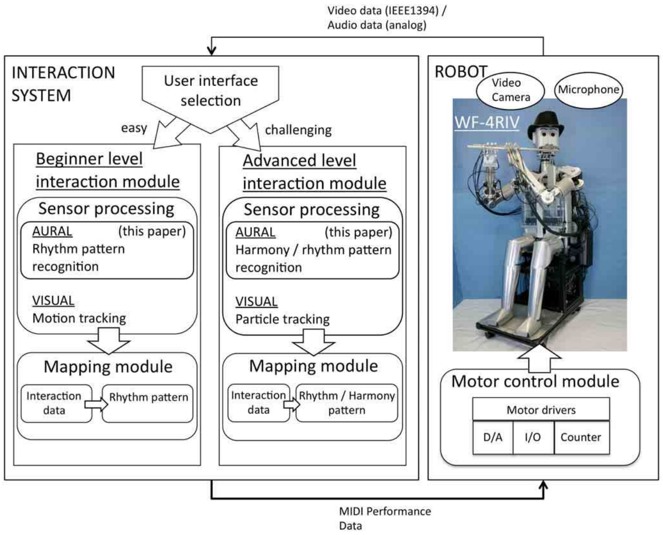

Fig. 1. On the left hand side the Musical-Based Interaction System (MbIS) is shown. The MbIS basically processes the information from the robot in the form of audio and video data. It processes this information employing the user‘s preferred interaction module. After mapping the result of the analysis to MIDI performance data, this information is transferred to the robot control module.

Two microphones are attached to the sides of the head for stereo-acoustic perception.

For this purpose, we proposed the Musical-Based Interaction System (MbIS) to allow for two levels of interaction (Figure 1). The purpose of the two level design is to make the system usable for people with different skill levels. Considering a situation of two human musicians intending to play together, the more advanced person would always have to adjust his way of interaction to the less advanced person. Even in case of having the same skill level two players need to get used to the way they interact and musically communicate with each other. We want to introduce the same kind of behavior for our humanoid robot. A person who has no experience in playing together with the robot will need more time to adjust to the particularities of this type of human-machine interaction. For that reason we designed the *beginner level* interaction system that provides easy-to-learn controllers having a strong resemblance to established studio equipment. Considering an advanced level player we want the robot to offer a way of interaction that satisfies more refined ways of creative expression. The *advanced level* interaction system thus allows for free control of the performance parameters.

Regarding the acoustic interaction we reduce the level of complexity by only using rhythmic ways of communication. The robot analyzes the timing of a tone sequence and reproduces this timing in its own performance. As the human player in this mode of interaction can concentrate on the rhythmic part of his play, without caring too much about the harmonic content, only the skill level of a beginner is required.

The advanced level of interaction requires more experience in working interactively with the robot, but also allows for more subtle control of the musical performance. In this mode we allow the user to additionally select a harmony for playing with the flutist robot. In other words, besides analyzing the musical content for rhythmic information, we also examine the tonal part of the sequence played by the human musician. The pattern being played by the robot as an answer contains the adapted rhythm and melody.

III. AURAL ANALYSIS INTERFACE DESCRIPTION

Various work has been done related to the field of aural human-robot interaction. Many of these developments involve speech recognition in order to be able to issue spoken commands to a robot [7]. Musical interaction with a robot that involves actual feedback to and from the robot has been studied, but not as extensively. [8] and [9] describe aural beat tracking systems, that enables their humanoid robots to perform dance-like motions synchronized to audio input. There is a number of robotic instruments (e.g. violin robot [10], piano robot in [11] or guitar robot in [12]), that are used to perform passively. That means that they perform a static score that during a performance is not influenced according to spontaneously developing musical context. Special interfaces to control the musical expression of a music robot have been presented in [13] and [14]. The cited articles present sensor-suits that allow the real-time alteration of performance parameters of various music robots. In previous research on the Waseda Flutist robot a system to enable the robot to teach flute playing to beginner students has been developed [4]. The implementation of an aural, musical human-robot interaction and performance system for a humanoid musician robot has so far very scarcely been studied.

In this paper, from an algorithmic point-of-view, we concentrate on real-time rhythm extraction and the real-time detection of harmonic structures within audio recordings. Up to now, extensive research has been done in this field. Various methods have been proposed to detect transients in musical data, a requirement for performing rhythmic analysis. Klapuri [15] uses division into frequency bands and linear regression to detect the starting point of a rhythmical feature. Division into high and low-energy peaks in addition to timing criteria [16] is applied to allow application on polyphonic sound data. Rao-Blackwellian Models [17], Online Onset Detection Models [18] and Brownian Motion Models [19] are commonly used to extract tempo and structural information.

A popular approach for harmonic analysis (phonic transcription) is the so called Blackboard Method. Originally not designed for audio analysis is has been applied in this area by various researchers ([20], [21], [22], [23], [24], [25], [26]). The method is somewhat similar to the algorithm described here as it also uses a ‘knowledge source‘ (which in our case consists of a library of possible melody patterns) to reinforce recognition. Multiple model based techniques have been proposed using approaches ranging from spectral template matching ([27], [28]) to sequential Monte Carlo methods ([29], [30], [31]).

The idea of using histograms for characterizing audio material has been used mainly in the context of efforts to archive large amounts of musical data ([32], [33], [34]). Based on this previous research we try to exploit the technique as a simple method to characterize and match smaller pieces of harmonic and rhythmic information.

We chose our algorithmic approach among other options due to its practicability. Our interaction system consists of several modules containing not only an aural interface but also computer vision and a cognitive center. To keep all parts maintainable and the whole structure efficiently usable, we decided to use techniques that not necessarily represent the state-of-the-art of the respective area but integrate well with whole system. In that sense we chose the method presented here not as a separate approach in aural analysis but rather in the context of the whole interaction system as a whole.

A. Rhythm Tracking (Beginner Level Interaction)

In this section we will describe the rhythm identification algorithm‘s principle of operation. Our purpose is to extract real-time,rhythmic information from the recorded sound data. The analysis result is matched with a library of timing patterns that are saved as previous knowledge in the robot. The algorithm determines the best matching pattern and passes this information on to the mapping module, in order to generate an output performance by the robot.

The performance of an instrumentalist is recorded and the data directly streamed to the analysis algorithm. In the beginner level interaction mode, because we are interested in the rhythm information contained in the acquired data, we examine it for timing characteristics. In the sound waveform separate notes are represented as distinguishable amplitude peaks. We isolate these peaks by defining a threshold, as it is shown in Eq. (1).

$$a_t = \begin{cases} 0 & \text{if } i_t \leq m \\ i_t & \text{if } i_t > m \end{cases} \tag{1}$$

a_t: thresholded sound wave value
m: threshold level
i_t: input sound wave

The duration of one tone impulse naturally is longer than a certain minimum time. In order to prevent very short noise peaks from falsely triggering the threshold we smoothen the sound wave with a running average calculation (Eq. (2)). This computation acts, from a signal processing point of view, similar to a low-pass filter ([35]):

$$p_r = \alpha * p_p + (1 - \alpha) * p_c \tag{2}$$

p_r: average for the resulting pixel
p_p: pixel at the same position in the previous difference image
p_c: same pixel in the current image
α: averaging factor

The rhythm patterns have a certain length. To identify the most recently played pattern we do not need to analyze all of the previous sound input. We rather use a window that always contains only the most up-to-date part of the recorded music information. This window continuously slides forward as new data is acquired. The size of the window is the length of the longest rhythm pattern in the robot‘s pattern library. Regardless which pattern is currently played by the interacting musician, it will always completely fit inside the window.

Each positive edge of the threshold sound wave in the time window represents a rhythmic pulse. To characterize the timing of this sequence of pulses as a whole we calculate the time differences between adjacent pulses. Utilizing this information we can construct a histogram, with one bin representing one certain time difference. Both axis of the histogram are normalized, with the result that the maximum and minimum bin of the histogram always relates to maximum and minimum pulse delta time. This histogram is then compared to the histograms of the timing patterns in the library of the robot. The similarity between two histogram is determined using the Bhattacharyya [36] coefficient (Eq. (3)).

$$\rho[p^i, q] = \sum_{u=1}^{m} \sqrt{p^i_u q_u} \tag{3}$$

with p^i being the histogram of one library pattern, q resembling the sampled rhythm pattern and m expressing the histogram size. The sum is indexed by u.

To prevent patterns from being falsely detected we apply a threshold to the similarity coefficient. If the result of the pattern comparisons falls below this threshold the robot does not recognize the input as a known rhythm. The result of the rhythmic analysis is the best match from the rhythm pattern library.

B. Harmony Tracking (Advanced Level Interaction)

Our approach for the aural processing of the advanced level interaction aims to create an interface that extends the amount of freedom a player has in controlling the robot. At the same time the usage of the system becomes more demanding for the human player in terms of skill level. However, it also allows a wider scope of musical expressiveness. This advanced level approach, additionally to extracting timing information, analyzes the harmonic components of the recorded sound. The method for recovering the rhythmic context is the same as the one used for the beginner level interaction. In the following we describe the unique part of the system, the harmonic component analysis. An overview of the structure of the system can be seen in Figure 2

Pitch information is recovered from the input data stream by applying a discrete Fast Fourier Transformation (FFT). As we sample sound data in windows of 1024 samples we apply the Hann windowing function (Eq. (4)) to smoothen spectral leakage.

$$w_i = a_i 0.5 \left(1 - \cos\left(\frac{2\pi i}{N-1}\right)\right) \tag{4}$$

w_i: resulting amplitude for sample i
a_i: input amplitude indexed with i
N: number of samples in the window

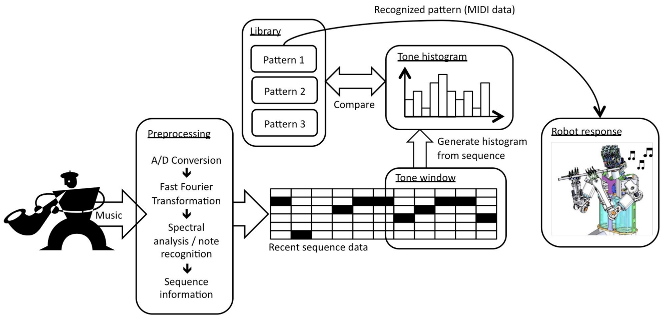

Fig. 2. Principle of harmony recognition in the advanced interaction level. Sound data is recorded from the instrumentalist by the microphone. After the spectral analysis a melody histogram is created. Comparing this histogram with the harmony library, the best match is selected.

Similar to the timing analysis, we apply a running average to adjacent frequency spectra and perform thresholding operations to reduce noise. If the threshold amplitude is retained by one or more peaks of the spectrum for long enough not to be suppressed by the low-pass filter, the peak with the highest amplitude is identified as lead-frequency. A recently registered pitch frequency is approximated by the twelve-tone system note with the closest frequency. The value of this note is queued into the sequence window.

When looking for harmonic information we look into the past only for the number of notes contained in the longest library pattern. The note information in the sequence window is gathered by generating a histogram from the pitch values. Again we match this histogram to the library histogram in order to find the best match. Information regarding which pattern was recognized is then forwarded to the mapping module.

C. Mapping

In our vision approach we mapped the result values of the computer vision analysis to control parameters like song tempo or vibrato amplitude. The gestures performed by the interacting musician do by themselves not have an influence on the music (as they do not produce any sound), so we can create a direct connection between instrument movement and the parameter to modulate without producing noise just by the means of information transmission. In contrast, given the case of intending to control the robot through music, the modulation data needs to be extracted from the actual musical content. Given the aural analysis that we perform now, the resulting information identifies at least a part of the musical intention of the partner instrumentalist. The robot is to use this data to complement the performance with his own play in compliance with the present rhythm and harmony.

The knowledge about which timing and harmony pattern is being currently used, can be utilized to create active feedback to the immediate song context. For that reason we mapped rhythmic as well as timing recognition to trigger one-to-one corresponding patterns from the robot‘s pattern library. The intended result is that the robot echos the performance of partner.

IV. EXPERIMENTS AND RESULTS

The purpose of our experiments is to show how well a user can express his musical intention using the provided interaction setup. The interaction system itself resembles a closed control loop, with the robot on one side and the human musician on the other side. The fact that decides about the quality of the output is how responsive the robot is to the actions of its interaction partner. In the same way, naturally it also depends on the skill level of the human player. To accommodate for these parameters our interaction system is separated into beginner level interaction module and advanced level interaction module. For each of these, we propose separate experiments. Although we examine the two modules in different experiments we in principle use the same setup and evaluation method for both.

In case of the beginner level interaction interface, the robot is programmed with a library of three rhythm patterns. We chose these patterns specifically to be easily discernible to prevent false recognition. For the current state of our research our purpose is to verify the functionality and also practicability of the system as such. As we will write in more detail in the future works section we consider several approaches to enhance our recognition technique.

The experiment itself consist of a human saxophone player situated in front of the robot at a distance of about 2 meters. The human musician has knowledge about the three patterns that are contained in the robot‘s library. With the start of the experiment, he will begin to deliberately play one of these rhythm patterns on a single note. He will repeat playing this rhythm until the robot responds. After that he will, again randomly, choose the next pattern. This procedure is repeated for several times. The responses of the robot as well as the play by the musician that triggered a response are recorded. The two recorded sound waves are examined using a FFT spectral analysis algorithm to find pitch and amplitude of the music data. Looking at the resulting graph we have the possibility to analyze the robot‘s choice of a response. Also the time-relationship of input and output are to be examined. The quality of the response can be characterized by how quickly (time difference between musician‘s first complete play of the pattern and the robot‘s response) it is produced and how accurately (compliance of the response timing pattern with the input pattern).

We perform the same method of experimentation for evaluating the aural recognition algorithm of the advanced level interaction system. In this case additionally to the three rhythm patterns, three harmony structures are contained in the robot‘s library. During the experiment, the saxophone player performs deliberate combinations of one timing and one harmony pattern. The human musician‘s play and the recognition response of the robot are recorded and analyzed in the same way as in the previous experiment. We chose all patterns to have the same length of 1 bar. This makes the recognition of the end of one pattern more easy. As can be examined in Fig. 3a, Pattern 1 consists of one bar of four equally spaced quarter notes. The instrumentalist plays

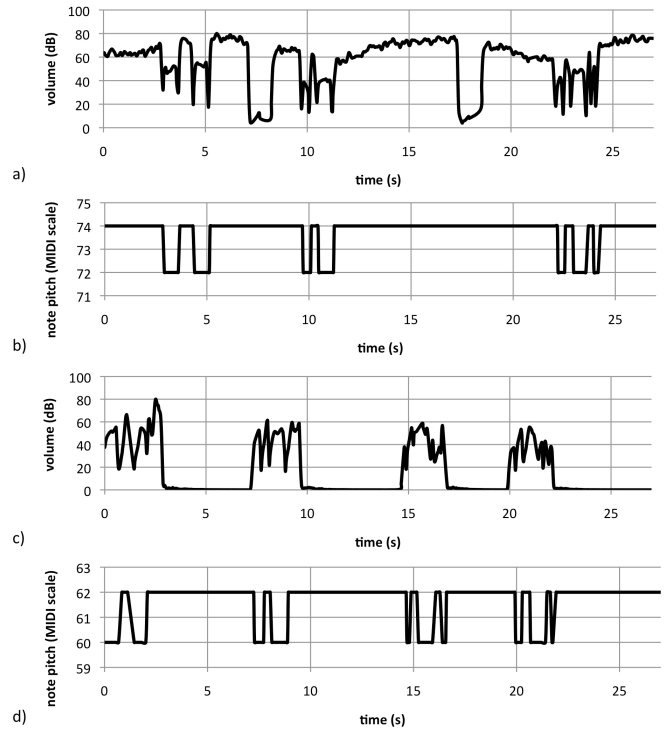

Fig. 3. Recorded input and output of the *beginner level* interaction system. a) is the amplitude plot of the flute robot response, b) the pitch analysis of this response, c) the amplitude plot of the *question* by the robot‘s partner musician and d) the pitch analysis of the musician‘s phrases. Because the flutist robot plays legato notes, alteration between the notes C and D was pre-programmed.

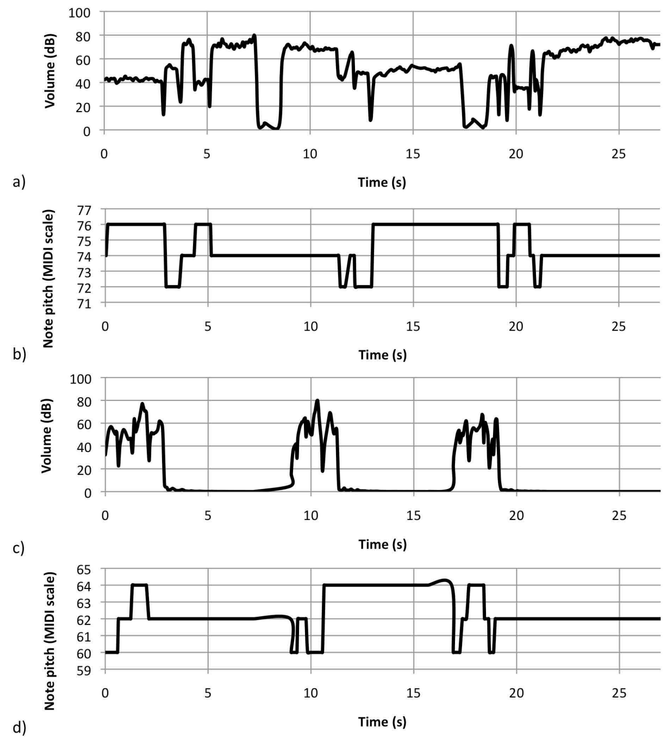

Fig. 4. Recorded input and output of the *advanced level* interaction system. As in the previous figure, a) is the amplitude plot of the flute robot response, b) the pitch analysis of this response, c) the amplitude plot of the *question* by the robot‘s partner musician and d) the pitch analysis of the musician‘s phrases.

this pattern alternating between the notes C and D at a tempo of 80bpm. After the first pattern has been performed (from $0s$ to $3s$), the robot answers by reproducing the same pattern (in beginner level interaction mode pre-programmed to alternate between C and D). The duration between the beginning of the instrumentalist‘s *question* and the robot‘s *answer* is approximately $3s$. The robot continues playing the last tone of the rhythm pattern until the next successful recognition. During this time the robot is in an *idle* state. The purpose of playing the last tone of a pattern continuously, is to give the interacting partner of the robot a feeling of the breathing cycles of the robot. When we look at the volume plot we can identify areas, where the level suddenly drops for a certain duration. These moments are called *breathing points* and relate to the time when the robot‘s lung system is deflated and needs to pull air in order to be able to produce the air-beam necessary to generate the flute sound. As the lung breathing speed is constant we see these events regularly happening at $7.5s$ and $14s$ in the graph. The duration of one breathing phase is $\approx 10s$ long. At $7.0s$ the musician starts playing rhythm pattern 2. This pattern is more complicated than the first one, consisting of one bar, containing two eighth-notes and two quarter notes. Again, this rhythm pattern is played without consideration to harmony, as variation of notes C and D. The answer of the robot is observed after the pattern has been finished, at $\approx 9.5s$. The duration between the last tone of the question pattern and the robots answer is $\approx 0.1s$. The third rhythm pattern is the most complicated of the three sequences used in this experiment. Besides a quarter and an eighth note it contains a triplet. The instrumentalists plays it two times, at $14.5s$ and $19.8s$. When played for the second time it is correctly recognized by the aural processing algorithm and the robot gives a response at $24.5s$.

For the advanced level interaction we used the same setup of microphones as for the previous experiment. Now the robot‘s library contains, besides the three rhythm patterns we already used for the previous experiment, three melody structures. The instrumentalist can now freely choose a combination of one of the rhythms and one of the harmonies as a question pattern. A graph of the results of the advanced interaction level aural recognition experiment is shown in Fig. 4. The following rhythm pattern / melody pattern combinations are chosen as *questions* to the robot:

- Combination 1: rhythm 1 / melody 1
- Combination 2: rhythm 2 / melody 2
- Combination 3: rhythm 3 / melody 3

The first combination pattern is played at $0s$. The further two patterns follow at $11.5s$ and $17.5s$. The robot generates an answer for each of these questions, reproducing the rhythm and melody combination that has been chosen by the musician ($3.5s$, $11.5s$ and $19.5s$). Again, in the volume plot of the output of the robot we observe the artificial lung‘s breathing rhythm. Breathing points are registered at $7.5s$ and $17.5s$. The break in audio output takes a duration of $\approx 1s$.

V. CONCLUSIONS AND FUTURE WORK

We have shown that after implementing our algorithm the robot is able to correctly recognize a number of rhythms and harmonies. It is able to engage in a simple form of *question and answer* play with a human musician. In case a pattern fails to pass the similarity threshold level that was described earlier, it will not be recognized. If this threshold is high, the human musician has to be very precise in his performance. If the threshold is low, false positive recognitions can occur. We experimentally chose a threshold that in most cases lead to satisfying results. We documented results for both, the beginner level and the advanced level interaction system. The results show, that, given certain circumstances during the experimentation, our method does lead to the intended outcome. This justifies further research in this direction, based on the development we explained in this paper. Several parts of the method leave room for improvement and we plan to address these problems in future research.

REFERENCES

[1] J. Solis, K. Chida, K. Suefuji, and A. Takanishi, "The development of the anthropomorphic flutist robot at waseda university," *Int. Journal of Humanoid Robots*, vol. 3, pp. 127–151, 2006.

[2] J. Solis, K. Taniguchi, T. Ninomiya, and A. Takanishi, "Understanding the mechanisms of the human motor control by imitating flute playing with the waseda flutist robot wf-4riv," *Mechanism and Machine Theory Journal*, vol. 44(3), pp. 527–540, 2008.

[3] J. Solis, S. Isoda, K. Chida, A. Takanishi, and K. Wakamatsu, "Anthropomorphic flutist robot for teaching flute playing to beginner students," *Proc. of the IEEE International Conference on Robotics and Automation, pp. 146-150*, 2004.

[4] J. Solis, K. Suefuji, K. Taniguchi, and A. Takanishi, "Towards an autonomous musical teaching system from the waseda flutist robot to flutist beginners," *Conference on Intelligent Robots and Systems - Workshop: Musical Performance Robots and Its Applications*, pp. 24–29, 2006.

[5] K. Petersen, J. Solis, and A. Takanishi, "Development of a real-time instrument tracking system for enabling the musical interaction with the waseda flutist robot," *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 313–318, 2008.

[6] K. Petersen, J. Solis, and A. Takanishi, "Toward enabling a natural interaction between human musicians and musical performance robots: Implementation of a real-time gestural interface," *The 17th IEEE International Symposium on Robot and Human Interactive Communication*, pp. 340–345, 2008.

[7] S. Yamamoto, J. Valin, K. Nakadai, J. Rouat, F. Michaud, T. Ogata, and H. Okuno, "Enhanced robot speech recognition based on microphone array source separation and missing feature theory," *Conference on Robotics and Automation (ICRA)*, pp. 1477–1482, 2005.

[8] M. Michalowski, S. Sabanovic, and H. Kozima, "A dancing robot for rhythmic social interaction," *Conference on Human Robot Interaction*, pp. 89–96, 2007.

[9] K. Yoshii, K. Nakadai, T. Torii, Y. Hasegawa, H. Tsujino, K. Komatani, T. Ogata, and H. Okuno, "A biped robot that keeps steps in time with musical breats while listening to music with its own ears," *Conference on Intelligent Robots and Systems*, pp. 1743–1750, 2007.

[10] K. Shibuya, "Analysis of human kansei and development of a violin playing robot," *Conference on Intelligent Robots and Systems - Workshop: Musical Performance Robots and Its Applications*, pp. 13–17, 2006.

[11] E. Hayashi, "Development of an automatic piano that produce appropriate touch for the accurate expression of a soft tone," *Conference on Intelligent Robots and Systems - Workshop: Musical Performance Robots and Its Applications*, pp. 7–8, 2006.

[12] E. Singer, K. Larke, and D. Bianciardi, "Lemur guitarbot: Midi robotic string instrument," *Conference on New Interfaces for Musical Expression*, pp. 3188–3191, 2003.

[13] S. Goto and F. Yamasaki, "Integration of percussion robots 'robotmusic' with the data-suit 'bodysuit': Technological aspects and concepts," *IEEE International Conference on Robot and Human Interactive Communication*, pp. 775–779, 2007.

[14] S. Jorda, "Afasia: the ultimate homeric one-man-multimedia-band," *Conference on New Instruments for Musical Expression*, pp. 201–206, 2002.

[15] A. Klapuri, "Automatic transcription of music," Master's thesis, Audio Research Group, University of Tampere, Finland, 1998.

[16] S. Dixon, "Monte carlo methods for tempo tracking and rhythm quantization," *Journal of New Music Research*, vol. 30, pp. 39–58, 2001.

[17] A. Cemgil and B. Kappen, "Monte carlo methods for tempo tracking and rhythm quantization," *Journal of Artical Intelligence Research*, vol. 18, pp. 45–81, 2003.

[18] G. Monti and M. Sandler, "Automatic polyphonic piano note extraction using fuzzy logic in a blackboard system," *Proceedings Digital Audio Effects Workshop (DAFx)*, pp. 39–44, 2002.

[19] J. Bello, "Towards the automated anaylsis of simple polyphonic music: A knowledge based approach," Ph.D. dissertation, Queen Marys University, London, 2003.

[20] K. Martin, "Automatic transcription of simple polyphonic music: Robust front end processing," *3rd Joint Meeting of the Acoustical Societies of America and Japan*, pp. 101–111, 1996.

[21] D. Martin, "A blackboard system for automatic transcription of simple polyphonic music," Media Laboratory, MIT, Tech. Rep., 1996.

[22] D. Godsmark and G. Brown, "A blackboard architecture for computational auditory scene analysis," *Speech Communication*, vol. 27, pp. 351–366, 1999.

[23] J. Bello, G. Monti, and M. Sandler, "An implementation of automatic music transcription of monophonic music with a blackboard system," *Proceedings of the Irish Signals and Systems Conference*, pp. 217–223, 2000.

[24] J. Bello and M. Sandler, "Blackboard system and top-down processing for the transcription of simple polyphonic music," *Proceedings Digital Audio Effects Workshop (DAFx)*, pp. 55–64, 2000.

[25] G. Monti, "Signal processing and music analysis," Department of Electronic Engineering, Queen Marys, University London, Tech. Rep., 2000.

[26] M. Plumbley, S. Abdallah, J. Bello, M. Davies, G. Monti, and M. Sandler, "Automatic music transcription and audio source separation," *Cybernetics and Systems*, vol. 33, pp. 603–627, 2002.

[27] K. Kashino and H. Murase, "Music recognition using note transition context," *Proceedings ICASSP*, vol. 6, pp. 3593–3596, 1998.

[28] K. Kashino and H. Murase, "A sound source identication system for ensemble music based on template matching and music stream extraction," *Speech Communication*, vol. 27, pp. 337–349, 1999.

[29] P. Walmsley, "Signal separation of musical instruments - simulation-based methods for musical signal decomposition and transcription," Ph.D. dissertation, Cambridge University Engineering Department, 2000.

[30] P. Walmsley, S. Godsill, and P. Rayner, "Bayesian graphical models for polyphonic pitch tracking," *Diderot Forum*, pp. 1–26, 1999.

[31] P. Walmsley, S. Godsill, and P. Rayner, "Multidimensional optimisation of harmonic signals," *Proceedings EUSIPCO*, pp. 2033–2036, 1998.

[32] G. Tzanetakis and P. Cook, "Musical genre classification of audio signals," *IEEE Transactions on Speech and Audio Processing*, vol. 10, pp. 293–302, 2002.

[33] G. Tzanetakis and P. Cook, "A quick search method for audio and video signals based on histogram pruning," *IEEE Transactions on Multimedia*, vol. 5, pp. 348–357, 2003.

[34] G. Tzanetakis, A. Ermolinskyi, and P. Cook, "Pitch histograms in audio and symbolic music information retrieval," *Journal of New Music Research*, vol. 32, pp. 143–152, 2003.

[35] C. Wren, A. Azarbayejani, T. Darrell, and A. Pentland, "Pfinder: Real-time tracking of the human body," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 780–785, 1997.

[36] K. Nummiaro, E. Koller-Meier, and L. V. Gool, "An adaptive color-based particle filter," *Journal of Image and Vision Computing*, vol. 21, pp. 99–110, 2003.