Anthropomorphic Musical Performance Robots at Waseda University: Increasing Understanding of the Nature of Human Musical Interaction

Jorge Solis
Waseda University, Faculty of Science and Engineering
Waseda University, Humanoid Robotics Institute
3-4-1 Ookubo, Shinjuku
Tokyo, Japan
solis@kurenai.waseda.jp

Takeshi Ninomiya, Klaus Petersen, Maasaki Takeuchi
Waseda University, Graduate School of Advanced Engineering and Science
3-4-1 Ookubo, Shinjuku
Tokyo, Japan

Atsuo Takanishi
Waseda University, Faculty of Science and Engineering
Waseda University, Humanoid Robotics Institute
3-4-1 Ookubo, Shinjuku
Tokyo, Japan
takanisi@waseda.jp

Abstract
During several decades, the research at Waseda University has been focused on developing anthropomorphic robots capable performing musical instruments. As a result of our research efforts, the Waseda Flutist Robot WF-4RIV and the Waseda Saxophonist Robot WAS-1 have been designed to reproduce the human player performance. As a long-term goal, we are proposing to enable the interaction between musical performance robots as well as with human players. In general the communication of humans within a band is a special case of conventional human social behavior. Rhythm, harmony and timbre of the music played represent the emotional states of the musicians. So the development of an artificial entity that participates in such an interaction may contribute to the better understanding of some of the mechanisms that enable the communication of humans in musical terms. Therefore, we are not considering a musical performance robot (MPR) just as a mere sophisticated MIDI instrument. Instead, its human-like design and the integration of perceptual capabilities may enable to act on its own autonomous initiative based on models which consider its own physical constrains. In this paper, we present an overview of our research approaches towards enabling the interaction between musical performance robots as well as with musicians.

Keywords: Anthropomorphic Robots, Musical Performance Robot, Wind Instruments, Musical Interaction.

1. Introduction

The relation between art and robots has a long history dating since the golden area of automata. As a result from the great efforts from researchers from both musical engineering and biomechanical engineering fields, nowadays we may distinguish two basic research approaches: developing human-like robots and developing robotic musical instruments [1-2].

The first approach, formally named Musical Performance Robots, is based on the idea of developing anthropomorphic robots capable displaying musical skills similar to human (from the point of view of intelligence and dexterity). The first attempt of developing an anthropomorphic musical robot was done by Waseda University in 1984. In particular, the WABOT-2 was capable of playing a concert organ. Then, in 1985, the WASUBOT built also by Waseda, could read a musical score and play a repertoire of 16 tunes on a keyboard instrument [2]. Prof. Kato argued that the artistic activity such as playing a keyboard instrument would require human-like intelligence and dexterity. Other examples can be found in [4-7]

From the second research approach, a robotic musical instrument is a sound-making device that automatically creates music with the use of mechanical parts, such as motors, solenoids and gears. By implanting algorithms of Musical Information Retrieval (MIR), the robotic musical instruments are simple mechanisms designed to embed sensors to analyze the human behavior and to provide physical responses on the actuated musical instrument. In other words, this approach may facilitate the introduction of novel ways of musical expression that cannot be conceived through conventional methodologies. A number of engineers and artist have made headway in this area. The art of building musical robots has been explored and developed by musicians and scientists such as [8-11].

More recently, few researchers have been focused on integrating basic perceptual modules to the musical performance robots in order to interact with human musicians. In particular, Singer et al. [11] developed the GuitarBot which it has been designed to create new of musical expression. In particular, their approach is based in developing robotic instruments that can play in way that humans can't or generally don't play. The instruments provide composers with an immediacy of feedback, similar to composing on synthesizers. However, as opposed to synthesizers, physical instruments resonate, project and interact with sound spaces in richer, more complex ways. All robotic instruments are controlled by custom developed MIDI hardware and software, based around PIC microcontrollers. Another example is the Haile developed by Weinberg et al. [13]; which is a robot designed to utilize autonomous behaviors that support expressive collaboration with human musicians. Haile is composed by a robotic arm that can hit the drumhead in different locations, speeds and strengths. The mechanism of the arm is reproduced by a sliding mechanism controlled by a solenoid. From the musical perceptual level, different Musical Information Retrieval algorithms have been implemented to modify the performance of Haile.

Even though GuitarBot and Haile are able to interact with musicians using conventional MIR algorithms, their physical mechanisms are too simple. However, if we want to understand in more detail the human while interacting in a musical way, we may require to increase the complexity of the mechanism of the musical performance robots as well as enhancing the perceptual capabilities of the robot (not only to process aural information but also visual, etc.) while considering physical constrains (such as breathing points, etc.).

Since 1990 at Waseda University we have been performing the research on musical performance robots. As a result, we have been developing an anthropomorphic robot that is capable of producing the flute sound similar to an intermediate player. In order to add expressiveness to the flute performance, we have implemented musical performance rules based on Neural Networks so that the robot can extract the musical content of the human player before doing the interaction. However, when we tried to perform experiment where the robot interacting in real-time with a human musician (band context), still the robot lacks of cognitive capabilities to process the coming musical information from the performance of the partner.

Up to now, several researchers have been providing advanced techniques for the analysis of human musical performance. However, in our case, we are talking about not just analyzing the human performance, but also we are required to map those musical parameters into control parameters of the robot. This means that we are also required to take into account the physical constraints of the robot. Due to the complexity of doing this task, we have proposed to continue our research based on two approaches: enhancing the cognitive capabilities of the Waseda Flutist Robot to process visual/aural information and developing a new musical performance robot such as a Waseda Saxophonist Robot. Therefore, as a long-term goal, we would like to enable the interaction between two human-like performance robots that are able to interact at the same level of perception as humans. From this, we may understand more in detail, from a scientific point of view, how humans can interact in musical terms. This may also contribute in finding new ways of musical expression that have been hidden behind the rubric of musical intuition.

In this paper we provide an overview of the current research achievements on the Waseda Flutist Robot WF-4RIV and the Waseda Saxophist Robot WAS-1. Then, a preliminary experiment is proposed to analyze the possibilities of interacting between both robots trough the MIDI communication.

2. Principles of Sound Production

In order to develop of human-like performance robots, we are required to understand in detail the principle of sound production of the instrument as well as the mechanism of humans to control different kinds of properties of the sound. In particular, in this section, we will provide a general overview of the differences on the principles of sound production between the flute and the saxophone. The flute is an air reed woodwind which only takes into consideration the width, thickness, angle, velocity of the air beam due to the absence of a reed (Figure 1a). Slight of any of these parameters are reflected in the pitch, volume, and tone of the flute sound. On the other hand, in the case of the saxophone, the sound is produced by controlling the differential pressure before the mouth piece (Uf) and after it (U). Such a differential pressure produces a vibration on the reed located inside the mouth piece so that a sound is produced (Figure 1b). Depending on the frequency of the vibration, the correct pitch of the sound is produced. Therefore, the pressure of the lips on the reed and the air beam coming from the lungs are important in order to control the pressure inside the mouth.

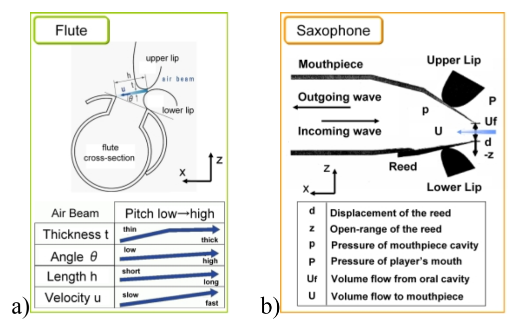

Figure 1. a) Principle of flute sound production [15]; b) Principle of saxophone sound production (single-reed) proposed by Schumacher [16].

3. Current research approach on the Waseda Flutist Robot

The research on the anthropomorphic Waseda Flutist Robot has been focused on mechanically reproducing the anatomy and physiology of the human organs involved during the flute playing [3]. In addition, we have focused on enabling the interaction with humans at the emotional level of perception. As a result of this research, in the 2008, the WF-4RIV has enhanced its musical performance thanks to the improvements on reproducing the lips and tonguing mechanisms (Figure 3). The WF-4RIV is composed by a total of 41-DOFs; which mechanically simulate the human organs involved during the flute playing. The WF-4RIV mechanically reproduces the anatomy and physiology of the following organs: lips (3-DOFs), neck (4-DOFs), lungs and valve mechanism (2-DOFs and 1-DOF respectively), fingers (12-DOFs), throat (1-DOF), tonguing (1-DOF), two arms (each with 7-DOFs) and eyes (3-DOFs). The WF-4RIV has height of 1.7m and a weight of 150kg. In particular, this new version has improved the mechanical design of the lips (to produce more naturally the shape of human lips so that more natural sounds can be produced), the tonguing mechanism (to reproduce the double tonguing so that smoother transitions between notes can be done), the vibrato, and the lung system.

In addition, research efforts have been done in order to enable the robot to enhance its expressiveness during the flute playing by implementing expressive performance rules by using Neural Networks. Moreover, an auditory feedback system (AFS) has been implemented on the WF-4RIV. The AFS enables the WF-4RIV to autonomously detect incorrect sounds produced during a performance and correct them. In particular, the proposed auditory feedback system is composed by three main modules (Figure 4): Expressive Music Generator (ExMG), Feed Forward Air Pressure Control System (FFAiPC) and Pitch Evaluation System (PiES).

The *ExMG* uses as an input the musical parameters (i.e. pitch, volume, tempo, etc.) from the performance of a professional flutist. Those parameters are analyzed and extracted by using our FFT tool [3]. As an output, a set of musical performance rules (which defines the deviations introduced by the performer) are produced (offline). The process of modeling the expressiveness features of the flute performance is done by using Neural Networks. The *FFAiPC* was implemented by a feed-forward control system to control the air pressure coming out from the lungs. Such an improvement is related to enable the robot to produce an attack time of the note more similar to the human. For this purpose, we compute the inverse model of the lung system to control of the air pressure during the attack time. The inverse model was computed by the feedback error learning. The feedback error learning is a computational theory of supervised motor learning proposed by Kawato [17]; which is inspired by the way the central neverous system. In addition, Kawato extended that the cerebellum, by learning, acquires an internal model of inverse dynamics of the controlled object.

The *PES* has been designed to detect both the pitch of the flute sound as well evaluation its quality. As a first approach, we have considered implementing the Cepstrum method. The Cepstrum is calculated by taking the Fourier transform (STFT) of the log of the magnitude spectrum of sound frame (time resolution is 23ms). In order to assure the accuracy of the pitch detection, the MIDI-data of the score was used to provide information to the pitch detection algorithm about where the pitch is supposed to be located. By tracing the peaks, we are capable of identifying the pitch of the note. After the detection of the pitch, we are then capable of evaluating the quality of the sound. Basically, the quality of the sound is determined, based on Ando's experimental results [15], by considering the relation among the harmonics structure content (Eq. 1).

$$Eval = \frac{(M - H) + (L_e - L_o)}{Volume} \tag{1}$$

M: Harmonic level [dB] *H*: Semi-Harmonics level [dB]
L_e: Even-harmonics level [dB] L_o: Odd-harmonics level [dB]
Volume: Volume level [dB]

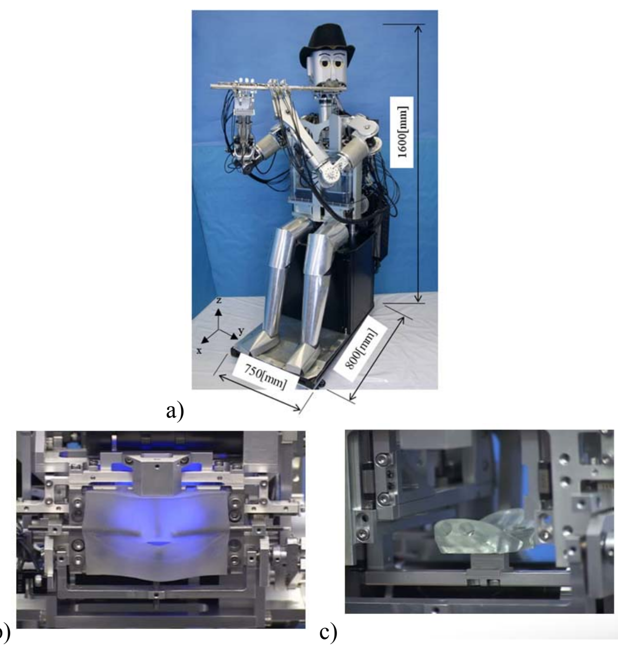

Figure 3. The Waseda Flutist Robot No.4 Refined IV (WF-4RIV) has been designed to play a classical flute: a) The Waseda Flutist Robot No.4 Refined IV (WF-4RIV) is composed by 41-DOFs, b) artificial lips; c) artificial tongue.

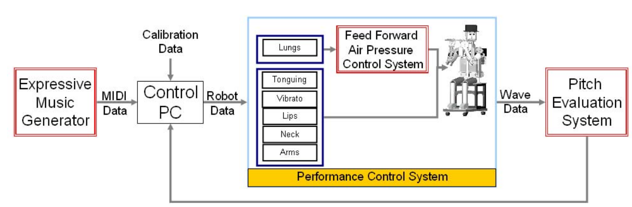

Figure 4. Diagram of the proposed Auditory Feedback System.

As the robot is a humanoid we emulated two of a human's most important perceptual organs: the eyes and the ears. We integrated two miniature video cameras in the head mechanism of the robot [13]. The viewing angle of these cameras can be controlled in three degrees of freedom. Furthermore two microphones are attached to the sides of the head for stereo-acoustic perception. Here, we concentrate on the visual interaction between a human instrument player and the flutist robot. Regarding this interaction we specifically consider the case that parameters of the performance of the robot are controlled by the options of an instrument. We consider that players do not only communicate acoustically but also visually, showing cue gestures to their fellow players at the end of a song part or an improvisation. More recently, advanced techniques of vision processing have been implemented to enhance the interaction with musical partners [18]. In particular, a novel human-robot interaction system for the WF-4RIV has been implemented. The proposed interaction system is based on the principle that musical performance parameters can be manipulated visually in real-time by instrument gestures of the interaction partner of the robot. Through this, natural and meaningful musical cooperation is facilitated. The tracking instrument algorithm has been implemented based on particle filter.

4. Current research approach on the Waseda Saxophonist Robot

In this year; we have developed the WAseda Saxophonist No.1 (WAS-1) which is composed by 15-DOFs required to play an alto saxophone (Figure 5a). The reason why we chose an alto saxophone, instead of the tenor saxophone, is due to its physical properties. The alto saxophone's height is 630mm and its weight is 2.4kg (in contrast, the tenor is 785mm and 3.3kg). In particular, the lips (1-DOF's lower lip), tongue (1-DOF), oral cavity, artificial lungs (1-DOF's air pump and 1-DOF's air flow valve) and fingers (11-DOFs) were developed.

The mouth of WAS-1 has been designed with 1-DOF that controls the motion of the lower lips (Figure 5b). The actuation of the lower lips enables the control of the threshold pressure and the production of vibrato. The lower lip is then connected to the artificial lips made of Septon; which is a thermoplastic rubber.

The artificial lips have been modeled by using Septon (Kuraray Co. Ltd.) which is an elastomer with high elasticity, high stiffness thermoplastic [19]. Such properties make possible the design of an artificial lip similar to the human lips in terms of shape and elasticity. In order to reproduce the motion of the lower lips, a T-shaped metallic pin (artificial tooth) has been embedded into the septon (Figure 5b). In contrast, even that a metallic pin has been embedded into the upper lips.

The oral cavity of WAS-1 has been also designed by using the septon. In addition, the strength of the oral cavity has been modelled to support pressures upper to 8kPa [20]. On the other hand, the tonguing mechanism is shown in Fig. 5c.

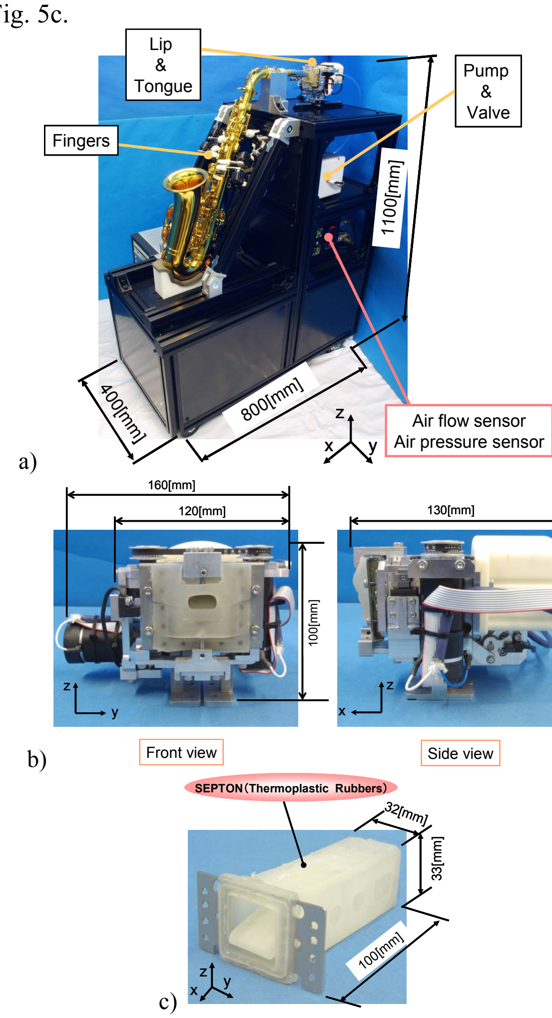

Figure 5. a) The WAseda Saxophonist Robot No.1 (WAS-1) was designed to play an alto saxophone. The WAS-1 is composed by a) artificial lungs and fingers, b) lips and c) oral cavity and tongue.

The motion of the tongue tip is controlled by a DC motor which is connected to a link attached to the motor axis. Thanks to this tonguing mechanism of the WAS-1, the attack and release of the note can be reproduced. Regarding the air source of WAS-1, a DC servo motor has been used to control the motion of the diaphragm of the air pump. By changing the rotational speed of the motor axis, the air flow quantity can be accurately controlled by measuring it with a flow meter (Figure 5a).

Moreover, a DC servo motor has been designed to control the motion of an air valve so that the delivered air by the air pump is effectively rectified. In order to enable the WAS-1 to play from C3 to C#5 (two octave fingering), 11-DOFs have been implemented. The musical performance control of WAS-1 is shown in Fig. 6.

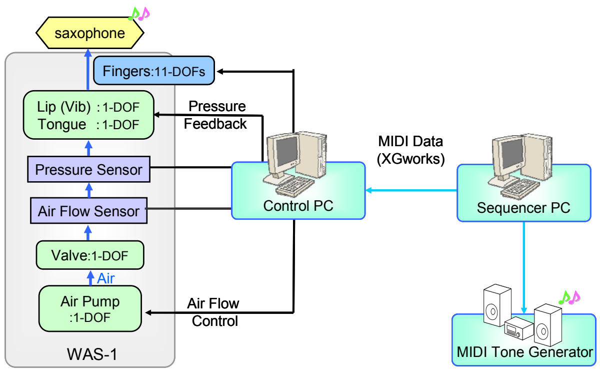

Figure 6. Musical performance control system of WAS-1

5. Experiments and Results

5.1 Evaluating the performance of WF-4RIV

Regarding the WF-4RIV, we have analyzed the flute sound compared with that produced by a professional flutist. For this purpose, we have used the Eq. (1) to compare their quality. The results are shown in Fig. 7. As we may observe, the flute sound quality of the WF-4RIV has been improved thanks to the mechanical improvements as well as the proposed auditory feedback system. In average, a 52% of improvement of the evaluation score was found while comparing the previous version of the flutist robot and the WF-4RIV.

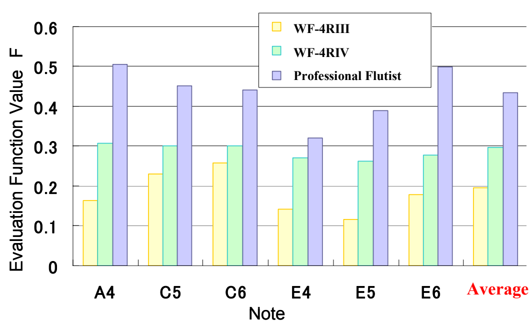

Figure 7. Experimental results while comparing the quality of the flute sound between the WF-4RIV and a professional flutist by using Eq. 1 (the previous version of the flutist robot WF-4RIII is also analyzed).

5.2 Evaluating the performance of WAS-1

Regarding the WAS-1, we have compared the pitch and volume produced by the WAS-1 and the intermediate level player. The experimental results are shown in Fig. 8. As we may observe the produced pitch by WAS-1 was quite similar to the human player (Fig. 8a); however, further improvements are still required to improve dynamic transitions between notes (Fig. 8b).

5.3 Preliminary Experiments: Duet Performance

In this preliminary experiment, we have focused in verifying the possibility of performing a duet between the WAS-1 (main voice) and the WF-4RIV (second voice). For this purpose, we have programmed both robots to perform the *Trois Duos de Mendelssohn et Lachner* composed by Felix Mendelssohn Bartholdy (Figure 9). In order to achieve the duet performance, as a preliminary approach, we have synchronized the performance of both robots by means of MIDI signal.

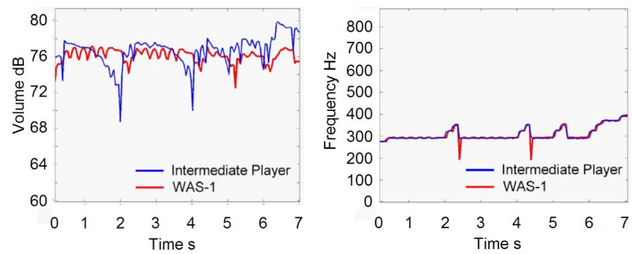

Figure 8. Experimental results while comparing the both performances: volume and pitch.

Figure 9. Duo between the WF-4RIV and WAS-1.

In this performance, we have recorded the performance of each robot separately by using two microphones. The recorded data was then analyzed by means of SSUM developed by Sturm et al. [21]. The SSUM is a tool to demonstrate essential principles and concepts of media signal processing to students.

The experimental results are shown in Fig. 10, where the pitch and volume of each performance during the duo are shown. As we may observe, both performance were synchronized and it is clearly observable the differences between the main and second voice by comparing the volume of the performance as well as the pitch pattern. However, in both cases still some difficulties are found while doing the dynamic transitions between notes, particularly during the breathing points of the robot. From these results, we may confirm the possibility of enabling the musical interaction between musical performance robots. In the near future, we hope we can enable a more natural interaction between both robots by means of processing not only the MIDI signal, but also the exchanged perceptual information among them.

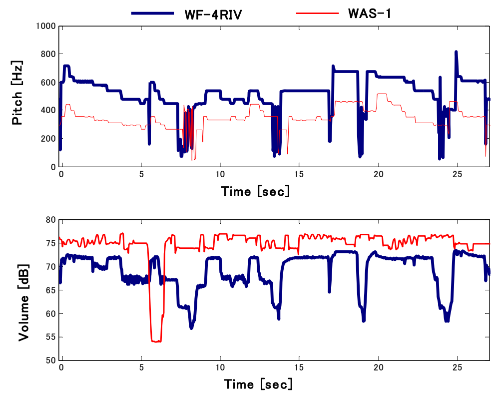

Figure 10. Experimental results while analyzing the duet performance of the WF-4RIV and WAS-1.

6. Conclusions and Future Work

In this paper, an overview of the research carried out at Waseda University towards enabling the interaction between musical performance robots has been detailed. In particular the efforts to enhance the perceptual capabilities of the Waseda Flutist Robot WF-4RIV and the development of the Waseda Saxophonist Robot WAS-1 have been detailed. A preliminary experiment has been presented to realize a duet between the WF-4RIV and WAS-1; where the musical performance was synchronized by means of MIDI data.

As a future work, we are planning to perform more detailed experiments to implement higher cognitive perceptual capabilities in both robots, so that in the future a natural duet performance between both robots can be done.

7. Acknowledgments

A part of this research was done at the Humanoid Robotics Institute (HRI), Waseda University.

References

[1] A. Kapur, "A History of Robotic Musical Instruments," in Proc. of the International Computer Music Conference, pp. 21-28, 2005.

[2] Solis, J., Takanishi, A. "An overview of the research approaches on Musical Performance Robots," in Proc of the International Computer Music Conference, pp. 356-359, 2007.

[3] Sugano, S., Kato, I., "WABOT-2: autonomous robot with dexterous finger-arm coordination control in keyboard performance," in Proc. of the Int. Conference on Robotics and Automation, pp. 90-97, 1987

[4] Solis, J., Takanishi, A., et al.: The Waseda Flutist Robot WF-4RII in Comparison with a Professional Flutist, Computer Music Journal, Vol. 30(4), pp. 12-24, 2007.

[5] Takashima, S., Miyawaki, T., "Control of an automatic performance robot of saxophone: Performance control using standard MIDI files," in Proc. of the IROS Workshop on Musical Performance Robots and Its Applications, pp. 30-35, 2006.

[6] Shibuya, K., "Toward developing a violin playing robot: bowing by anthropomorphic robot arm and sound analysis," in Proc. of ROMAN, pp. 763-768, 2007.

[7] Kuwabara, H., Shimojo, M., et al., "The development of a violin musician robot," in IROS06 Workshop on Musical Performance Robots and Its Applications, pp. 18-23, 2006.

[8] Kapur A., Singer E., Benning, M, Tzanetakis G., "Integrating HyperInstruments, Musical Robots & Machine Musicianship for North Indian Classical Music," in Proc. of the New Interfaces for Musical Expression, pp. 238-241, 2007.

[9] Dannenberg, R.B., Brown, B., Zeglin, G., and R. Lupish, "McBlare: A Robotic Bagpipe Player", In Proc. of the NIME2005, Vancouver, 2005.

[10] Beilharz, K. "Interactively Determined Generative Sound Design for Sensate Environments: Extending Cyborg Control", Y. Pisan (eds), Interactive Entertainment, pp. 11-18. 2004.

[11] Hayashi, E., "Development of an automatic piano that produce appropriate: touch for the accurate expression of a soft tone," in Proc. of IROS Workshop on Musical Performance Robots and Its App., pp. 7-12, 2006.

[12] E. Singer, et al., "LEMUR's musical robots," in Proc. Of the Conference on New Interfaces for Musical Expression, pp. 183-184, 2004.

[13] Solis, J., Takanishi, A., et al., "Towards an automated transfer skill system," in Proc. of the International Conference on Music Computer, pp. 423-426, 2005.

[14] Weinberg, G., Driscoll, S., Parry, M., "Musical interactions with a perceptual robotic percussionist," in Proc. of the ROMAN2006, pp. 456-461, 2006.

[15] Ando Y., "Drive conditions of the flute and their influence upon harmonic structure of generated tone", Journal of the Acoustical Society of Japan, pp. 297-305, 1970.

[16] Schumacher, R. T., "Ab Initio Calculations of the Oscillations of a Clarinet," Acustica, 48(2), 1981, pp. 71-85.

[17] Kawato M., Furukawa K., Suzuki R., "A hierarchical neural network model for control and learning of voluntary movement," Byological Cybernetics, vol. 57, pp. 169-185, 1987.

[18] Petersen, K., Solis, J., Takanishi, A., "Development of a real-time gestural interface for hands-free musical performance control," in Proc. of the International Computer Music Conference, 2008

[19] http://www.septon.info/en/septon/what_septon.html.

[20] Fuks L & Sundberg J. "Blowing pressures in reed woodwind instruments", KTH TMH-QPSR 3/1996, Stockholm, pp. 41-56, 1996.

[21] Sturm, B. L., and J. Gibson, "Signals and Systems Using MATLAB: An Effective Application for Teaching Media Signal Processing to Artists and Engineers," in Proc. of the International Computer Music Conference, pp. 645–649, 2004.