Development of the Waseda Saxophonist Robot and Implementation of an Auditory Feedback Control

Jorge Solis[†, ‡], Klaus Petersen[±], Tetsuro Yamamoto[±], Masaki Takeuchi[±], Shimpei Ishikawa[±], Atsuo Takanishi[†, ‡], Kunimatsu Hashimoto[*]

[†]Waseda University, Faculty of Science and Engineering, Tokyo, Japan, +81-3-5369-7329

[±]Waseda University, Graduate School of Advanced Science and Engineering, Tokyo, Japan

[‡]Humanoid Robotics Institute (HRI), Tokyo, Japan

[*]Toyota Motor Corporation, Partner Robot Division, Toyota, Japan

solis@ieee.org

ABSTRACT

Since 2007, our research is related to the development of an anthropomorphic saxophonist robot, which it has been designed to imitate the saxophonist playing by mechanically reproducing the organs involved for playing a saxophone. Our research aims in understanding the motor control from an engineering point of view and enabling the communication. In this paper, the Waseda Saxophone Robot No. 2 (WAS-2) which is composed by 22-DOFs is detailed. The lip mechanism of WAS-2 has been designed with 3-DOFs to control the motion of the lower, upper and sideway lips. In addition, a human-like hand (16 DOF-s) has been designed to enable to play all the keys of the instrument. Regarding the improvement of the control system, a feed-forward control system with dead-time compensation has been implemented to assure the accurate control of the air pressure. In addition, the implementation of an auditory feedback control system has been proposed and implemented in order to adjust the positioning of the physical parameters of the components of the robot by providing a pitch feedback and defining a recovery position (off-line). A set of experiments were carried out to verify the mechanical design improvements and the dynamic response of the air pressure. As a result, the range of sound pressure has been increased and the proposed control system improved the dynamic response of the air pressure control.

Keywords

Humanoid Robot, Auditory Feedback, Music, Saxophone.

1. INTRODUCTION

The development of anthropomorphic robots is inspired by the ancient dream of humans replicating themselves. However, human behaviors are difficult to explain and model. Owing to the evolution of computers, electronics, and signal processing, this ancient dream is becoming a reality. In fact, current humanoid robots are able to perform activities such as dancing and playing musical instruments.

However, these mechanical devices are still far from understanding and processing emotional states as humans do. Research on musical performance robots seems like a particularly promising path toward helping to overcome this limitation [1], because music is a universal communication medium, at least within a given cultural context. Furthermore, research into robotic musical performance can shed light on aspects of expression that traditionally have been hidden behind the rubric of "musical intuition" [2]. In 1984, at Waseda University, the WABOT-2 was the first attempt of developing an anthropomorphic musical robot; it was able to play 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 [3]. The late Prof. Ichiro Kato argued that the artistic activity such as playing a keyboard instrument would require human-like intelligence and dexterity [4]. Nowadays, different kinds of musical performance robots (MPRs) and robotic musicians (RMs) have been developed. MPRs are designed to closely reproduce the required motor skills displayed by humans in order to play musical instruments ([5]-[8]).

Some examples of MPRs are described as follows. Shibuya is developing an anthropomorphic arm which reproduced the movement required to play a violin [6]. In particular, this violin robot is designed to produce expressive sounds by considering kansei (sensitivity). The arm has a total of 7-DOFs actuated by DC motors. From experimental results, the violin robot is able of playing notes with a high level of repetitiveness. Takashima has been developing different music performance robots that are able of playing wind instruments such as [7]: saxophone, trumpet, trombone and shakuhachi (traditional Japanese bamboo flute). In particular, the saxophone playing robot has been developed under the condition that the musical instrument played by robots should not be change or remodeled at all. This robot is composed of an artificial mouth, fingering mechanisms and air supplying system. Due to the complexity of replicating the motion of human fingers, the fingering mechanism is composed by twenty-three fingers so that each finger can press each key of the saxophone. Shimojo has worked on a violin-playing robot, which it is composed by a commercial 7-DOFs manipulator and a 2-DOFs fingering mechanism [9]. The end effector of the manipulator has been designed to hold a bow. A force/torque sensor has been attached to the end effector to control the bowing pressure. As a result, the violin-playing robot is able of performing simple musical scores.

In resume, the research on MPRs has been particularly intensified in recent decades. In fact, we may distinguish four different researches approaches [2]: Enabling the Human and Robot interaction, understanding the human motor control, introducing new ways of art/entertainment and introducing new methodologies of music teaching. Even that the above anthropomorphic musical robots have achieved promising results; up to now, only few of them are able to perform as human musicians (in terms of perception and motor dexterity). Moreover, none of the above robots are able of playing different kinds of musical instruments which could be useful to improve our understanding of the nature of human musicians.

For this purpose at Waseda University, since 2007, we have proposed the development of an anthropomorphic saxophone robot. In [7], the requirements for developing a tenor saxophone performance robot were introduced at Hosei University. Such automatic performance saxophone robot is composed by three main components: mouth mechanism (as a pressure controlled oscillating valve), the air supply mechanism (as a source of energy), and fingers (to make the column of air in the instrument shorter or longer). Such automatic performance saxophone robot has been designed under the principle that the instrument played by the robot should not be changed. However, a total of twenty-three fingers have been used to play the saxophone's keys (actuated by solenoids), a modified mouth mechanism has been designed (composed by a flexible artificial lip and a reed pressing force control mechanism were developed) to attach it with the mouthpiece, and no tonguing mechanism has been implemented (normally reproduced by the tongue motion). Moreover, in [8], the metasaxophone has been introduced. The Metasaxophone is an acoustic tenor saxophone retrofitted with an onboard computer microprocessor, and an array of sensors that convert performance data into independent continuous control messages for a computer. The instrument has additionally been outfitted with a unique microphone system allowing for detailed control of the amplified sound. While maintaining full acoustic functionality, the metasaxaphone is also a versatile computer controller and an electric instrument.

From the above research approaches, it could be rather difficult to understand the human motor control mechanism. Instead; based on our experience in developing the WF-4RIV, we proposed the development of an anthropomorphic saxophonist robot as an approach to extend our knowledge on the motor control skills required by players to play woodwind instruments. In addition, we would like to enable the interaction with musical partners to study in more detail the HRI in a musical context. As a matter of fact, we are aiming as a long-term goal two basic issues: enabling the interaction between two human-like robots (by developing two different robots able of performing different wind instruments), and enabling a single human-like robot to play different kind of wind instruments (our ability to enable a single human-like robot to play different kind of wind instruments can be studied in detail).

As a result of our research, in [10], we have presented the Waseda Saxophonist Robot No.1 (WAS-1), which was composed by 15-DOFs required to play an alto saxophone. In particular, the mouth (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. Both lips and oral cavity were made of a thermoplastic rubber. The tongue was implemented to reproduce the tonguing technique; which is an important source for adding expressiveness to the saxophone performance. Even that the lip mechanism of WAS-1 was useful in order to adjust the pitch of the saxophone sound, the range of sound pressure was too short. Moreover, the finger mechanism was designed only to play from C3 to C#5.

Therefore, in this paper, the mechanical design of the lip and finger mechanisms were improved to increase the range of sound pressure and to enable the saxophone robot to play all the keys of the alto saxophone (A#2 to F#5). From the control system point of view, a cascade feedback control system has been implemented. However, a considerable delay in the attack time to reach the desired air pressure was detected when playing musical scores at fast tempo. Thus, we describe the mechanical improvements of the simulated organs involved during the saxophone playing and the implementation of a feed-forward air pressure control system with dead-time and an auditory feedback control system.

2. WASEDA SAXOPHONIST ROBOT

In this year, we have developed the Waseda Saxophonist Robot No. 2 (WAS-2) which it has been designed to increase the range of sound by improving the design of the artificial lips and increase the range of fingering by designing a human-like hand. In particular, the WAS-2 is composed by 22-DOFs that reproduce the physiology and anatomy of the organs involved during the saxophone playing as follows (Figure 1): 3-DOFs (from which 1-DOF is passively controlled) to control the shape of the artificial lips, 16-DOFs for the human-like hand, 1-DOF for the tonguing mechanism and 2-DOFs for the lung system (1-DOF for the air pump and 1-DOF for the valve mechanism).

2.1 Mechanical Simulation of Human Organs

The artificial lip of the mouth mechanism of the WAS-1 was

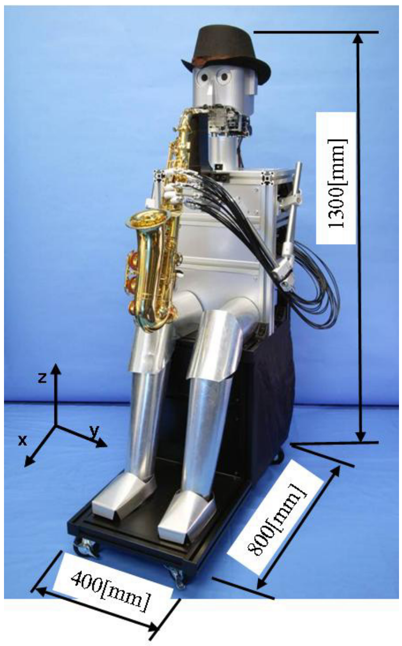

Figure 1. The Waseda Saxophonist Robot No. 2.

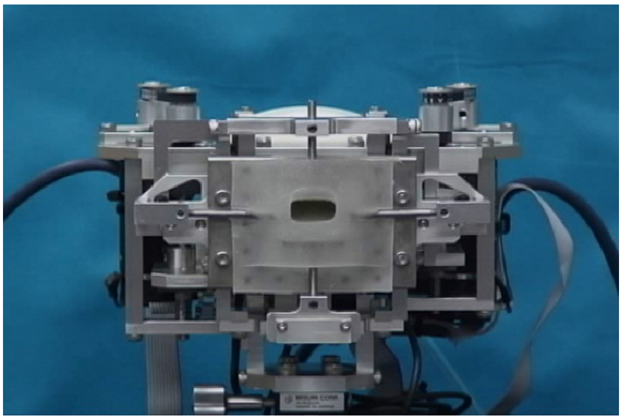

Figure 2. New mouth mechanism designed for the WAS-2, where the motion of upper, sideway and lower lips were implemented.

Figure 3. Arrangement of DOFs: a) right hand; b) left hand.

designed with 1-DOF in order to control the vertical motion of the lower lip [10]. Based on the up/down motion of the lower lip, it became possible to control the pitch of the saxophone sound. However, it is difficult to control the sound pressure by means on 1-DOF. In addition, as we have previously described, in the future the saxophonist robot should be able of stand-up as a human player does. Therefore, it is difficult to hold the instrument with the artificial mouth. For this purpose, the improved version of the mouth mechanism has been designed to expand the range of sound as well as to reduce the movement of the instrument when the robot holds the instrument with its mouth. The new artificial lip of the WAS-2 is shown in Fig. 2. As we may observe, the lip mechanism consists of 2-DOFs designed to control the up/down motion of both lower and upper lips. In addition, a passive 1-DOF has been implemented to modify the shape of the side-way lips. The material of the artificial lips is thermoplastic elastometer (Septon), which reproduces the elasticity and stiffness of human lips.

On the other hand, in order to produce the saxophone sound, it is required to control the motion of each of the fingers to push the correspondent keys. The finger mechanism of the WAS-1 was composed by a link connected directly to the RC motor axis. In particular, eleven motors were used in order to push each of the keys required to play from the C3 to C#5.

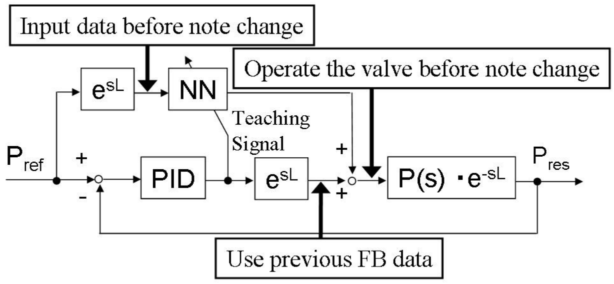

Figure 4. Block diagram of the proposed feed-forward control system with dead-time compensation implemented to assure the accuracy of the air pressure control during a performance.

However, with the alto saxophone is possible to play from A#2 to F#5. For this purpose, a new human-like hand has been designed, which it is composed by 16-DOFs (Figure 3). In order to reduce the weight on the hand part, the actuation mechanism is composed by a wire and pulley commented to the RC motor axis. In order to control the motion of each single finger, the RS485 communication protocol has been used.

2.2 Air Pressure FF Control System with Dead-Time Compensation

In our previous research, a cascade feedback control system was implemented to assure the accuracy of the air pressure during a musical performance [10]. Basically, based on the measurements of the pressure sensed at the output of the air pump and the position of the lower lips, the air pressure has been controlled. However, during the attack time the target air pressure is reached around 100ms later during a musical performance. Mainly, this effect is related to the way the musical performance control is implemented. Basically, the signal of the note to be played is sent to the control system through a MIDI message. As soon as message of a note change is received, the air pressure as well as the position of the lower lips is adjusted. Thus, a delay on the control of the air pressure is observed.

Actually, if we analyze the performance of a human playing the saxophone, the distance between the lungs and the oral cavity there are a few dozens of centimeters. This distance provokes the existence of dead-time. However, musicians when playing a musical performance, in order to avoid any delay on the adjustment of the air pressure located inside the oral cavity, controls the required parameters of the lungs and the mouth beforehand the notes changes.

Inspired on the above principle, a modified version of the feedback error learning has been used. The feedback error learning is a computational theory of supervised motor learning proposed by Kawato [11]; which is inspired by the way the central nervous system. In addition, Kawato extended that the cerebellum, by learning, acquires an internal model of inverse dynamics of the controlled object [12]. From this extension, the feedback error learning can be also used as training signal to acquire the inverse dynamics model of the controlled system based on Neural Networks. On the other hand, the dead-time compensation is used to control devices that take a long time to show any change to a change in input. A dead-time compensation control uses an element to predict how changes made now by the controller will affect the controlled variable in

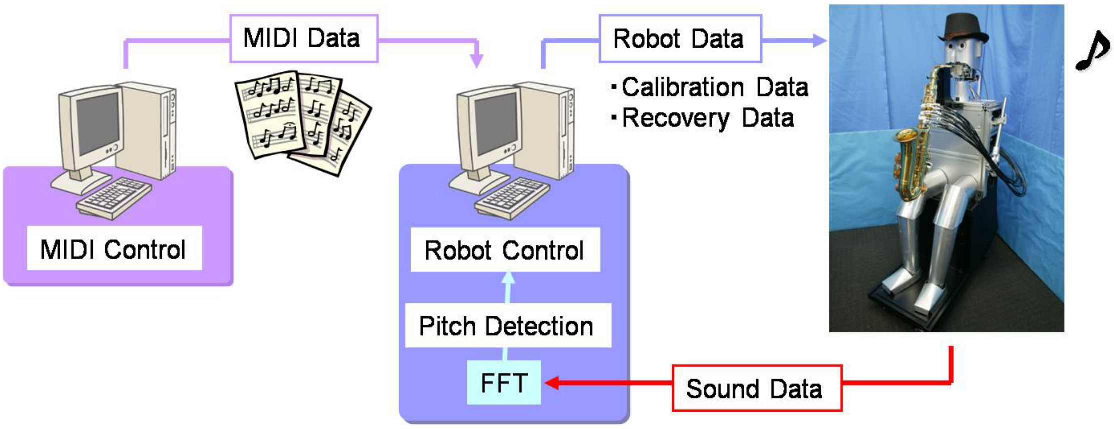

Figure 5. Block diagram of the proposed auditory feedback control system implemented for the WAS-2.

the future [13]. For our purpose, we have proposed the implementation of a feed-forward error learning control system with dead-time compensation as it is shown in Fig. 4. The inputs of the ANN are defined as follows (the input is based on the difference with the previous played note): pressure reference, note, and lower/upper lips position. In this case, a total of six hidden units were used (experimentally determined while varying the number of hidden units). As an output, the position of the air valve is controller to assure the accurate control of the required air pressure to blow a sound. In addition, a dead-time factor (referred as e^{sL}) is introduced to compensate the delay during the attack time.

2.3 Auditory Feedback Control

From the experimental results obtained in [10] with the WAS-1, we have determined the importance of correcting the pitch of the sound produced by the saxophonist robot. Inspired on the way professional saxophonist improve their sound, we have proposed the implementation of an auditory feedback control system, as it is shown in Fig. 5. In order to adjust the pitch of the sound produced by the robot, we use a contact microphone (CM-100L commercialized by Korg) to feedback the frequency of vibration of the instrument (which it is considered to be the pitch of the sound produced by the sound). The reason of using a contact microphone (which is commonly used for acoustic tuning) instead of a conventional microphone is based on the principle that the environmental noise (i.e. inside a concert hall, etc.) will not be captured by the contact microphone. Therefore, we may assure the correct recognition of the frequency of the pitch by means of the Fast-Fourier Transform (FFT).

In order to implement an auditory feedback control system, in addition to the calibration data (defined as the optimal physical parameter of each of the mechanical component of the robot [10]), we have defined the adjustment data (defined as the physical parameter change required for recovering from a miss-tone). As a first approach, both the calibration and recovery data are obtained before the saxophone performance (offline).

As a result, in order to implement the auditory feedback control system, we proposed two different procedures to be carried out before and during the performance (Figure 6). Regarding the procedure before the performance, the detected pitch is classified as correct or miss tone (over blowing or no sound, etc.). As a first approach, when a miss-tone is found, the recovery data is manually determined by the operator. After that, during a performance, the robot automatically detects a miss- tone and the recovery data is loaded to assure its correct tuning of the sound.

3. EXPERIMENTS AND RESULTS

3.1 Range of Sound Pressure

In order to verify if the designed new mouth mechanism enables to extend the range of sound pressure; we have compared the previous mechanism with the new one while playing the notes from C3 to C5. The experiment results are shown in Fig. 7. As we may observe, the new mechanism has effectively increased the range of sound pressure (an average increment of 33%). Even though the range of sound pressure was expanded, still there are differences with the one measure by an intermediate level saxophonist. Thanks to this improvement, we could perform experiments with the WAS-2 in order to vary the dynamic properties of the sound such as decrescendo.

3.2 Air Pressure FF Control

In order to determine the effectiveness of the proposed control system implemented on the WAS-2, we have programmed the saxophonist robot to perform the moonlight serenade composed by Glenn Miller. In order to define the setting of the training of

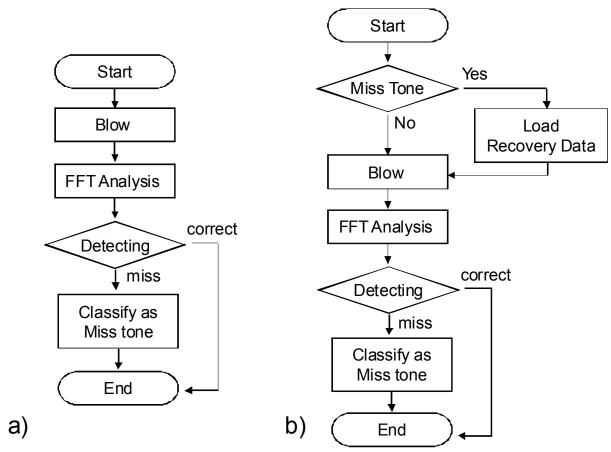

Figure 6. a) Procedure proposed to carry out before the performance in order to determine the recovery data; b) Proposed procedure to carry out during the performance.

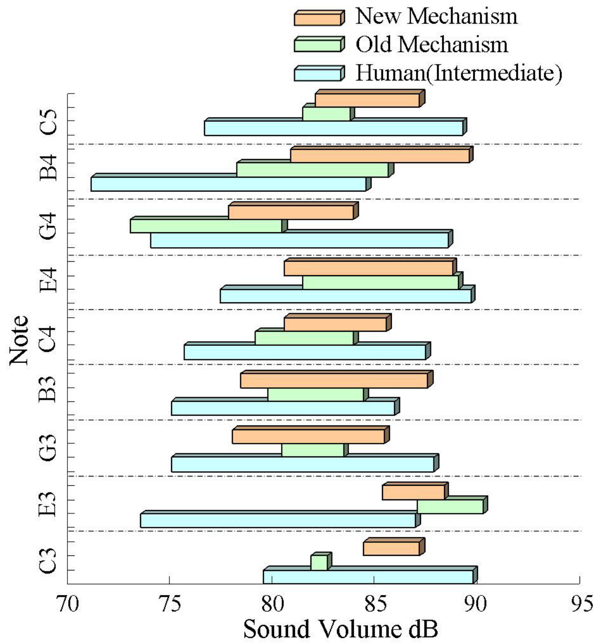

Figure 7. Comparison of the range of air pressure between the previous mouth mechanism of WAS-1 and the new one of WAS-2.

the ANN, we perform a preliminary experiment for varying the number of hidden units. The experimental results are shown in Fig. 8. As we may observe, the best fitting was found when six hidden units were used. Thus, by defining the number of hidden units (6) and learning steps to 523), we compare the performance while using the previous and proposed control system strategies. The experimental results are shown in Fig. 9. As we may observe, we can clearly observe that the proposed feed-forward control system with dead-time compensation presented a more stable dynamic response to the air pressure reference (in particular during the first 5sec of the musical performance).

In order to compare both dynamic responses, we have computed the correlation coefficient respect to the target signal (P_{ref}). The correlation coefficient is a quantity that gives the quality of a least squares fitting to the original data (in this case, the target signal). As a result, we found that the resulted air pressure with the feed-forward control system with dead-time compensation was more similar to the target one (correlation coefficient of 0.636) than the previous control system (correlation coefficient of 0.459). With these results, we can assure the improvements done respect to the previous control system (cascade feedback control).

3.3 Qualitative Performance Evaluation

In addition, we have performed a subjective analysis of the improvements thanks to the implementation of the propose control system. For this purpose, we have recorded the performance of WAS-2 while playing the moonlight serenade with the proposed control system and with the previous one. A total of twelve subjects were asked to compare the above recordings with the performance of a professional saxophonist. The evaluation criterions are: pitch quality, tone stability and overall performance. The maximum score (10) was considered the professional one.

The experimental results are shown in Fig. 10. As we may observe, a higher evaluation was given to the performance in all the evaluation parameters with the proposed auditory control system.

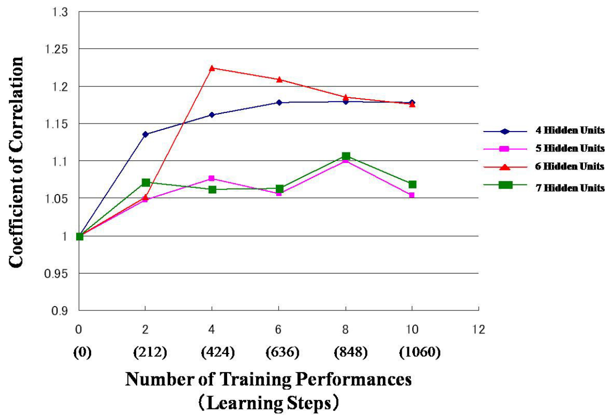

Figure 8. Experimental results with the NN while varying the number of Hidden Units.

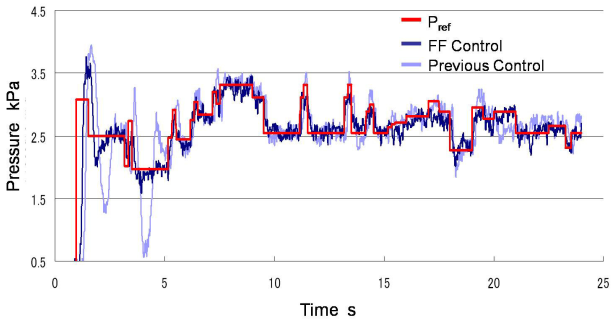

Figure 9. Experimental results with the Feed-Forward Control System with Dead Time Compensation.

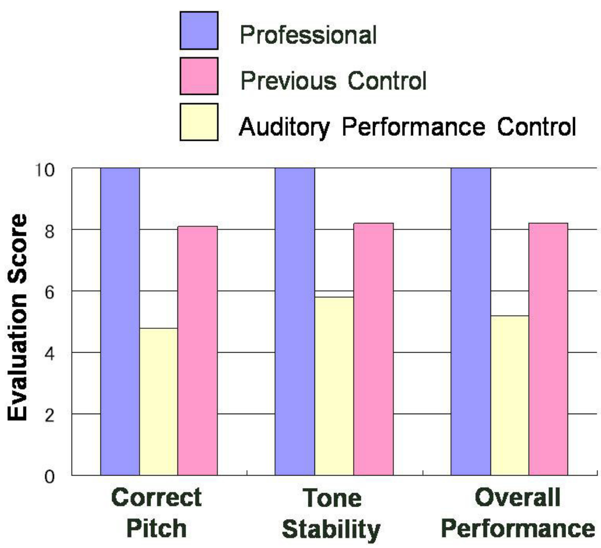

Figure 10. Experiments while evaluating the proposed performance control implemented on the WAS-2 and compared with the previous one and the performance of a professional saxophonist.

4. ACKNOWLEDGMENTS
A part of this research was done at the Humanoid Robotics Institute (HRI), Waseda University and at the Center for Advanced Biomedical Sciences (TWINs). This research was supported (in part) by a Gifu-in-Aid for the WABOT-HOUSE Project by Gifu Prefecture. This work was supported in part by Global COE Program "Global Robot Academia" from the Ministry of Education, Culture, Sports, Science and Technology of Japan. WAS-2 has been designed by 3D CAD design software SolidWorks. Special thanks to SolidWorks Japan K.K. for the software contribution.

5. CONCLUSIONS AND FUTURE WORK
In this paper, the development of the Waseda Saxophonist Robot No.2 (WAS-2) has been presented. In particular, the improvements of the mechanical simulated organs involved in the saxophone playing were introduced. In addition the implementation of an auditory feedback system has been described. A set of experiments were carried out to verify the improvements of the musical performance of the WAS-2. As a result, we could confirm the increase of the sound pressure range, the imitation of the crescendo/decrescendo and the improvements on the correctness of the pitch, tone stability and overall performance.

As a future work, the proposed feed-forward control system implemented for the air pressure will be also extended to control the pitch (there is still some deviations of the pitch during the performance). In addition, further studies will carried out to study the influence on the shape of the oral cavity while producing the saxophone sound.

6. REFERENCES

[1] Solis, J., Chida, K., Suefuji, K, Takanishi, A., The development of the anthropomorphic flutist robot at Waseda University, International Journal of Humanoid Robots, Vol. 3(2), pp. 127-151 (2006).

[2] Solis, J., Takanishi, A., An overview of the research approaches on Musical Performance Robots," in Proc. of the International Computer Music Conference, Copenhagen, 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 ICRA, New York, pp. 90-97, (1987).

[4] Kato, I., Ohteru, S., Kobayashi, H., Shirai, K., Uchiyama, A., Information-power machine with senses and limbs, in Proc. of the CISM-IFToMM Symposium on Theory and Practice of Robots and Manipulators, Udine, pp. 12-24, (1973).

[5] Solis, J., Chida, K.; Suefuji, K., Taniguchi, K., Hashimoto, S.M., Takanishi, A., The Waseda Flutist Robot WF-4RII in Comparison with a Professional Flutist, Computer Music Journal, Vol. 30(4), pp. 12–24 (2006)

[6] Shibuya, K., Toward developing a violin playing robot: bowing by anthropomorphic robot arm and sound analysis, in Proc. of the 16th IEEE International Conference on Robot and Human Interactive Communication, Jeju Island, pp. 763-768, (2007).

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

[8] Burtner, M., The Metasaxophone: Concept, Implementation and Mapping Strategies for a New Computer Music Instrument, Organised Sound, Vol. 7(2), pp. 201-213, 2003

[9] Kuwabara, H., Seki, H., Sasada, Y., Aiguo, M., Shimojo, M., The development of a violin musician's robot," in Proc. of the IROS Workshop on Musical Performance Robots and Its Applications, Beijing, pp. 18-23, (2006).

[10] Solis, J., Ninomiya, T., Petersen, K., Yamamoto, T., Takanishi, A., Anthropomorphic Musical Performance Robots at Waseda University: Increasing Understanding of the Nature of Human Musical Interaction, in Proceedings of the 9th International Conference New Interfaces for Musical Expression, pp. 64–69 (2009)

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

[12] Kawato, M., Gomi, H. A computational model of four regions of the cerebellum based on feedback-error-learning," Biological Cybernetics, vol. 68, pp. 95-103 (1992).

[13] Kim, H., Kim, K., Young, M., On-Line Dead-Time Compensation Method Based on Time Delay Control. IEEE Transactions on Control Systems Technology, Vol. 11(2), pp. 279-286 (2003).