The 2010 IEEE/RSJ International Conference on
Intelligent Robots and Systems
October 18-22, 2010, Taipei, Taiwan

Implementation of an Overblowing Correction Controller and the Proposal of a Quantitative Assessment of the Sound's Pitch for the Anthropomorphic Saxophonist Robot WAS-2

Jorge Solis, *Member, IEEE*, Klaus Petersen, Tetsuro Yamamoto, Masaki Takeuchi, Shimpei Ishikawa, Atsuo Takanishi, *Member, IEEE*, and Kunimatsu Hashimoto

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 a previous paper, the Waseda Saxophone Robot No. 2 (WAS-2) which is composed by 22-DOFs has been presented. Moreover, a feedback error learning with dead time compensation has been implemented to control the air pressure of the robot. However, such a controller couldn't deal with the overblowing effects (unsteady tones) that are found during a musical performance. Therefore; in this paper, the implementation of an Overblowing Correction Controller (OCC) has been proposed and implemented in order to assure the steady tone during the performance by using the pitch feedback signal to detect the overblowing condition and by defining a recovery position (off-line) to correct it. Moreover, a saxophone sound evaluation function (sustain phase) has been proposed to compare the sound produced by human players and the robot. A set of experiments were carried out to verify the improvements on the musical performance of the robot and its sound has been quantitatively compared with human saxophonists. From the experimental results, we could observe improvements on the pitch (correctness) and tone stability.**

I. Introduction

Since the golden era of automata, the development of mechanical dolls served as a mean to understand how the human brain is able of coordinating multi-degrees of freedom in order to play musical instruments [1-2]. A particular interest was given to wind instruments as a research approach for understanding human breathing mechanism. Nowadays, different kinds of wind playing-instrument automated machines and humanoid robots have been developed for playing wind instruments to understand the human motor control from an engineering point of view [1-6]. In this paper, we particular deal with the development of a saxophone-playing robot.

Up to now, different saxophone-playing robots have been built. In particular, the "Saxophone Playing Robot" was developed by Takashima [4]. This robot; named APR-SX2, 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). The artificial mouth consists of a flexible artificial lips and a reed pressing mechanism. The artificial lips are made of a rubber balloon filled with silicon oil with a proper viscosity. The air supplying system (lungs) consists of an air pump, a diffuser tank with a pressure control system (the supplied air is controller at the pressure from 0.0 MPa to 0.02 MPa). The APR-SX2 has been designed under the principle that the instrument played by the robot should not be changed. A total of twenty-three fingers have been used to play the saxophone's keys (actuated by solenoids) and 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). The control system implemented for the APR-SX2 is composed by one computer dedicated to the control of the key-fingering, air pressure and flow, pitch of the tones (related to the applied force in the reed), tonguing and pitch bending. In order to synchronize all the performance, the musical data was sent to the control computer through MIDI in real-time. In particular, the SMF format was selected to determine the status of the tongue mechanism (on or off), the vibrato mechanism (pitch or volume) and pitch bend (applied force on the reed).

On the other hand, the authors have developed the Waseda Saxophonist Robot No.1 (WAS-1), which was composed by 15-DOFs required to play an alto saxophone [8]. 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

Manuscript received February 20, 2010. 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 is supported (in part) by a Gifu-in-Aid for the WABOT-HOUSE Project by Gifu Prefecture. This work is also 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

Jorge Solis is with the Faculty of Science and Engineering, Waseda University; and a researcher at the Humanoid Robotics Institute (HRI), Waseda University (solis@ieee.org)

Klaus Petersen, Tetsuro Yamamoto, Masaki Takeuchi, Shimpei Ishikawa are with the Graduate School of Advanced Science and Engineering, Waseda University

Atsuo Takanishi is with the Faculty of Science and Engineering, Waseda University; and one of the core members of the Humanoid Robotics Institute (HRI), Waseda University (contact@takanishi.mech.waseda.ac.jp)

Kunimatsu Hashimoto is with the Partner Robot Division, Toyota Motor Corporation.

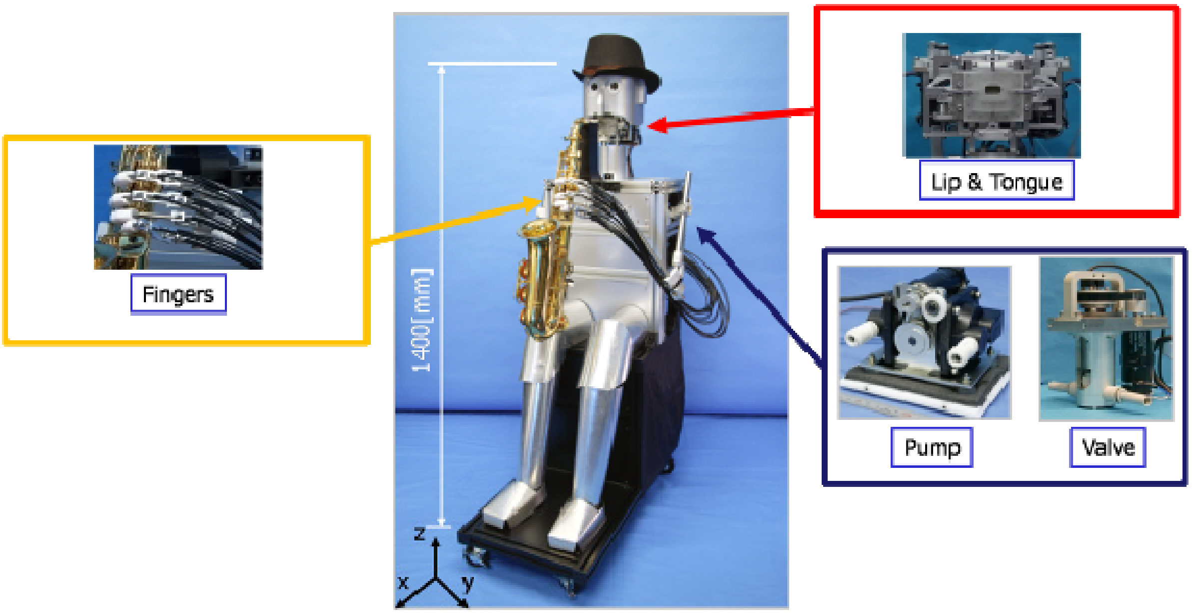


Figure 1. The WAseda Saxophonist Robot No.2 (WAS-2) was designed to play an alto saxophone [9]. The WAS-2 is composed by an artificial lung (air pump with 1-DOF and air valve with 1-DOF), tongue (1-DOf), fingers (16-DOFs) and mouth mechanism (3-DOFs).

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. From the control point of view, a cascade feedback control system was implemented to assure the accuracy of the air pressure during a musical performance. 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.

From the mechanical design point of view, in [9], the authors presented the improved Waseda Saxophonist Robot No.2 (WAS-2). 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). The mouth mechanism of the WAS-2 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 finger mechanism of the WAS-2 is composed by 16-DOFs to push the correspondent keys (A#2 to F#5.). 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 addition, in order to assure the accurate control of the air pressure, a modified version of the feedback error learning has been implemented [9]. For this purpose, a feed-forward error learning control system with dead-time compensation was implemented. 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. For this case, a total of six hidden units were used. 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.

However, during a performance, the above implemented air pressure control system cannot correct the sound itself produced by the robot. In fact, during a performance of WAS-2, we identified overblowing cases. Overblowing is a typical problem found while playing bagpipes, saxophones, etc. that is related to unsteady tones [10]. For this purpose, in this paper, we proposed the implementation of an Overblowing Correction Controller (OCC) implemented by using the pitch feedback signal to detect the overblowing condition and by defining a recovery position to correct its effect on the pitch. Moreover, we also proposed a specific evaluation function to quantitatively compare the quality of the sound produced by human saxophonists and the robot.

II. Overblowing Correction Controller

The control system of the WAS-2 is composed by a PC Control and a PC Sequencer. The PC Control is used to acquire and process the information from each of the degrees of freedom of the saxophonist robot as well as controlling the air flow/pressure to produce the desired sound. The PC Control has as inputs the MIDI data and Music Pattern Generator (calibration data). The Music Pattern Generator is

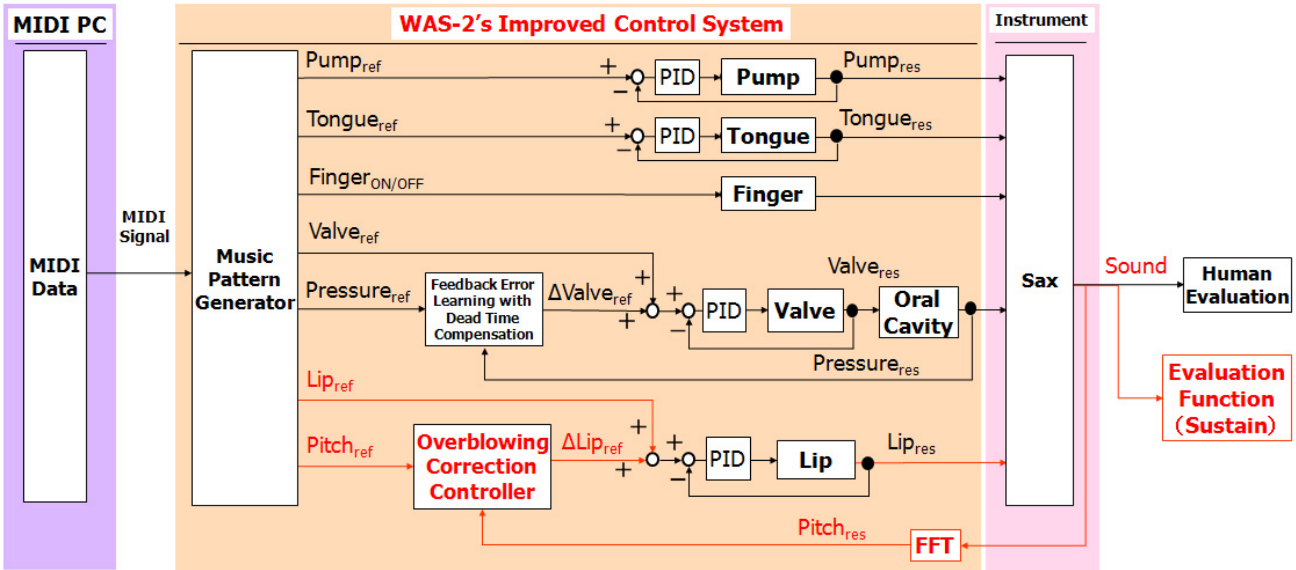


Figure 2. Block diagram of the improved control system implemented for the WAS-2 (the added modules are indicated in red color).

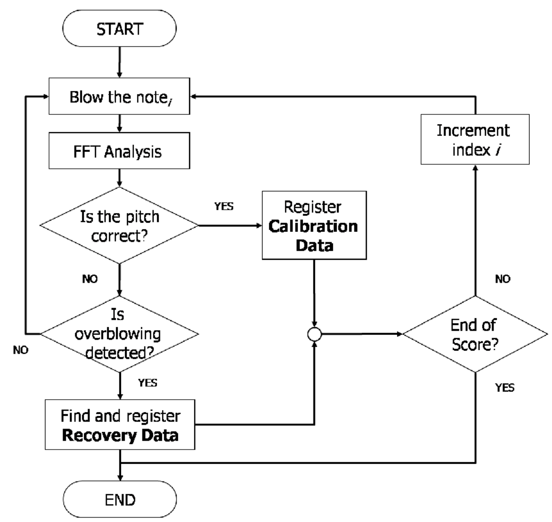
designed to output the calibration parameters required in order to produce the desired saxophone sound.

Inspired on the way professional saxophonist improve their sound, we have proposed the implementation of an Overblowing Correction Controller (OCC), as it is shown in Fig. 2. As we may observe, the $Pitch_{ref}$ and the $Pitch_{res}$ are used as an inputs to compute the required amount of compensation of the lower lip's positioning (ΔLip_{ref}). Such amount compensation is the added to the predefined position of the lower lip (Lip_{ref}) defined during the calibration process of the robot.

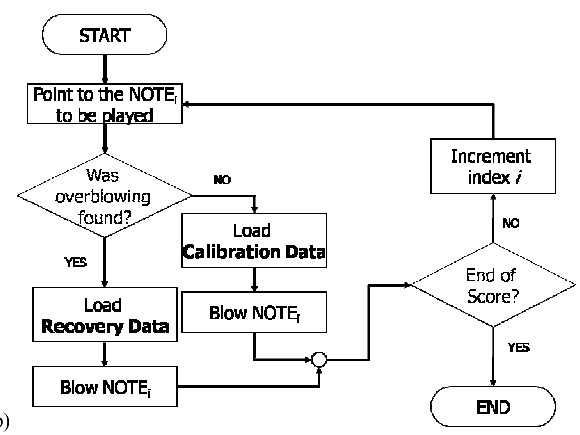
In order to compute the pitch of the sound produced by the robot ($Pitch_{res}$), we use a contact microphone (CM-100L commercialized by Korg) to feedback the frequency of vibration of the instrument (which it is considered to be the pitch of the sound produced by the sound). The reason of using a contact 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).

Due to the complexity of determining the amount of required compensation of the lower lip's position (ΔLip_{ref}), as a first approach, we proposed to include the recovering data as a part of the calibration procedure of the robot. For this purpose, the following procedure was implemented for the overblowing correction controller based on two phases (Figure 3a): calibration and performance.

Regarding the calibration phase, basically the operator is required to record not only the calibration data related to each of the mechanical components of the robot (as commonly was done for the WAS-1) but also the recovery data in order to produce the sound is recorded. During this process, the operator manually detects an overblowing case and readjusts



a)



b)

Figure 3. a) Calibration Phase is done by the operator where both the calibration data as well as the recovery one are determined; b) Performance Phase is done by determining the when to use either the calibration data or the recovery one..

the parameters to produce the correct pitch. Those required adjustments are recorded as “Recovery Data”.

Regarding the performance phase; at first, the operator loads the recovery data into the music pattern generator. After that, the signal of starting for the correspondent musical performance is activated. During the performance, the robot is able of automatically detecting overblowing cases by simply analyzing the harmonic contents of the sound. As it is shown in Fig. 4, we may identify the correct pitch and overblowing one by observing the amplitude relation between the fundamental frequency (f_0) and the next harmonic (f_1). When the amplitude of the f_0 is greater than f_1, the produced sound is blown in tune. In contrast, the produced sound is blown with an overblowing effect. Basically, during a performance, if a saxophone is detected with overblowing effect, the recovery data is automatically used to avoid its effect on the sound produced by the robot (of course, based on our initial approach, this will be exclusively effective for the musical score considered during the calibration phase of the robot).

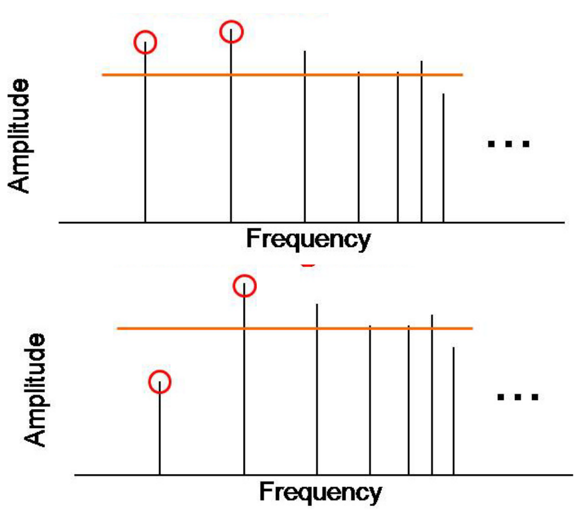


Figure 4 a) When the sound is played with a steady tone, the peak amplitude of the fundamental frequency is nearly the biggest; b) An overblowing case is detected when the peak amplitude of the fundamental frequency is quite different from the biggest one. .

III. SAXOPHONE SOUND EVALUATION FUNCTION

In this year, we have focused on proposing an evaluation function in order to quantitatively evaluate the saxophone sound as an approach to understand better how to improve the musical performance of the anthropomorphic saxophonist robot. For this purpose, we have proposed the following evaluation function during the sustain phase (E_s); as it is shown in Eq. (1). As we may observe, the proposed evaluation function is composed by three main parameters: harmonic structure (*M-H*), noise (*N*) and fluctuation (*1/f*).

The harmonic structure has been proposed to evaluate if the produced sound is played as it is expected. For this purpose, basically we are able of measuring the level of deviation of the pitch. On the other hand, the noise is referred as the high frequency components contained on the spectrum of the saxophone sound.

$$E_S = \frac{w_1 \times (M - H) - w_2 \times (N)}{Volume} + w_3 \times \frac{1}{f} \quad (1)$$

M: Harmonic level [dB] *H*: Semi-Harmonics level [dB]
Volume: Volume level [dB] *N*: Noise [dB]
1/f: Fluctuation

Finally, even that the previously introduced two parameters determine the difference between sounds played correctly or not, there is still the impression that the saxophone sound produced by the WAS-2 is machine-like. Therefore, we have proposed the addition of the fluctuation parameter. The fluctuation is defined as the inclination degree of the power spectrum in music and it has been related to a unique property of humans while playing music. The fluctuation parameter is obtained by analyzing the power spectrum of the sound (Frequency vs. Amplitude) and by computing the slope of such a relationship. From the result of several researchers, it has been demonstrated that in the case of humans, it is supposed that the slope is nearly to -1 [11-12]. As an example, we can observe the difference on a sound played by WAS-2 and a professional saxophonist by computing the FFT of the produced sound (Figure 5). As we may observe, the resultant slope is quite different between them and the professional one is nearly -1.

Regarding the weighting coefficients w_1, w_2 and w_3 of Eq. (1), we have determined experimentally by using the discriminant analysis method [13]. The Discriminant analysis is a statistical technique for classifying a set of observations into predefined classes. The purpose is to determine the class of an observation based on a set of variables known as predictors or input variables. The model is built based on a set of observations for which the classes are known (training set). For our case, we have collected the data from two professional and two beginner saxophonists. As a result, we

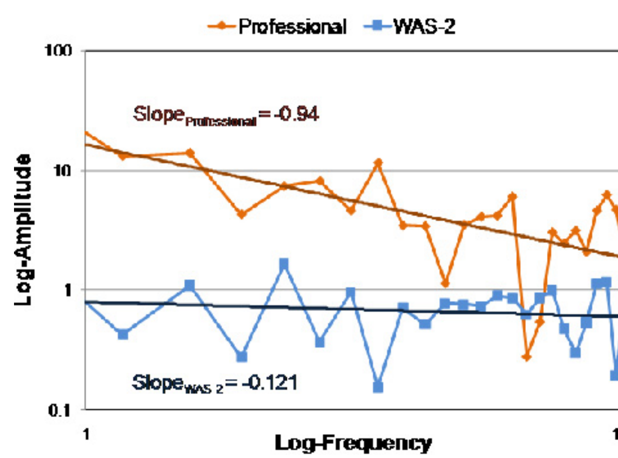


Figure 5 Comparison of the fluctuation by plotting the power spectral obtained from the saxophone sound produced by the WAS-2 and a professional saxophonist.

have determined the weighting coefficients values as 0.92, -0.05 and 0.29 respectively.

IV. Experiments & Results

A. Overblowing Correction Controller

In this experiment, we proposed to confirm the improvements achieved by the addition of the overblowing correction controller. For this purpose, we have asked twelve subjects to evaluate the performance of the robot (with and without the overblowing correction controller) and compare it with the performance of a professional saxophonist recording based on the following evaluation parameters: pitch correctness, tone stability and overall performance. In particular, twelve subjects and two professional saxophonists were asked to evaluate the different two recordings in a scale from 1 to 10.

The experimental results are shown in Fig. 6. As we may observe, a higher evaluation scoring was given by both common subjects and professionals to the performance of the robot when the overblowing correction controller is used. The experimental results were divided in two parts depending on the subjects: unskilled evaluators (Figure 6a) and professional saxophonists (Figure 6b).

Regarding the first case, the collected data from the unskilled evaluators was subjected to a two-way ANOVA analysis (Factors: Recorded Sounds and Evaluation Parameters). As result, we found a statistical significant difference among the evaluation from the three recorded sounds ($P<0.001$). On the other hand, there is no significant difference among the three evaluation parameters ($P=0.3561$). While comparing the differences among the three cases (Bonferroni test was used), we may observe a statistical difference between the WAS-2 with OCC and the one without OCC for the pitch evaluation parameters ($P<0.01$). Moreover, for the tone stability and overall evaluation parameter, a statistical difference was also detected ($P<0.05$). On the other hand, we also found a statistically difference between the WAS-2 with OCC and the professional one for all the three evaluation parameters ($P<0.05$).

Regarding the second case, the collected data from the professional evaluators was also subjected to a two-way ANOVA analysis. As result, we also found a statistical significant difference among the evaluation from the three recorded sounds ($P<0.001$). On the other hand, there is also no significant difference among the three evaluation parameters ($P=0.25$). While comparing the differences among the three cases, we may observe a statistical difference between the WAS-2 with OCC and the one without OCC was observed for the pitch and tone stability evaluation parameters ($P<0.01$). Moreover, for the overall evaluation parameter, a statistical difference was detected ($P<0.05$). However, we still found a statistically significant difference between the WAS-2 with OCC and the professional one for all the three evaluation parameters ($P<0.0001$).

From the above, results, we may confirm the improvements on the WAS-2 while using the OCC during the performance. However, still further improvements are required to enhance the WAS-2's sound based on the evaluation results obtained from the professional evaluators.

B. Evaluation Function (Sustain Phase)

In this experiment, we have focused on quantitatively determining the differences between different levels of saxophonist players and the WAS-2. For this purpose, we have collected the sound data (C3~A3) from one professional, one beginner, one intermediate and the WAS-2. The collected data was then analyzing by using the Eq. 1 to determine the quantitative difference among all of them .The experimental results are shown in Fig. 7. As we may observe, the beginner

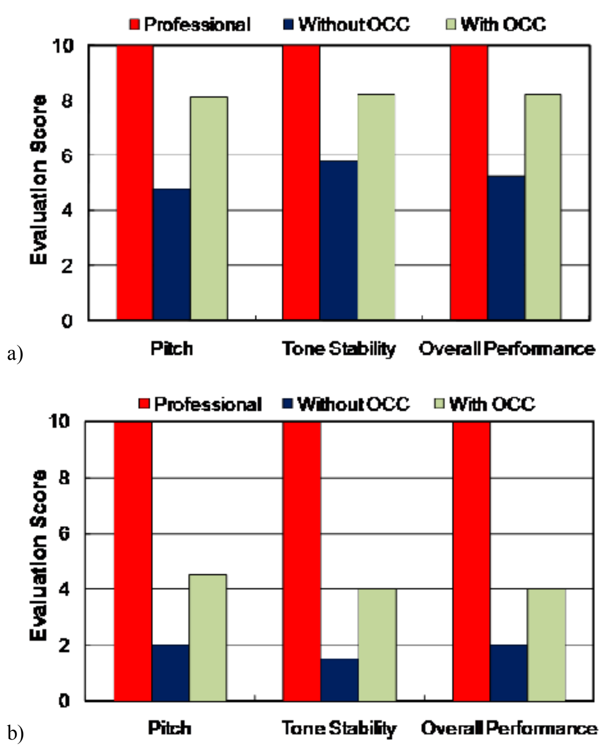


Figure 6 Experimental results obtained while subjectively evaluating the WAS-2's performance when the proposed overblowing correction is used or not: a) evaluation results from twelve subjects (without musical experience); b) evaluation results from two professional saxophonists

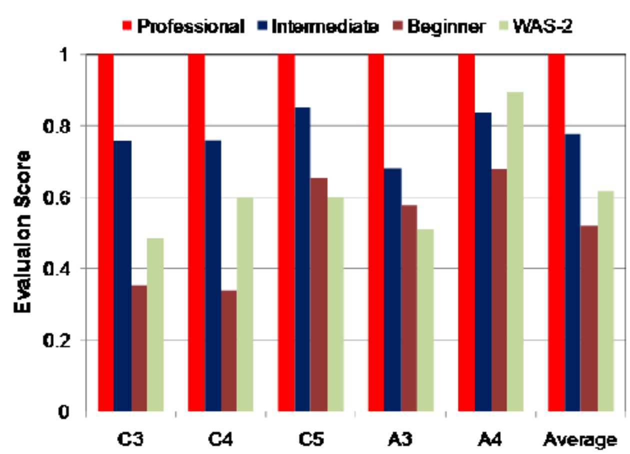


Figure 7 Experimental results obtained while quantitatively comparing the performance of WAS-2 and human flautist players (professionals, intermediate and beginners).

player is found with the lowest scoring average (E_s = 0.52) and the professional one with the highest one (E_s = 1.00). Then, the intermediate player and the WAS-2 are found in between them (E_s = 0.78 and E_s = 0.61 respectively); where the WAS-2 obtained a lower scoring (but still higher than the beginner one).

The collected data was also subjected to the two-way ANOVA analysis (Factors: Player and Notes). As result, we found a statistical significant difference among the considered saxophonist players ($P<0.001$). On the other hand, there is no quite statistically significant difference among the evaluated notes ($P=0.0774$). However, while comparing the differences among the four players (Bonferroni test was used), we observed no statistical difference among the WAS-2, Intermediate and Beginner ($P>0.05$) for all the evaluated notes. Moreover, while comparing the performance of WAS-2 and the professional one, we found a statistically difference while considering the notes C3 and A3 ($P<0.05$). In contrast; for the other considered notes, no significant statistically difference was found ($P>0.05$).

From the above results, we could confirm the effectiveness of quantitatively evaluating the saxophone sound during the sustain phase. However, we still require performing further experiments with higher amount of data to calculate a more accurate value for the weighting coefficients from Eq. 1.

V. Conclusions & Future Work

In this paper, we have presented the implementation details of an Overblowing Correction Controller for the Waseda Saxophonist Robot No.2 (WAS-2). The OCC has been implemented by using the pitch feedback signal to detect the overblowing condition and by defining the recovery data to correct the pitch. Furthermore, an evaluation function of the saxophone sound was proposed for the sustain phase; which it is composed by three evaluation parameters (harmonic structure, noise and fluctuation). Finally, a set of experiments were proposed to compare the improvements achieved on the pitch correctness, tone stability and overall performance. From the experimental results, we could confirm its improvements when the proposed control system was used. Moreover, we could also quantitatively compare the differences among different saxophonist players and the WAS-2 based on the proposed evaluation function.

As a future work, we would implement a MIMO-based pressure-pitch controller to improve the correctness of the pitch (particularly when some deviations are presented). The proposed MIMO system will be implemented based on a modified version of the Feedback Error Learning. In addition, the proposed evaluation function will be compared with other methods in the quantitative assessment of the sound pitch. Furthermore, an evaluation function for the attack phase of the sound will be also proposed.

References

[1] Doyon, A., Liaigre,L. (1966). Jacques Vaucanson: Mecanicien de Genie, PUF.

[2] Klaedefabrik, K.B., (2005). Martin Riches - Maskinerne / The Machines. Kehrer Verlag, pp. 10-13.

[3] Solis, J., Taniguchi, K., Ninomiya, T., Takanishi, A. (2008) Understanding the Mechanisms of the Human Motor Control by Imitating Flute Playing with the Waseda Flutist Robot WF-4RIV," Mechanism and Machine Theory, Vol. 44 (3), pp. 527-540.

[4] Takashima S., Miyawaki, T. (2006). Control of an automatic performance robot of saxophone: performance control using standard MID files, Proc. of the IEEE/RSJ Int. Conference on Intelligent Robots and Systems - Workshop: Musical Performance Robots and Its Applications, pp. 30-35.

[5] Vaucanson, J. de. (1979). *Le Mécanisme du Fluteur Automate*; *An Account of the Mechanism of an Automation: or, Image Playing on the German-Flute*. The Flute Library, First Series No. 5, Ed. Frans Vester. Intro. David Lasocki. Buren (GLD), The Netherlands: Uitgeverij Frits Knuf.

[6] Solis, J., Taniguchi, K., Ninomiya, T., Takanishi, A. (2008) Understanding the Mechanisms of the Human Motor Control by Imitating Flute Playing with the Waseda Flutist Robot WF-4RIV," Mechanism and Machine Theory, Vol. 44 (3), pp. 527-540

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

[8] 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)

[9] Solis J., Petersen, K., Yamamoto, T., Takeuchi, M., Ishikawa, S., Takanishi, A., Hashimoto, K., Design of New Mouth and Hand Mechanisms of the Anthropomorphic Saxophonist Robot and Implementation of an Air Pressure Feed-Forward Control with Dead-Time Compensation, Proceedings of the International Conference on Robotics and Automation, 2010.

[10] Kool, J., The Saxophone. Ed. J. J. Weber, Leipzig. 1931

[11] Sakamoto, H., Hayashi, G., Sugiura, S., Tsujikawa, M., "Psycho-circulatory responses caused by listening to music, and exposure to fluctuating noise or steady noise", Journal of Sound and Vibration, Vol. 250(1), pp. 23-29, 2002.

[12] http://www.asahi-net.or.jp/~HB9T-KTD/music/English/Research/Youragi/music_and_fractuation.htm

[13] Friedman, J.H.: "Regularized Discriminant Analysis", Journal of the American Statistical Association, vol. 84, no. 405, pp. 165-175, March 1989.