This is the published version of a paper published in *International Journal of Advanced Robotic Systems*.

Citation for the original published paper (version of record):

https://doi.org/10.5772/59232

Access to the published version may require subscription.

N.B. When citing this work, cite the original published paper.

Permanent link to this version:
http://urn.kb.se/resolve?urn=urn:nbn:se:kau:diva-33926
Biologically-inspired Control Architecture for Musical Performance Robots

Invited Review Article

Jorge Solis¹,²,³, Kenichiro Ozawa⁴, Maasaki Takeuchi⁴, Takafumi Kusano⁴, Shimpei Ishikawa⁵, Klaus Petersen² and Atsuo Takanishi³,⁵

¹ Department of Engineering and Physics, Karlstad University, Karlstad, Sweden
² Research Institute for Science and Engineering, Waseda University, Tokyo, Japan
³ Humanoid Robotics Institute, Waseda University, Tokyo, Japan
⁴ Graduate School of Advanced Science and Engineering, Waseda University, Tokyo, Japan
⁵ Department of Mechanical Engineering, Waseda University, Tokyo, Japan

* Corresponding author E-mail: solis@ieee.org

Received 14 May 2014; Accepted 13 Sep 2014

DOI: 10.5772/59232

© 2014 The Author(s). Licensee InTech. This is an open access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract At Waseda University, since 1990, the authors have been developing anthropomorphic musical performance robots as a means for understanding human control, introducing novel ways of interaction between musical partners and robots, and proposing applications for humanoid robots. In this paper, the design of a biologically-inspired control architecture for both an anthropomorphic flutist robot and a saxophone playing robot are described. As for the flutist robot, the authors have focused on implementing an auditory feedback system to improve the calibration procedure for the robot in order to play all the notes correctly during a performance. In particular, the proposed auditory feedback system is composed of three main modules: an Expressive Music Generator, a Feed Forward Air Pressure Control System and a Pitch Evaluation System. As for the saxophone-playing robot, a pressure-pitch controller (based on the feedback error learning) to improve the sound produced by the robot during a musical performance was proposed and implemented. In both cases studied, a set of experiments are described to verify the improvements achieved while considering biologically-inspired control approaches.

Keywords Biologically-inspired Robotics, Feedback Error Learning Control, Music

1. Introduction

The development of wind instrument playing anthropomorphic robots has interested researchers ever since the golden era of automata. As an example, we may find some classic examples of automata displaying human-like motor dexterity to play instruments such as the “Flute Player” [1]. More recently, we find the first attempt to develop an anthropomorphic musical robot, the WABOT-2. The WABOT-2 was capable of playing a concert organ, built by the late Prof. Ichiro Kato. In particular, Prof. Kato argued that artistic activity such as playing a keyboard instrument would require human-like intelligence and dexterity [2].
Compared to other kinds of instruments, research into wind instruments has interested researchers from the point of view of human science, motor learning control, musical engineering, etc. Thanks to the advances in electronics, computers science, etc., different musical-instrument playing robots have been developed to produce live performances by enhancing their dexterity and perceptual capabilities. Nowadays, research into musical robots opens up the opportunity of studying several aspects of humans, such as understanding human motor control, how humans communicate ideas, finding new ways of musical expression, etc. As a result, different kinds of wind instrument-playing automated machines and anthropomorphic robots have been developed [3-9]. Other researchers have focused on analysing playing wind-instruments from a musical engineering approach by performing experiments with simplified mechanisms [10-11] and from a physiological perspective by analysing medical imaging data from professional players [12-13].

The authors deal particularly with the development of an anthropomorphic flutist robot and a saxophone-playing robot designed to mechanically emulate the required organs during the flutist and saxophone playing. Due to the interdisciplinary nature of this research, our collaboration with musicians, musical engineers and medical doctors has been of vital importance for better reproducing and understanding human motor control from an engineering perspective.

One of the first attempts to develop a flute playing robot came during the golden era of automata by Jacques de Vaucanson. He designed and constructed a “Flute Player” as a means to understand the human breathing mechanism. The design principle for the “Flute Player” was that every single mechanism corresponded to every muscle [1]. Thus, Vaucanson arrived at those sounds by mimicking the very means by which a man would make them.

One of the first attempts to develop a saxophone-playing robot was by Takashima at Hosei University [6]. This robot, named APR-SX2, is composed of three main components: a mouth mechanism, the air supply mechanism, and fingers. The artificial mouth consisted of flexible artificial lips and a reed pressing mechanism. The artificial lips were made of a rubber balloon filled with silicon oil with the proper viscosity. The air supplying system consists of an air pump and a diffuser tank with a pressure control system. The APRSx2 was designed under the principle that the instrument played by the robot should not be changed. A finger mechanism was designed to play the saxophone’s keys (actuated by solenoids), and a modified mouth mechanism was designed to attach it to the mouthpiece, and no tonguing mechanism was implemented. The control system implemented for the APR-SX2 is composed of one computer dedicated to the control of the key-fingering, air pressure and flow, pitch of the tones, tonguing, and pitch bending. In order to synchronize all the performances, the musical data were sent to the control computer through a Musical Instrument Digital Interface (MIDI) in real-time. In particular, the Standard MIDI File (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). Hosei University developed the APR-SX2, the design of which was based on the concept of reproducing melodies on a tenor saxophone. Therefore, the saxophone playing robot was developed under the condition that the musical instrument played by robots should not be changed or remodelled at all. However, a total of 23 fingers were used to play the saxophone’s keys, a modified mouth mechanism was designed to attach it to the mouthpiece, and no tonguing mechanism was implemented.

The authors have developed the Waseda Flutist Robot No. 4 Refined IV (WF-4RIV) [14]. The WF-4RIV has a total of 41 DOFs, which mechanically simulate the human organs involved during the playing of the flute. The WF-4RIV mechanically reproduced the anatomy and physiology of the following organs: lips (three DOFs), neck (four DOFs), lungs and valve mechanism (two DOFs and one DOF, respectively), fingers (12 DOFs), throat (one DOF), tonguing (one DOF), two arms (each with seven DOFs) and eyes (three DOFs). The WF-4RIV has a height of 1.7 m and a weight of 150 kg. In addition, in [15], the authors proposed as a long-term goal enabling the flutist robot to interact more naturally with musical partners in the context of a Jazz band by implementing a Musical-Based Interaction System (MBIS).

The authors have also developed the Waseda Saxophonist Robot No. 2 Refined (WAS-2R). The WAS-2R is composed of 22 DOFs that reproduce the physiology and anatomy of the organs involved during the playing of a saxophone as follows [16]: three DOFs to control the shape of the artificial lips, 16 DOFs for the human-like hand, one DOF for the tonguing mechanism and two DOFs for the lung system. In addition, to improve the stability of the pitch of the sound produced, a pressure-pitch controller system was implemented.

In this review article, we focus on describing in detail the biologically-inspired control architectures proposed and implemented in the Waseda Flutist Robot No. 4 Refined VI (WF-4RVI) and the Waseda Saxophonist Robot No.2 Refined III (WAS-2RIII). In Section 2, an overview of the mechanism design and control system for the WF-4RVI is described. In Section 3, an overview of the mechanism design and control system for the WAS-2RIII is described.
Finally, in Section 4, a set of experiments is presented in order to verify the improvements achieved with the proposed biologically-inspired control architectures implemented in each case studied.

2. The Waseda Flutist Robot No. 4 Refined VI

2.1 Mechanism Overview

In [17], the Waseda Flutist Robot No.4 Refined VI (WF-4RVI) was introduced. The WF-4RVI is composed of 41 DOFs that reproduce the physiology and anatomy of the organs involved during the flute playing as follows (Figure 1): lips, lungs, arms, neck, oral cavity, vocal cord, tongue, fingers and eyes. The previous version of the flutist robot was characterized by a low conversion efficiency ratio and the appearance of husky sounds due to a speed reduction of the air flow inside the oral cavity and therefore the air flow is partially disrupted. Therefore, we have focused on redesigning the oral cavity without increasing the complexity of the mechanism.

![Figure 1. The Waseda Flutist Robot No. 4 Refined VI [17]](image)

In order to re-design the oral cavity, MR images from two professional flutist players (one man and one woman) were obtained [17]. For this experiment, the subjects were asked to hold the dummy flute and simulate playing low-pitch and high-pitch notes and for each of cases they played two types of tones (low and high). Therefore, a total of four types of cases were registered in the coronal direction. Furthermore, a mouthpiece made of SEPTON was also developed and the subjects were asked to attach the mouthpiece to their tooth. From this, four different types of cases were also registered in the sagittal direction.

The 3D model of the tooth was obtained by using the volume rendering function of the software 3D Mimics. From this a 3D model of the oral cavity is obtained for each of the registered cases. By using 3D Mimics functions, the cross-sectional centreline is obtained and the minimal cross-section of the 3D model is obtained. In addition, the rate of change of the cross-sectional area respective to the position of the lip and the cross-sectional area respective to the tip is identified so that the approximate shape of the cross section of an ellipse is drawn. By obtaining the circumscribed circle radius and the maximum radius of the inscribed circle, we have to evaluate the change from the position of the cross-sectional shape of the lips and the tip radius (both long and short radius). Regarding the changes of the cross-sectional area of the female professional flutist player, no significant differences were detected in the collected data (Figure 2a). Regarding the changes of the cross-sectional area of the male player, a difference was detected in the collected data (Figure 2b).

![Figure 2. Experimental results of the evaluation of the cross-section area: a) women subject; male subject [17]](image)

The 3D model of the oral cavity shape obtained by using the volume rendering function of the 3D mimics from the MR images from a male and female professional player [17].
From these, a total of four different types of prototypes for the oral cavity shape were obtained from the male subject and one prototype based on the average from the female one (Figure 3). Based on the experimental results [17], the prototype labelled m-HF shows the best results both in terms of air conversion efficiency ratio and sound evaluation function score. Therefore, the new oral cavity and lip of the WF-4RVI were constructed and integrated into the flutist robot.

2.2 Auditory Feedback Control System

In order to search for the best set of parameters to control the mechanically simulated organs of the flutist robot, two approaches can be considered: by manually fixing the parameters or by analysing a human’s performance. Of course, the first option may be the easiest to implement, but it will be take a long time to find those parameters. In the literature, we basically find two approaches for the calibration procedure [18]: model-based or non-parametric kinematic calibration. Model-based kinematics calibration is based on an error model representing the relationship between the error of the kinematics parameters and the end-effector pose error [19-20]. However, the process is complex and high precision measurement devices are required [18]. Non-parametric kinematics calibration uses intelligent algorithms to represent the relationship between Cartesian space and joint space instead of the formulas established by kinematics parameters [21-22]. The benefit is that the position of the end-effector of the robot can be compensated directly, which can simplify the calibration process. However, the calibration efficiency and accuracy are limited by the arrangement of intelligent algorithms and the calibration space is limited by the convergence ability of intelligent algorithms [18].

Due to the fact that both approaches present difficulties for the implementation in the flutist robot, in order to improve the calibration procedure of the robot in order to play all the notes correctly during a performance, a biologically-inspired control approach has been considered. In particular, if we consider the way a professional flutist prepares for a musical performance, they use a basic position in which he/she can easily create all the tones, only by changing the shape of his/her lips and the speed of his/her breath. This position, called the general position, is found before the performance and it is implemented by adjusting the parameters of the lips and lungs while continuously blowing a simple etude until all the notes are produced with a uniform sound quality [24]. Inspired by this principle, the flutist robot also performs a calibration phase in order to find the general position.

For this purpose, the authors have proposed and implemented an auditory feedback control system for the flutist robot, which is composed of three main modules (Figure 4): feed-forward air pressure control, Expressive Music Generator and Pitch Evaluation System.

2.2.1 Expressive Music Generator

Up to now, the implementation of AI approaches has been demonstrated as being capable of generating high-quality human-like monophonic performances based on the examples of human performers. However, all of these systems have only been tested by computer systems or MIDI-enabled instruments, which limit the unique experience of a live performance. Therefore, the authors proposed implementing an AI approach to the performance control system of the flutist robot to take advantage of a live performance. In particular, a feed-forward neural network was implemented to model the musical expressiveness from a performance of a professional player [23]. From such a model, a set of musical performance rules can be created that then can be used by the flutist robot to generate an expressive performance even from a different nominal score. Basically, a musical score is defined by the note information and its duration. By defining these two parameters, it is possible to play any kind of instrument. However, the sound becomes monotonous. Therefore, human players add expressiveness to the performance by adding other parameters such as vibrato, staccato (tonguing), dynamic marks, slurs, etc.

Up until now, there has been no widely accepted scientific method to describe how human players add such kinds of parameters to express emotions, feelings, etc. For this purpose, we have implemented an Expressive Music Generator (ExMG) that has been designed to produce the musical information required to produce an expressive performance [18]. As an input, the ExMG uses the musical parameters (i.e., pitch, volume, tempo, etc.) from the performance of a professional flutist. Those parameters are analysed and extracted by using our fast Fourier transform (FFT) tool [14]. As an output, a set of musical performance rules (which define the deviations introduced by the performer) are produced (offline). The process of modelling the expressive features of the flute performance is done by using neural networks. In particular, we have considered the note duration, vibrato frequency and vibrato duration, attack time, and tonguing as part of the musical performance rules. The resulting musical information output is then sent

![Image](313x639 to 535x711)
directly to the robot’s control system by sending MIDI messages from a sequencer device (PC). As a result, the flutist robot is capable of performing a musical performance with expressiveness.

In order to create an expressive performance, a feedforward neural network trained with the error back propagation algorithm was implemented using Borland Builder C++. Feed-forward neural networks are the most widely used models in many practical applications [25].

The network considered for our purpose is a three layered neural network that consists of $n$ input units, $m$ hidden units and $k$ output units. Each input unit $j$ ($j = 1, 2...n$) accepts an input signal $x_j$ and transmits it to hidden units. The output $s_i$ of the hidden unit $i$ is obtained by a weighted summation of the inputs, which are then passed through a nonlinear function know as an activation or transfer function. The output function $F$ is a sigmoid function defined as (2). As a result, we can compute the output of the network $y_k$ by (3).

$$s_i = F\left(\sum_{j=0}^{n} v_{ij}x_j\right) \quad (1)$$

$$F(s) = \frac{1}{1 + e^{-s}} \quad (2)$$

$$y_k = F\left(\sum_{i=0}^{m} w_{ki}s_i\right) \quad (3)$$

In our case, the artificial neural network (ANN) was implemented as follows. Our proposal is to train three different networks to model the expressiveness of the performance. For this purpose, three different musical parameters were considered: duration rate, vibrato frequency and vibrato amplitude. For each of the networks, a number of inputs were defined based on [26] as follows:

- Duration rate (14 inputs): slur, crescendo, decrescendo, normalized pitch, pitch difference, note, note time, before breath, after breath, second beat, note from down to up, note accessibility, note up, note down.
- Vibrato duration (19 inputs): slur, slur on, tie, crescendo, decrescendo, normalized pitch, note, note time, before breath, after breath, first beat, second beat, third beat, note length change, pitch change from down to up, note accessibility, note up, note down, pitch leap.
- Vibrato frequency (18 inputs): crescendo decrescendo, normalized pitch, note, note time, before breath, after breath, first beat, second beat, third beat, note length change, pitch change from down to up, note accessibility, note up, note down, pitch leap.

We have experimentally determined different numbers of hidden layers for each of them. In particular, seven units were defined for the duration rate, five units for the vibrato frequency and 17 units for the vibrato amplitude. In order to train the ANN, a back-propagation algorithm was considered. This kind of supervised learning incorporates an external teaching signal (the performance of a professional flutist has been used), so that each output unit is told what its desired response to input signals ought to be. During learning, the weight vectors ($w$) are updated using (4), where $E(t)$ is the error between the output value (5) and the desired one, and $\eta$ is the learning rate. $E(t)$ is basically computed by using the difference between the actual output $y_k$ and the teaching signal $d_k$. On the other hand, the coefficient $\eta$ affects the network teaching speed. There are a few techniques for selecting this parameter so that in this paper we have experimentally determined $\eta = 0.75$. The ANN was trained to learn the extracted performance rules obtained from the analysis of the professional flutist’s performance. Based on the previous setting of the ANN, the duration rate converges at 148 steps, the vibrato frequency at 72 steps and the vibrato amplitude at 49 steps. It is worth mentioning that in our implementation, in order to prevent over-fitting (one of the critical issues while designing a neural network is the generalization), we have not considered that by defining a small number of hidden layer units and by limiting the number of leaning steps (less than 10 000).

$$w_i(t+1) = w_i(t) - \eta \left(\frac{\delta E(t)}{\delta w_i}\right) \quad (4)$$

$$E(t) = \frac{1}{2} \sum_{j=0}^{n} (y_k(t) - d_k(t))^2 \quad (5)$$

![Block diagram of the feedback error learning used to determine the inverse model](image)

**Figure 5.** Block diagram of the feedback error learning used to determine the inverse model [25]

### 2.2.2 Feed-forward Air Pressure Control

As a first approach, we proposed to implement the feedforward control system by means of the feedback error learning [18]. The feedback error learning (FEL) is a computational theory of supervised motor learning proposed by Kawato et al. [27], which is inspired by the central nervous system. In addition, Kawato and Gomi
considered that the cerebellum, by learning, acquires an internal model of the inverse dynamics of the controlled object [28]. From this extension, the feedback error learning can also be used as a training signal to acquire the inverse dynamics model of the controlled system (Figure 5). In our case, the teaching signal is the data collected from a pressure sensor placed at the simulated lungs. In our case, \( y(t) \) is defined as the air flow velocity in the lungs and \( u(t) \) is defined as the air pressure [23].

### 2.2.3 Pitch Evaluation System

In [29], we proposed a method to analyse the robot’s performance based on the experimental results of Ando [30] who analysed the acoustical properties of the flute sound by using a mechanical blowing apparatus. In order to evaluate the flute sound produced by the robot, we have developed the Pitch Evaluation System [23]. The Pitch Evaluation System (PES) was designed to detect the pitch of the flute sound as well as to evaluate its quality. The PES is designed to estimate the pitch or fundamental frequency of a musical note. For this purpose, as a first approach, we have considered implementing the Cepstrum method. This method is probably the most popular pitch-tracking method in speech as it can be computed in real-time. First, the Cepstrum is calculated by taking the Fourier transform (STFT) of the log of the magnitude spectrum of the sound frame. In our case, the sound is sampled at 44.1 Hz, where the frame size was defined as 2048 with 50% overlapping (Hanning window) so that the time resolution is 23 ms. Then, the Cepstrum is inspected for a peak by dividing the frame into bins, corresponding to the period of the signal, as well as for a second or third peak an equal distance (period) away from the first or second peak. However, this method presents the problem of deciding how to divide the frequency. Therefore, we proposed implementing the Cepstrum method and synchronizing with the MIDI data of the score.

This approach provides information to the pitch detection algorithm about where the pitch is supposed to be located. From this analysis, we were able to analyse the pitch (by tracking the maximum amplitude of each FFT frame), the sound intensity level, the harmonic structure content (by localizing the harmonics, semi-harmonics, and even- and odd-harmonics level) and, finally, integrating this information into the sound quality evaluation function (6), where \( M \) and \( H \) are the harmonic and semi-harmonic level; \( L_e \) and \( L_o \) are the even-harmonics and odd-harmonics level; and \( SIL \) is the sound intensity level that can be calculated by using (7).

\[
\text{EvalF} = \frac{(M - H) + (L_e - L_o)}{\text{SIL}} \quad (6)
\]

\[
\text{SIL} = 10 \log_{10} \left( \frac{\sum_{i=1}^{N} |\text{FFT}_i|^2}{N} \cdot \frac{1}{4\pi^2 \cdot I_0} \right) \quad (7)
\]

Basically, by using (7), the flutist robot is capable of detecting when a note is incorrectly played. During a performance, the musical information is stored in a database where the pitch, volume, etc., are indicated. After each note is played by the robot, the sound quality is determined (the \( \text{EvalF} \) range is from 0 to 10). When the PES detects a note, the sound quality of which is lower than a threshold (less than 6), a fault flag is produced as an input within the database to all the notes that appear with the same pitch. In order to improve the sound quality of the flute sound, in case the same note should be played again, we change the parameters of the robot (e.g., one DOF of the lungs, three DOFs of the upper lip, and one DOF of the lower lip; etc.). In the field of robotics research, different optimization techniques have been proposed to find the set of parameters that improve the performance of the robot (e.g. [31-33]; etc.).

As described in the previous sections, we have proposed that the flutist robot determine those parameters by finding the general position. For this purpose, the principle of searching for an orthogonal table was considered. The orthogonal table is described in [24] in order to find the best suitable set of parameters to play all the notes using the same lip shape and lung velocity that can be found. The proposed algorithm is as follows [24]: at first, a phrase is selected and performed by the robot using a set of initial blowing parameters (home position). The parameters used for this algorithm are: one DOF of the lungs, three DOFs of the upper lip (inside, outside and forward), and one DOF of the lower lip (inside). The remaining degree of freedom of the lower lip (forward) cannot be changed as this part should always remain in contact with the flute. Furthermore, the position of the arms, fingers, and neck are supposed to be unique by defining a home position of these components.

3. Waseda Saxophonist Robot No. 2 Refined III

3.1 Mechanism Overview

In [5], the Waseda Saxophonist Robot No. 2 Refined III (WAS-2RIII) was introduced. The WAS-2RIII has improved the air compressor design for reducing the effect of the ripple to improve the sound quality.
Furthermore, an eye mechanism was designed to enable the saxophonist robot to acquire visual information from the musical partner. In particular, the WAS-2RIII is composed of 27 DOFs that reproduce the physiology and anatomy of the organs involved during the playing of the saxophone as follows (Figure 5): two DOFs to control the shape of the artificial lips, one DOF for the control of the shape of the oral cavity, one DOF for the tongue control, 19 DOFs for the human-like hand, two DOFs for the lung system and two DOFs for the eye mechanism.

3.2 Pressure-Pitch Control System

In a previous research study [34], a feed-forward air pressure controller with dead-time compensation was implemented to ensure the accurate control of the air pressure during the attack time. Moreover, for the control of the finger mechanism, an open-loop controller was implemented. In particular, the feedback error learning during the attack phase of the sound was used to create the inverse dynamics model of the Multiple-Input Single-Output (MISO) controlled system based on Artificial Neural Networks (ANN). In addition, an Overblowing Correction Controller (OCC) was proposed and implemented in order to ensure 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. However, we still detect deviations on the pitch while playing the saxophone [34].

It is known that the performance of a robot can be improved by including the robot dynamics into its controller. However, the complexity and the lack of knowledge about the dynamic parameters of the robot, lead robots to be controlled mostly by PID control, where the control is done independently for each joint [35]. On the other hand, in order to deal with the uncertainties of complex robot systems (e.g., the saxophone robot), two nonlinear control methods have been popular [36]: adaptive control and deterministic robust control. The adaptive control achieves asymptotic tracking for reasonably large classes of nonlinear systems without using discontinuous or infinite-gain feedback terms (e.g., [37], etc.). However, adaptive controllers only deal with the ideal case of constant parametric uncertainties and the adaptation law may lose stability even when a small disturbance appears. In contrast, the deterministic robust control can be used to achieve a guaranteed transient performance and a guaranteed final tracking accuracy in the presence of uncertainties (e.g. [38], etc.). However; it usually involves switching or infinite-gain feedback terms, which introduces control chattering.

Due to the fact that both approaches present difficulties for the implementation in the saxophone robot, in order to improve the accuracy of the control of the pressure and pitch during a performance, a biologically-inspired control approach has been considered. In particular, in order to improve the accuracy of the control in [16], we have proposed the implementation of the control system shown in Figure 6. In particular, the proposed control system includes a dead-time compensation controller for the finger mechanism (to reduce the effect of response delay due to the wire-driven mechanism) and a Pressure-Pitch Controller (PPC) for the control of the valve and lip mechanism (to assure the accurate control of the pitch).

Regarding the implementation of the dead-time compensation control (which uses an element to predict how changes made now by the controller will affect the controlled variable in the future [39]), for each finger of WAS-2R, the pressing time of the saxophone’s key is measured by means of the embedded potentiometer sensor (defined as L_N; where N represents the total number of DOF designed for the finger mechanism). By including the dead-time factor (referred as e^{-st}), it is possible to compensate for the finger’s response delay during the playing of the saxophone.

Due to the fact that both approaches present difficulties for the implementation in the saxophone robot, in order to improve the accuracy of the control of the pressure and pitch during a performance, a biologically-inspired control approach has been considered. In particular, in order to improve the accuracy of the control in [16], we have proposed the implementation of the control system shown in Figure 6. In particular, the proposed control system includes a dead-time compensation controller for the finger mechanism (to reduce the effect of response delay due to the wire-driven mechanism) and a Pressure-Pitch Controller (PPC) for the control of the valve and lip mechanism (to assure the accurate control of the pitch).

Due to the fact that both approaches present difficulties for the implementation in the saxophone robot, in order to improve the accuracy of the control of the pressure and pitch during a performance, a biologically-inspired control approach has been considered. In particular, in order to improve the accuracy of the control in [16], we have proposed the implementation of the control system shown in Figure 6. In particular, the proposed control system includes a dead-time compensation controller for the finger mechanism (to reduce the effect of response delay due to the wire-driven mechanism) and a Pressure-Pitch Controller (PPC) for the control of the valve and lip mechanism (to assure the accurate control of the pitch).
On the other hand, regarding the pressure-pitch controller during the sustain phase of the sound, a Multiple-Input Multiple-Output (MIMO) system has been proposed not only to ensure the accurate control of the air pressure during the attack phase of the sound, but also to ensure the accurate control of both air pressure and sound pitch during the sustain phase of the sound. For this purpose, we implemented a feedback error learning method [28] to create the inverse model of the proposed MIMO system which is computed by means of an ANN. During the training process (Figure 7), the inputs of the ANN are defined as follows: pressure reference (PressureREF), pitch reference (PitchREF). 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 (ΔValve) and lower lip (ΔLip) are controllers to ensure the accurate control of the required air pressure and pitch to produce the saxophone sound. Moreover, during the training phase, the air pressure (PressureRES) and sound pitch (PitchRES) are used as feedback signals and both outputs from the feedback controller are used as teaching signals to effectively train the ANN. As a result of the training phase, during a saxophone playing performance, the created inverse model is used.

4. Experiments and Results

4.1 Evaluation of the auditory feedback system for the flutist robot

In order to verify the effectiveness of the proposed auditory feedback system, we have proposed a set of experiments with the flutist robot [23]. For this purpose, the flutist robot was programmed to perform three different scores with different levels of difficulty: Le Cygne (beginner level), Flute Quartet kv. 298 (intermediate level) and Polonaise (advanced level). At first, we focused on verifying the effectiveness of the proposed auditory feedback control system to enhance the quality of the produced flute sound and to adjust the parameters of the robot when an incorrect note is detected by the PES. In particular, the quality of the sound was determined by using (6). The experimental results are shown in Figure 8. Based on the results, the average number of notes produced of high quality was higher with the proposed auditory system. The collected data were also subjected to a two-ANOVA statistical analysis to compare the differences between the control methods (open-loop and auditory feedback system) and the differences between the melodies (Le Cygne, Flute Quartet kv. 298 and Polonaise). As a result, the effect of the control system on the performance is considered significant ($P = 0.0340$). On the other hand, the effect of different levels of melodies on the performance of the robot is considered not quite significant ($P = 0.0663$). Furthermore, in Table 1, the number of incorrect notes (low sound quality) detected by PES are shown. As we may observe, the flutist robot enhanced the quality of its performance while using the proposed auditory feedback system.

In order to confirm our experimental results, we asked 14 subjects to listen to the recordings of the performances of the robot while using the open-loop based system and the auditory feedback based system [18]. Each of the subjects was asked to evaluate from one to 10 the performance of the flutist robot while comparing the performance of a professional flutist. As a result, subjects evaluated with an average of 3.9 to the performance with the open-loop based system and an average of 6.1 to the performance with the auditory feedback based system.

In order to determine the effectiveness of the proposed pressure-pitch controller in reducing the pitch deviations while playing the saxophone, we programmed the saxophone-playing robot to play the main theme of the “Moonlight Serenade” composed by Glenn Miller before and after training the inverse model [16]. In particular, as for the neural network parameters, a total of six hidden units were used. For the training process, a total of 144 steps were made. The experimental results are shown in Figure 9; where 1[cent] is defined as (equi-tempered
As we could observe, the deviations of the pitch after the training (Standard Error is 41.7) are considerably less than before training (Standard Error is 2372.8).

Figure 9. Experimental results while comparing the deviations of the pitch before and after training the inverse model of the proposed MIMO system with the saxophone-playing robot [16]

On the other hand, in order to determine the improvements in the performance level of the saxophone-playing robot, we proposed to carry out a qualitative performance evaluation [16]. For this purpose, we asked three professional saxophone players to evaluate and compare the performance of the saxophone-playing robot with respect to the previous version. In particular, the audio recordings from both robots as well as a professional one were carried out while playing the “Moonlight Serenade”. For this purpose, each subject was requested to evaluate each performance (presented randomly without providing information on the player’s source) while considering the following parameters: attach phase, crescendo and decrescendo effects and overall performance. Each of the evaluation parameters were evaluated from a range of 0 to 10 (where the highest value belongs to a recording from a professional one). The experimental results are shown in Figure 10. As we may observe, all the subjects evaluated with a higher score the performance of the saxophone-playing robot with respect to the previous version.

Figure 10. Experimental results while qualitatively comparing the performance of the saxophone-playing robot and the previous version by three professional saxophone players [16]

As a future work, a modified version of the feedback error learning will be proposed by including the sound evaluation quality score in order to improve the performance of the proposed control system in both cases.

5. Conclusion and Future Work

In this paper, the design of two different biologically-inspired control architectures proposed and implemented for musical performance robots have been described. In particular, the details of the proposed auditory feedback system, and a pressure-pitch controller implemented into the flutist robot and the saxophone-playing robot respectively have been given. A set of experiments were carried out in order to verify the effectiveness of the proposed methods. From the experimental results, we were able to verify the improvements, based on the proposed biologically-inspired control approaches, achieved on the calibration control of the flutist robot (a significant difference was observed with respect to the open-loop control approach) and the performance control of the saxophonist robot (less deviation of the pitch was detected after the training procedure).

6. Acknowledgments

Part of the research on Musical Instrument-Playing Robot was done at the Humanoid Robotics Institute (HRI), Waseda University and at the Center for Advanced Biomedical Sciences (TWINs). This study is supported (in part) by a Grant-in-Aid for Young Scientists (B) provided by the Japanese Ministry of Education, Culture, Sports, Science and Technology, No. 23700238 (Jorge Solis, PI).

7. References


