Robo-Sax Quartet: A Semi-Automatic Robotic Saxophone System for Augmented Ensemble Performance

Gou Koutaki
koutaki@cs.kumamoto-u.ac.jp
Kumamoto University
Japan

Masatoshi Hamanaka
masatoshi.hamanaka@riken.jp
Riken
Japan

Figure 1: Overview of the robotic saxophone quartet. The left panel shows the four robotic saxophones—soprano, alto, tenor, and baritone—developed in this work. The right panel illustrates a live ensemble performance, where human players perform a saxophone quartet using the robotic instruments.

Abstract

We present a semi-automatic robotic saxophone system that assists saxophone key operations using servo-motor–driven actuators. The system augments conventional acoustic saxophones by supporting fingering actions while leaving breath control and musical expression to the human performer. We developed four robotic saxophones—soprano, alto, tenor, and baritone—which can be performed simultaneously to enable a robot-assisted saxophone quartet. Due to differences in instrument size, key layout, and performer posture across the four saxophone types, the placement and mechanism of actuators required careful design. Through iterative prototyping and experimentation, we identified suitable actuator configurations for each instrument. To support intuitive performance, we developed a dedicated music-game–style graphical user interface, enabling wireless communication and battery-powered operation. This design allows performers to handle the instruments in a manner comparable to conventional saxophone performance. The robotic saxophone offers the advantage of reliable and error-free execution of rapid fingering patterns that are challenging for human players. Through performance demonstrations, we show that the proposed system achieves practical usability as an augmented musical instrument.

Keywords

Sax, semi-automatic musical instrumental robot, music ensemble system

---

NIME '26, June 23–26, 2026, London, UK

1 Introduction

The saxophone is a widely recognized wind instrument, valued for its expressive sound and broad use across musical genres. Although many people are motivated to learn the saxophone, performance is widely regarded as challenging for beginners. This difficulty stems from several factors: the need to memorize complex fingering patterns involving approximately 32 keys, the coordination of fingering with musical notation and rhythm, and the precise control of breath and embouchure required to produce sound. Together, these cognitive, motor, and physiological demands create a steep learning curve that often discourages novice players.

To address these challenges, we have been developing a semi-automatic robotic saxophone that supports fingering and score-related operations through a robotic system, while leaving breath input and expressive control to the human performer. By delegating complex key operations and timing control to the robot, performers can focus solely on supplying breath at appropriate moments, enabling musical performance without mastering traditional fingering techniques or music reading skills. This approach allows beginners to experience advanced performance at an early stage and to concentrate their practice on breath control and embouchure, which are central to musical expressivity.

In this paper, we extend the robotic saxophone system to a robot-assisted saxophone quartet consisting of soprano, alto, tenor, and baritone saxophones. Because these instruments differ in size, key layout, required actuation force, and playing posture, instrument-specific robotic mechanisms are required to ensure sufficient power while avoiding physical interference with the performer. We present optimized designs for each instrument and address ensemble-specific challenges such as communication, synchronization, and compensation for mechanical delays.

Mechanical delays are measured and incorporated into MIDI-based control data to achieve coordinated ensemble performance. Finally, we demonstrate the practical feasibility of the proposed system through a public concert featuring a robo-sax quartet, confirming its suitability for real-world musical performance.

Our contributions are summarized as follows:

- We extend a semi-automatic robotic saxophone framework to a multi-instrument ensemble, realizing a robot-assisted saxophone quartet that integrates human breath control with robotic key actuation.
- We design and implement instrument-specific robotic mechanisms that address physical constraints across different saxophone types, including size, key configuration, actuation force, and performer posture.
- We propose a delay-compensated synchronization scheme for robotic ensemble performance based on measured mechanical delays, and validate the system through a public concert demonstration.

2 Related Work

Research on automated musical instruments spans a long history and can be broadly categorized into *fully automatic* systems, in which a machine performs music independently, and *semi-automatic* systems, in which a human performer remains an active participant. This section reviews representative work in both categories and positions our approach within the context of semi-automatic, human-in-the-loop musical instruments.

2.1 Fully Automatic Musical Instruments

Early examples of fully automatic musical instruments date back more than two centuries. Mechanical music boxes [4] and hand-cranked organs encoded musical sequences using mechanical media such as cylinders and punch cards. In the early twentieth century, more elaborate systems such as the Hupfeld Phonoliszt Violina [2] were developed, enabling automatic performance on instruments including the piano and violin. These systems can be understood as early mechanical music sequencers.

From the 1960s onward, mechanical sequencing technologies were largely replaced by electronic sequencers, which became widespread alongside synthesizers and sound modules. Parallel to this development, research emerged on robots capable of performing *acoustic* musical instruments rather than electronic ones. Since the 1980s, robots capable of playing instruments such as the violin, piano, and guitar have been reported [1, 3, 6, 8, 12], and fully automatic pianos have since been commercialized.

Since the 2000s, humanoid robots capable of playing wind instruments such as the trumpet, flute, and saxophone have also been actively studied [7, 10, 11, 15]. In these systems, key or valve operations are typically performed by robotic fingers, while sound production is achieved through artificial lips and lung mechanisms. Although such robots demonstrate impressive technical achievements, their primary role is often demonstrative or performative, offering audiences a passive experience of listening to or observing a robot's performance.

2.2 Semi-Automatic Musical Instruments

In contrast to fully automated systems, semi-automatic musical instruments aim to support human performance rather than replace it. These systems are often motivated by pedagogical goals, accessibility, or the desire to lower technical barriers to musical expression.

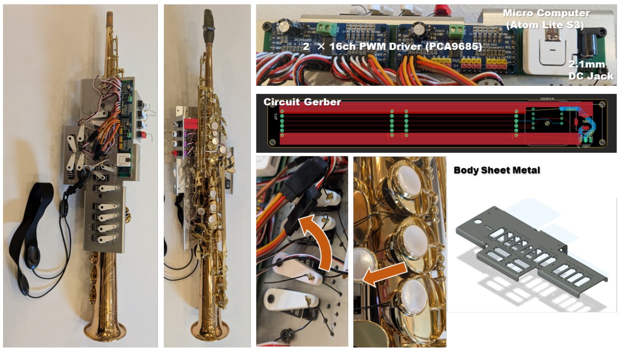


Figure 2: Robo-soprano saxophone with an attached robotic key-actuation module. Nineteen servo motors and a compact control circuit are integrated into a high-rigidity anodized aluminum body, enabling reliable key actuation

Various assistive devices have long been used in musical practice, such as mechanisms that simplify chord fingering on the guitar, support violin bowing, or attach to the mouthpiece of a flute to aid sound production. One-handed recorders and other modified instruments have also been developed to accommodate performers with physical constraints.

More recent research has explored robotic and computational approaches to performance assistance [9, 13, 14]. Some systems employ actuators to press guitar strings or assist with flute fingering [5], while leaving expressive control, timing, and sound production to the human performer. These approaches emphasize the performer's agency and embodiment.

Because the human remains the primary musical actor, semi-automatic instruments offer an active and participatory musical experience, in contrast to the largely passive consumption associated with fully automated performance robots. Our work builds on this line of research by exploring a semi-automatic robotic saxophone system that supports complex key operations while preserving human breath input and expressive control, and further extends this concept to ensemble performance.

3 Robo-Sax System

Figure 2 shows the developed robo-soprano saxophone. The robo-sax system consists of a high-rigidity sheet-metal body on which 19 servo motors (A20BHS) and a control circuit are mounted to an actual soprano sax (Yanagisawa SWO2). The robot assists saxophone key operations while allowing the performer to play the instrument in a conventional manner. In this section, we describe the mechanical structure, control electronics, actuator selection, key actuation mechanism, and overall weight of the system.

3.1 Body Design and Mechanical Properties

Because the servo motors exert relatively high forces to actuate saxophone keys, the robot body must have sufficient rigidity to withstand these forces without deformation. Although steel or stainless steel would provide high stiffness, their weight makes them unsuitable for a wearable musical instrument. Therefore, the body was fabricated from aluminum alloy A5052 sheet metal with a thickness of 2 mm. The total weight of the body plate is 195 g.

To ensure electrical safety, the surface of the aluminum body is anodized. This surface treatment is essential to prevent unintended electrical conduction and short circuits between the metal body, the control electronics, and the servo power lines. Because aluminum has lower stiffness than steel, a thickness of 2 mm was selected, and the edges of the sheet metal were bent to form a beam-like structure, thereby increasing structural rigidity.

The robot body is fixed to the saxophone using two M4 spacers. One spacer is attached to the existing threaded hole of the thumb hook, which is originally intended for the performer's right thumb. The second spacer is fixed near the upper octave key, at the center of the thumb rest. Initial attempts involved drilling holes or bonding the spacer using epoxy resin; however, these methods frequently failed due to mechanical shocks during transportation. To achieve sufficient durability, the spacer was soldered directly to the thumb rest. Because a soldering iron could not provide adequate heat, the saxophone keys were temporarily disassembled and the joint was soldered using a gas burner.

The robotic system is designed to be removable; once detached, the saxophone can be played as a conventional acoustic instrument. Small mounting jigs (approximately 1 cm^2) are permanently affixed to the instrument body at several attachment points, but these do not interfere with normal playing posture or technique.

3.2 Control Circuit

The control circuit is shown in the upper-right part of Figure 2. The main microcontroller is an M5Atom Lite S3, which receives MIDI data and controls the servo motors accordingly. Servo positions are controlled via pulse-width modulation (PWM). PWM signals are generated using two PCA9685 16-channel PWM driver ICs, which communicate with the microcontroller via serial communication.

Securing sufficient power for the servo motors is a critical design issue. Each servo can consume up to 0.5 A at peak load; therefore, driving 19 servos simultaneously may require instantaneous currents approaching 10 A. Although such peak currents rarely occur in practice, the power supply and circuit layout must be designed to handle these transient loads. Accordingly, the printed circuit board was designed with wide power and ground traces on both sides of the board, and a copper thickness of 2 oz was selected to reduce voltage drop and heat generation.

3.3 Actuator Selection

The selection of servo motors is crucial for achieving reliable and expressive key actuation. The actuators must be compact and lightweight while providing sufficient torque and high-speed response. We evaluated nine different servo motors from various manufacturers. As a result, the brushless micro servo motor A20BHS was selected.

According to the manufacturer's specifications, when driven at 7.4 V, the A20BHS provides an operating speed of 0.075 s/60°, a torque of 10.20 kg·cm, and a weight of 23.5 g. For comparison, the commonly used SG92R servo has a speed of 0.1 s/60° degrees and a torque of 2.5 kg·cm, which proved insufficient for reliably pressing saxophone keys. Some servos with superior nominal specifications were also tested; however, they were rejected due to excessive operating noise, high-frequency electrical noise, or unacceptable sound during both motion and stall conditions.

Long-term durability is also an important consideration. During repeated use, SG92R servos frequently experienced motor burnout, whereas the A20BHS servos have operated continuously for over one year without a single failure.

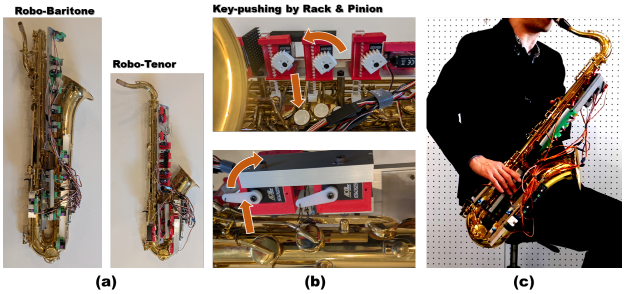


Figure 3: Robo-tenor and robo-baritone saxophones for ensemble performance. (a) The developed robotic saxophones with instrument-specific actuation mechanisms. (b) Rack-and-pinion-based key actuation mechanism designed to generate sufficient force for large keys. (c) A performance scene showing embodied use of the robo-tenor saxophone without physical interference, preserving conventional playing posture and breath-based expressivity.

3.4 Key Actuation Mechanism

Each servo motor is connected to a saxophone key using a nylon-coated wire (CENFILL FLEX 7×7, diameter 0.6 mm). The wire is attached to the tip of the servo horn, routed through a hole in the sheet-metal body, and fixed to the corresponding key. When the servo rotates, the wire is pulled, causing the key to close. When the servo rotates in the opposite direction, the wire slackens, and the key opens due to the restoring force of the key's spring.

The opening speed of each key is determined by the spring tension rather than the actuator speed. Therefore, even if a servo motor is capable of high-speed motion, the key-opening speed has an inherent mechanical limitation. Additionally, some saxophone keys exhibit inverse behavior, where pressing the key opens the tone hole rather than closing it; these characteristics were carefully considered in the actuator layout and control design.

3.5 Total Weight

The weight of the aluminum body plate is 195 g. Including 19 servo motors and the control circuit board, the total weight of the robotic unit is approximately 700 g. The weight of the saxophone itself is approximately 1.4 kg, resulting in a combined system weight of about 2.1 kg. For comparison, a typical alto saxophone weighs approximately 2.6 kg. Therefore, the proposed system remains lighter than a standard alto saxophone and can be comfortably held and played by a human performer.

4 Robo-Sax Quartet Ensemble System

The saxophone family includes instruments with different pitch ranges, such as soprano, alto, tenor, and baritone saxophones. Ensemble performance using these four instruments in a saxophone quartet is widely practiced and represents a rich musical experience. Although these instruments are transposing instruments with different keys, their basic fingering patterns are largely consistent. In contrast, breath control differs significantly across instruments. For example, soprano saxophones require a tighter embouchure and faster airflow to stabilize high pitches, whereas baritone saxophones demand slower and larger-volume airflow to produce low-frequency tones.

In this study, we aim to realize a robo-sax quartet by simultaneously operating multiple semi-automatic robotic saxophones. By combining robot-assisted key actuation with human breath input across four instruments, the system enables ensemble performance while preserving embodied and expressive aspects of wind instrument playing.

4.1 Robotization of Large Saxophones

Compared with the soprano saxophone described in the previous section, lower-pitched instruments such as the tenor and baritone saxophones are substantially larger and heavier. As a result, their keys require greater actuation force, and performers typically hold the instruments close to their torso. These characteristics impose additional constraints on robotic augmentation.

Figure 3 (a) shows the developed robo-tenor and robo-baritone saxophones. In the robotization of soprano and alto saxophones, the robotic body was mounted on the rear side of the instrument using the threaded hole of the thumb hook. However, applying the same rear-mounted configuration to tenor and baritone saxophones caused the robotic structure to interfere with the performer's body, particularly around the chest and abdomen. Moreover, due to increased key resistance and longer mechanical transmission paths, wire-driven pulling mechanisms from the rear side proved insufficient. [1]

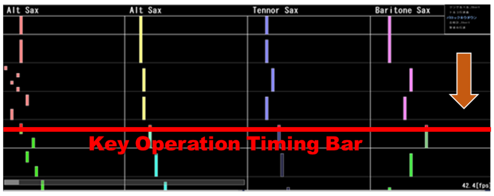
To overcome these limitations, we developed an alternative actuation mechanism in which a sheet-metal structure is installed between the front-side keys and the bell of the instrument, as illustrated in Figure 3 (b). In this configuration, servo motors press keys by pushing a rigid rod forward rather than pulling a wire. A rack-and-pinion mechanism attached to each servo converts rotational motion into linear displacement, allowing sufficient force to be applied to reliably actuate large keys. To avoid damaging the instrument, the tip of each rod is covered with a soft material fabricated using TPU filament.

Certain keys, such as octave keys and high-register auxiliary keys, require opening rather than pressing motions and do not demand large actuation forces. These keys are therefore actuated using wire-driven mechanisms similar to those used in the soprano saxophone system. By combining rack-and-pinion push mechanisms for high-force keys with wire-based pull mechanisms for low-force keys, the system adopts a hybrid actuation strategy tailored to the mechanical characteristics of each key.

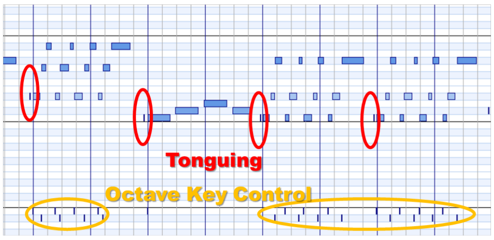
Figure 3 (c) shows a performer playing the robo-tenor saxophone. The robotic mechanism is positioned so as not to interfere with the performer's body, allowing the instrument to be held and played in a conventional posture. This demonstrates that the proposed design enables practical performance of large saxophones while maintaining comfort and playability for human performers.

4.2 Control Software

The robo-sax system is recognized by a PC as an external MIDI device, enabling straightforward control using standard musical software environments. To support ensemble performance, we developed a dedicated game-style graphical user interface (GUI) that allows simultaneous control of four robo-sax instruments, as shown in Figure 4(a).



(a) Gamification GUI System



(b) Additional MIDI note to control Robo-Sax

Figure 4: (a) Game-style graphical user interface (GUI) that enables simultaneous control of four robo-sax instruments. Musical notes scroll toward a timing bar, triggering key actuation, while performers provide breath input in synchronization. (b) Example of MIDI data conversion for the robo-sax system, where additional control notes outside the normal saxophone pitch range are embedded to specify tonguing, alternative fingerings, and octave key operations.

In the GUI, musical notes scroll vertically from the top of the screen toward a horizontal timing bar. When a note reaches the red timing bar, the corresponding key actuation is triggered on the robo-sax. By synchronizing breath input with this visual cue, performers can play the instrument in real time. This design allows performers to focus on breath control and ensemble timing, while the robotic system handles complex fingering operations.

The software was developed for the Windows platform using C++ and the game development library DXLIB. Because precise timing is critical for musical performance, native code was selected instead of higher-level languages such as Python. The system operates at 60 frames per second, synchronized with the display's vertical refresh rate to ensure stable visual timing.

Performance data can be directly derived from standard MIDI files. In our experiments, commercial sheet music for saxophone quartet was purchased and converted into MIDI data using Score Maker Zero (Kawai Co.). Although raw MIDI data can be used to drive the robo-sax system, additional preprocessing is required to achieve refined and idiomatic saxophone performance.

To this end, we embed robo-sax-specific control codes into the MIDI data as additional note events. For example, passages requiring tonguing are augmented by inserting repeated notes of the same pitch at slightly earlier timings than the main note events. Alternative fingerings, which are commonly used by human saxophonists to minimize finger motion between successive notes, are also explicitly encoded. Furthermore, in pitch leaps across octaves, appropriate advance or delayed actuation of the octave key is essential for stable sound production. While human performers perform these actions implicitly, the robo-sax system requires explicit specification.

---

[1] Early prototypes employed rear-mounted wire-driven mechanisms similar to those used for soprano saxophones. These designs suffered from insufficient actuation force and, in some cases, electrical short circuits caused by contact between wiring and the performer's body, leading to circuit failure.

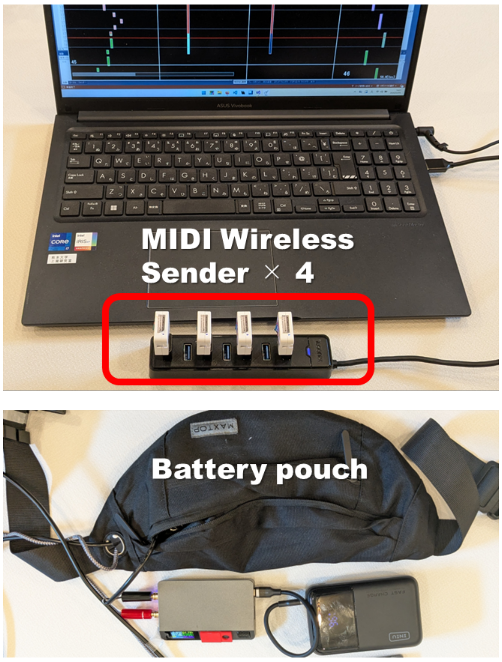


Figure 5: Wireless communication and portable power system for the robo-sax quartet. Top: Four wireless transmitters connected to a PC via a USB hub, each corresponding to one robo-saxophone (soprano, alto, tenor, and baritone). Bottom: Wearable battery pouch integrating a mobile power bank and DC regulator, enabling stable power supply without restricting performer movement.

These control notes for alternative fingerings and octave key operations are defined as MIDI note numbers outside the normal saxophone pitch range, specifically between note numbers 10 and 20. An example of this encoding scheme is shown in Figure 4(b). Based on this specification, we developed a conversion tool that preprocesses standard MIDI files into robo-sax-optimized MIDI data, enabling expressive and reliable ensemble performance.

4.3 Wireless System

The robo-sax system requires a personal computer (PC) to control four instruments simultaneously for saxophone quartet performances, while also providing a stable power supply to each unit. Although wired USB communication is technically feasible, it introduces several practical issues, including the risk of performers tripping over cables and physical interference between cables and the performers' bodies during performance. To address these issues, we implemented a fully wireless communication and portable power system.

4.3.1 Wireless Communication. As shown in the upper part of Fig. 5, a USB hub is connected to the control PC, to which four wireless transmitters (M5Atom Lite S3) are attached. Each transmitter corresponds to one robo-sax instrument: soprano, alto, tenor, and baritone. These devices are based on ESP32-series microcontrollers, which support ESP-NOW, a lightweight and high-speed wireless communication protocol operating in the 2.4 GHz band.

ESP-NOW enables direct peer-to-peer communication between microcontrollers with low latency and minimal overhead. In our system, the measured communication latency was approximately 5 ms, which is sufficiently low for real-time musical performance. Compared to BLE MIDI solutions, this approach is both more cost-effective and offers lower latency; therefore, ESP-NOW was selected for the wireless communication system.

4.3.2 Portable Power Supply. Each robo-sax instrument requires a power supply of approximately 7.4 V and up to 5 A to drive the servo motors. While a stationary stabilized power supply can provide sufficient power, it is heavy and necessitates wired connections, which reduces mobility and usability during performance. To overcome these limitations, we developed a portable power supply system, shown in the lower part of Fig. 5. A compact mobile battery capable of delivering up to 100 W is placed inside a waist pouch. Power is supplied via USB Type-C Power Delivery (PD) and converted to 7.2 V using a compact DC stabilized power module (DP100). The pouch is worn around the performer's waist and connected directly to the robo-sax instrument.

This configuration eliminates problematic power cables around the performer and enables greater freedom of movement. As a result, the system supports not only stage performances but also more dynamic performance styles such as marching and staged ensemble performances.

5 Evaluation and Demonstration

We conducted a series of evaluation experiments and public performance demonstrations to assess the mechanical performance, acoustic characteristics, and practical usability of the proposed robo-sax quartet system.

5.1 Key Actuation Latency

We first measured the opening and closing speeds of the keys on the Robo-Soprano-Sax. A high-speed camera (CASIO EX-ZR100) operating at 240 frames per second was used to record key motions. From the recorded video, the frames corresponding to the start and end of each key movement were manually identified, and the number of frames was converted into elapsed time in milliseconds.

Both opening and closing times were measured for each key. The results are summarized in Table 1. Across all measured keys, the actuation time ranged approximately from 12 ms to 40 ms, with an average opening/closing time of about 25 ms. For comparison, the typical manual key actuation speed of an experienced saxophonist without robotic assistance is reported to be around 50 ms.

Keys requiring a larger stroke length to close, such as the D, E, and F keys, exhibited slower actuation. In contrast, octave keys, palm keys, and table keys require only small displacements and therefore showed faster response times. In the closing phase, motion is primarily driven by the servo motors, whereas opening relies on the instrument's key springs. As a result, the opening speed is fundamentally limited by the spring tension rather than actuator performance.

Overall, the results indicate that the robotic system achieves faster key actuation than human manual operation. It should be noted, however, that human performers implicitly anticipate upcoming fingering changes, a factor not directly captured by this mechanical latency measurement.

5.2 Acoustic Analysis and Listening Test

We compared the acoustic characteristics of robotic and human performances. A short musical excerpt, illustrated in Fig. 6, was

Table 1: Measured delay times for closed and open key

Note	Close [ms]	Open [ms]
B	33.3	25.0
A#	33.3	25.0
A	33.3	29.2
G	37.5	33.3
G#	33.3	33.3
F	41.7	33.3
E	37.5	33.3
D	37.5	33.3
D#	20.8	20.8
C#	16.7	25.0
C	25.0	20.8
Low B	16.7	20.8
Low Bb	12.5	16.7
Octave	16.7	29.2
2nd Oct.	20.8	16.7
C1	16.7	12.5
C2	20.8	16.7
C3	20.8	20.8
C4	20.8	12.5
C5	16.7	16.7
Average	25.6	23.8

Table 2: Results of instrumental listening test

Question	Answer as		Unable to judge
	Human	Robot	
For Human musical performance	4	3	3
Which plays better?	2	7	1

performed both by a human saxophonist and using the Robo-Soprano-Sax. Both performances were recorded while listening to a metronome click to ensure consistent timing.

Figure 6 also shows the recorded waveforms and Constant-Q Transform (CQT) representations. The resulting waveforms and estimated pitch contours were largely consistent between the two performances. Although the robotic performance exhibited slightly stronger overtones, this difference is likely attributable to acoustic changes introduced by the sheet-metal structure mounted on the instrument.

A listening test was conducted in which participants were asked to identify whether a given recording was performed by a human or by the robot. Because the robotic performance contained mechanical noise from the servo motors, similar mechanical noise was artificially added to the human performance to avoid bias. The results are shown in Table 2.

Among the ten participants, four correctly identified the human performance, three incorrectly identified it as robotic, and three reported that they could not distinguish between the two. These results suggest that the difference between robotic and human performances was perceptually subtle. When asked which performance they preferred, seven participants reported a preference for the robotic performance. This preference may be attributed to the more stable sound quality achieved by the robotic system, which allows performers to focus exclusively on breath control and embouchure.

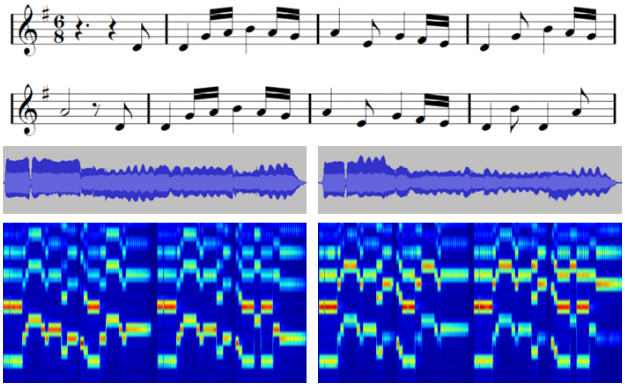

(a) Human (b) Robo

Figure 6: Comparison of Robot and Human waveforms

5.3 Concert Performances

We conducted multiple public concert performances and demonstrations using the proposed robo-sax quartet system. The ensemble consisted of a Robo-Soprano-Sax (YAMAHA YSS-61), Robo-Alto-Sax (YAMAHA YAS-62), Robo-Tenor-Sax (YAMAHA YTS-62), and Robo-Baritone-Sax (YAMAHA YBS-62). The performers were members of a university wind ensemble saxophone section; when necessary, the authors also participated as performers. The performed repertoire included saxophone quartet arrangements such as *Sir Duke*, *Supercalifragilisticexpialidocious*, and *Spain*. More than two months of rehearsal using the robotic instruments were conducted prior to the concerts, as the playing experience differs from that of conventional instruments and requires dedicated practice.

Figure 7(a) shows a performance in a small concert hall. The program consisted of four pieces with a total duration of approximately 15 minutes, performed for an audience of around 100 people. The main PC display was distributed via an HDMI splitter to two monitors placed on music stands for the performers and to a rear projection screen for the audience. Performers used both the proposed GUI and conventional sheet music. Timing synchronization among performers was achieved using the visual GUI cues and a wireless single-ear earphone providing a click track. A single-ear configuration was adopted to allow performers to hear the ensemble acoustically. After extensive rehearsals, the concert proceeded without major issues, and the audience reportedly did not perceive the use of robotic assistance.

Figure 7(b) shows an ensemble performance with a professional theatrical vocalist. Despite being a sight-reading performance, a high level of musical coordination was achieved. The vocalist reported that performing with the robo-sax quartet was more engaging than singing with pre-recorded accompaniment, as the robotic instruments produce real acoustic sound. Similarly, robo-sax performers reported increased engagement when performing with a live vocalist. At present, however, the system performs at a fixed tempo defined by the sequence data, requiring the vocalist to adapt to the robot-controlled timing.

Figure 7(c) shows a performance at a conference banquet held in a large hall with an audience of approximately 400 people. In this performance, drums and vocals joined the robo-sax quartet with minimal prior rehearsal. The drummer synchronized performance by following the GUI display on the PC. Although

(a) (b) (c)

Figure 7: Concert and demonstration performances using the robotic saxophone quartet: (a) a public concert in a small concert hall, where four robotic saxophones were performed by human players using the proposed GUI and click synchronization; (b) an ensemble performance with a professional theatrical vocalist, demonstrating real-time augmentation with a human singer; (c) a banquet performance at an academic conference in a large venue, involving additional musicians. These performances illustrate the practical stability and musical applicability of the proposed system in diverse performance contexts.

Table 3: Participant profiles for Robo-Sax interviews

ID	Gender	Age	Experience	Instrument
P1	F	23	8+ yrs (wind band)	Alto, Soprano
P2	M	21	8 yrs (wind band)	Tenor
P3	M	21	8 yrs (wind band)	Baritone
P4	F	18	6 yrs (wind band)	Alto
P5	F	19	6 yrs (wind band)	Alto
P6	F	19	6 yrs (wind band)	Alto, Soprano
P7	M	19	6 yrs (wind band)	Baritone
P8	M	19	6 yrs (wind band)	Tenor
P9	F	18	6 yrs (wind band)	Tenor
P10	M	40	Professional	Tenor, Baritone

slight timing deviations occurred due to the size of the venue, no critical issues were observed.

These demonstrations collectively indicate that the proposed robo-sax quartet system achieves stable operation and practical ensemble performance at a level suitable for real-world musical contexts.

5.4 Interviews with Robo-Sax Performers

To investigate the usability and limitations of the proposed system, ten saxophonists were invited to perform with Robo-Sax, followed by semi-structured interviews. Participant details are summarized in Table 3. The group included both student musicians with 6–8 years of wind band experience and one professional saxophonist, covering a range of saxophone types including soprano, alto, tenor, and baritone. The main feedback is summarized below.

Positive Feedback.

- **Reduced burden of score reading**
  Several performers noted that the system allows them to grasp the structure and musical flow of a piece more quickly. Traditionally, they relied on reading sheet music while listening to reference recordings or watching performance videos. With automated fingering, they could focus on the overall musical image without intensive score reading.
- **Ability to perform fast and accurate fingerings**
  Performers appreciated that Robo-Sax enabled fingering passages beyond their technical ability. Rapid phrases could be executed without fingering errors.
- **Increased focus on expressive elements**
  Since fingering is handled by the robotic mechanism, performers could concentrate on breathing, embouchure, vibrato, and pitch control. Some participants suggested that the system could also serve as a training tool for breath and embouchure control.
- **Ease of switching between instruments**
  Participants indicated that transitioning between different types of saxophones (e.g., alto to tenor) became easier, potentially lowering the barrier to instrument switching.

Negative Feedback.

- **Difficulty understanding the GUI**
  Some performers found the game-style graphical interface unintuitive. As many saxophonists are accustomed to traditional staff notation, they expressed a preference for incorporating standard musical score display.
- **Difficulty with tonguing timing**
  Because fingering and sound articulation are separated, synchronizing tonguing with the robotic fingering was reported to be challenging.
- **Challenges in ensemble synchronization**
  In ensemble contexts, performers experienced difficulty handling tempo fluctuations such as fermatas and collective timing adjustments. Familiarization and dedicated practice with the robotic system may be necessary.
- **Reduced sense of performance agency**
  Some participants reported feeling as though they were being "played by the robot," indicating a perceived shift of control from the performer to the system. This suggests the need to further consider performer agency in the design.

These findings indicate that Robo-Sax has the potential to support performance and expand expressive possibilities, while also revealing challenges in interface design, synchronization, and performer agency.

5.4.1 Novice Performer Experience.

Two students with no prior saxophone experience performed on the Robo-Alto-Sax. After a 30-minute introduction to reed sound production, both were able to play a melody without prior fingering knowledge or music-reading skills. Post-performance interviews indicated that participants enjoyed the experience and remained motivated to continue practicing breath control independently.

6 Discussion and Limitations

6.1 Discussion

The proposed robotic saxophone system demonstrates a form of human–robot augmentation in instrumental performance, in which musical control is selectively shared between the performer and the robotic mechanism. In this system, pitch-related key operations are delegated to the robot, while essential expressive elements such as embouchure, breath pressure, articulation, and musical phrasing remain under the control of the human performer. This division of roles allows performers to focus on sound production and musical expression rather than on complex fingerings.

The experimental results show that the robotic key actuation operates faster than typical human finger movements, enabling reliable execution of rapid pitch transitions. At the same time, acoustic analysis and listening tests indicate that the resulting performances are perceptually comparable to human-only performances, despite the presence of mechanical actuation. These findings suggest that the system functions not as a replacement for human performers, but as an instrumental interface that augments human performance capabilities.

Unlike fully autonomous musical robots, the proposed system preserves the embodied nature of wind instrument performance. Because sound generation is entirely acoustic and directly controlled by the performer, the system maintains a strong connection between physical gesture and musical outcome. In ensemble settings such as the saxophone quartet demonstrations, this design supports stable coordination while retaining the presence of live performance and human agency.

6.2 Limitations

The current system focuses primarily on key-based pitch control and does not address other important dimensions of saxophone performance. Expressive techniques that rely on subtle embouchure adjustments, tonguing variations, or fine-grained breath control are not mediated by the robot and remain entirely dependent on the performer. Consequently, the expressive scope of the system is limited to pitch changes achieved through key actuation.

In addition, the system operates according to a predefined tempo sequence and does not adapt in real time to expressive timing variations by other performers or vocalists. This requires human performers to synchronize with the robotic system rather than enabling mutual tempo adaptation. While this approach is sufficient for rehearsed ensemble performances, it restricts flexibility in improvisational or highly interactive musical contexts.

7 Conclusion

This paper presented a robotic saxophone augmentation system in which a robot assists fingering operations while a human performer remains responsible for breath control and musical expression. Robotic mechanisms were developed for soprano, alto, tenor, and baritone saxophones, employing a hybrid actuation strategy that combines rack-and-pinion and wire-driven mechanisms to accommodate differences in key characteristics and instrument size.

An integrated ensemble system was realized by combining a game-style GUI for four instruments, a MIDI conversion scheme tailored to robotic control, low-latency wireless communication, and a portable power supply. Measurements and acoustic analyses showed that robotic key actuation achieved speeds comparable to or faster than those of experienced human players, with acoustically similar results.

Public concerts and demonstrations confirmed that the system operates reliably in real performance settings and supports musically coherent saxophone quartet ensembles. By offloading complex fingering tasks to the robot, performers were able to focus on tone production and expressive control, highlighting the system's role as an instrument interface that extends, rather than replaces, human musical performance.

Future work will investigate adaptive tempo control, expanded sensing of performer intent, and tighter integration between human expressive input and robotic actuation, aiming to enhance collaborative performance between humans and robotic musical systems.

Ethical Statement

This study involved human participants in acoustic analysis listening tests, semi-structured interviews with saxophonists, and public concert performances. All participants were informed of the purpose of the study and provided verbal consent prior to participation. Performers in public concerts and conference events consented to the use of their names in official programs and publications. No personally identifiable information beyond what participants agreed to disclose was collected. The study did not involve vulnerable populations, deception, or any procedures that could cause physical or psychological harm. This research was conducted in accordance with the ethical guidelines of Kumamoto University.

References

[1] Kapur A. 2020. A history of robotic musical instruments. In *ICMC*.

[2] Q. David Bowers. 2012. *The Violin-Playing Machines: Hupfeld Phonoliszt-Violina, Mills Violano-Virtuoso: A Study and Appreciation.* Automatic Musical Instrument Collectors' Association.

[3] Gerelmaa Byambatsogt, Lodoiravsal Choimaa, and Gou Koutaki. 2021. Data Generation from Robotic Performer for Chord Recognition. *IEEJ Trans. Electronics, Information and Systems* 141, 2 (2021), 205–213.

[4] Neville H. H. Fletcher and Thomas D. Rossing. 2010. *The Physics of Musical Instruments.* Springer Science Business Media.

[5] Florian Heller, Irene Meying Cheung Ruiz, and Jan Borchers. 2017. An Augmented Flute for Beginners. *NIME'17* (2017), 15–19.

[6] Makoto Kajitani. 1989. Development of Musician Robots. *Journal of Robotics and Mechatronics* 1, 3 (1989), 254–255. https://doi.org/10.20965/jrm.1989.p0254

[7] Ajay Kapur and Eric Singer. 2006. A Retrieval Approach for Human/Robotic Musical Performance. In *ISMIR*. 363–364.

[8] Keita Kodama and Gou Koutaki. 2019. Development of guitar playing robot by PWM control of solenoid. In *IEEE GCCE*. 291–293.

[9] Yoshifumi Kurosawa and Kenji Suzuki. 2010. Robot-Assisted Playing With Fingering Support For A Saxophone. In *International Computer Music Conference, ICMC 2010*.

[10] Yoshihiro Kusuda. 2008. Toyota's violin-playing robot. *Industrial Robot:An International Journal* 35, 6 (2008), 504–506.

[11] Jia Yeu Lin, Kawai Mao, Nishio Yuya, Cosentino Sarah, and Takanishi Atsuo. 2019. Development of performance system with musical dynamics expression on humanoid saxophonist robot. *IEEE Robotics and Automation Letters* 4, 2 (April 2019), 1684–1690.

[12] Steven R. Ness, Shawn Trail, Peter F. Driessen, W. Andrew Schloss, and George Tzanetakis. 2011. Music Information Robotics: Coping Strategies for Musically Challenged Robots. In *ISMIR*. 567–572.

[13] J. Solis, M. Bergamasco, K. Chida, S. Isoda, and A. Takanishi. 2004. The anthropomorphic flutist robot WF-4 teaching flute playing to beginner students. In *ICRA*. 146–151 Vol.1.

[14] Jorge Solis, Yoshihisa Sugita, Klaus Petersen, and Atsuo Takanishi. 2016. Development of an anthropomorphic musical performance robot capable of playing the flute and saxophone: Embedding pressure sensors into the artificial lips as well as the re-designing of the artificial lips and lung mechanisms. *Robotics Auton. Syst.* 86 (2016), 174–183.

[15] Jorge Solis, Koichi Taniguchi, Takeshi Ninomiya, Tetsuro Yamamoto, and Atsuo Takanishi. 2008. Development of Waseda flutist robot WF-4RIV: Implementation of auditory feedback system. In *2008 IEEE International Conference on Robotics and Automation*. 3654–3659.