Playing Together with a Semi-Automated Robotic Flute Using a Gesture Cue Detection System

Jaeran Choi
jaeran.choi@kaist.ac.kr
Graduate School of Culture Technology, KAIST
South Korea

Juhan Nam
Graduate School of Culture Technology, KAIST
South Korea

Hikari Kuriyama
Kumamoto University
Japan

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

Figure 1: Overview of the ensemble performance setup using the developed semi-automated robotic flute and the gesture-cue-based control system with MIDI accompaniment.

Abstract

This study presents a semi-automated robotic flute system that coordinates performance onset timing based on a human performer's gesture cues and examines how such control influences the performance experience in a human–robot ensemble. In the proposed system, the performer produces sound by blowing while the robot actuates the flute's keys via a servo-driven mechanism, establishing a shared performance structure. A camera-based motion tracking system detects preparatory head gestures in real time and predicts the intended onset timing using a gesture cue–onset ratio model. We compared three conditions: timer-based onset, gesture-cue-based onset with visual feedback, and gesture-cue-based onset without visual feedback. Quantitative measures assessed onset asynchrony, and qualitative measures examined perceived partnership, agency, leadership, and trust. The results indicate that gesture-cue-based control enhances the sense of partnership and performer agency, while timer-based control yields higher timing stability. These findings suggest that gesture-driven semi-automated musical robots can shift performers' perception of the robot from a playback device to an ensemble partner.

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

Keywords

flute, semi-automated musical instrument robot, gesture cue, human–robot ensemble

1 Introduction

In recent years, research and development on robots capable of playing musical instruments have advanced. Robots that perform wind instruments [13, 20, 21, 27, 28], keyboard instruments [9], and string instruments [6, 10, 18, 24] have been reported. All of these systems are capable of fully automated performance. These systems have primarily been developed to analyze the mechanisms of musical performance or for musical appreciation.

On the other hand, robots have also been developed in which part of the performance is automated by the robot, while the remaining actions are carried out by a human [16, 30, 31]. These semi-automated systems are primarily intended to assist beginner musicians and individuals facing physical challenges and are expected to enable a wider range of people to play musical instruments more easily.

However, many semi-automated systems adopt a structure in which performers follow the robot's predetermined playback, leading the robot to be perceived as a utilitarian device rather than a musical partner. Moreover, how a robot should interpret and respond to a performer's intention at the critical moment of performance initiation remains underexplored. In ensemble performance, the beginning of a piece is a particularly crucial moment in which implicit agreement and leadership are negotiated [1, 29]. Performers coordinate onset timing through nonverbal

gestures, such as head nods, body movements, and breathing, that precede acoustic sound [2, 3]. Prior research shows that the temporal relationship between such gestural cues and musical onset can be modeled using relatively stable timing ratios [5].

Building on this foundation, this study investigates how gestural cues can determine onset timing in a semi-automated robotic flute system in which the human produces sound by blowing while the robot actuates the keys. Specifically, we examine whether gesture-based onset control leads performers to perceive the robot not merely as a reactive device but as an ensemble partner actively aligned with the performer at the moment of musical initiation. In this sense, the work also connects to a lineage of robotic musical instruments and musical mechatronics that frame mechanical systems as part of performing on acoustic instruments rather than as mere playback devices [11, 26].

2 Related Work

Many musical instrument robots and performance-support systems have been developed, but their purposes and target instruments differ. In this section, we describe various musical instrument robots and performance-support systems.

2.1 Musical Instrument Robots

In recent years, various musical instrument robots have been developed. Hoffman et al. developed a robotic marimba player that listens to a human musician and continuously adapts its improvisation and choreography while performing simultaneously with the human [8]. Lau et al. explored the formation of a musical band that includes both humanoid robots and human musicians, with the goal of achieving natural synchronization and collaboration during musical performances [19]. These robots are designed for collaborative performance with humans.

Furthermore, Zhao et al. developed a novel upper-limb rehabilitation device that integrates piano playing into task-oriented occupational therapy [33]. This device is intended for rehabilitation purposes rather than for assisting with music performance itself. Solis et al. developed a robot capable of fully automated flute performance [27, 28]. The system is equipped with mechanisms that reproduce the functions necessary for flute performance and consists of a total of 41 degrees of freedom. It mechanically reproduces the lips, neck, lungs and valve mechanism, fingers, throat, tonguing, two arms, and eyes, enabling the robot to perform the flute.

Regarding the Chinese bamboo flute, which has a shape similar to the Western flute, Li et al. developed a fully automated playing robot [20]. The robot is designed in the style of traditional Chinese wooden carving and is equipped with an air-blowing nozzle resembling a human mouth and six pneumatically actuated fingers. In addition to the fingers, it has several servo-actuated joints including the head, arms, and waist, enabling it to reproduce body movements synchronized with the performance. These robots perform autonomously and are classified as fully automated playing robots.

2.2 Performance-Support Systems

In the field of performance support, various systems have been developed for beginner players or to support specific movements. Chin et al. proposed a recorder tutoring system that integrates auditory, visual, and haptic modalities to support recorder learning [4]. Kato et al. proposed a breathing assist device for saxophone players [12, 14]. Xu et al. also developed a fully actuated and lightweight hand exoskeleton robot for piano playing, designed to assist novice piano players in maintaining correct finger technique [32].

Research has also been conducted on performance support for the flute. Kuroda et al. proposed a sensorless control parameter estimation system that estimates the physical control parameters from the sound of a flute played by humans. They also demonstrated that it is possible to provide guidance on wind instrument performance parameters based on the recorded sound without using special sensors [17].

Additionally, devices that assist with correct embouchure formation in flute playing have also been developed. Heller et al. developed an augmented flute for beginners that makes the parameters of the embouchure visible [7]. Playing the flute is determined by a range of parameters, including the holding angle, the angle of airflow, the embouchure, and the width and speed of the breath. Correcting these requires an experienced teacher, making it difficult for beginners to practice individually. To address this problem, they proposed an augmented flute that measures and visualizes the parameters of the embouchure.

Regarding gesture-based performance support, prior studies have examined how performers achieve synchronization through gestures, particularly after long pauses and at the beginning of a piece. These works show that kinematic features such as acceleration, periodicity, duration, and peak velocity convey beat position and tempo information [1–3, 15]. In flute performance, gesture velocity features have been shown to maintain a stable ratio relationship with actual onset timing, supporting the feasibility of predicting intended onsets from gesture cues [5]. In addition to analytical studies, gesture cues have also been incorporated into responsive systems for synchronization. For example, gesture-driven systems have explored adjusting the tempo of automatic accompaniment in real time based on performers' gesture cues [22].

However, these studies have primarily focused on analyzing synchronization mechanisms between human performers or improving prediction accuracy. The impact of gesture cues on performers' perceived agency and sense of collaboration in a human–robot ensemble context remains relatively underexplored.

3 Proposed System

Our system consists of two main parts: hardware and software. The hardware implements the flute's semi-automated mechanism, while the software senses human movement to control the robo-flute. Each is described below.

3.1 Hardware: Robo-Flute

3.1.1 Overview of the Flute. As shown in Figure 2, the flute consists of a head joint, a middle joint, and a foot joint. The player places their lower lip on the lip plate and blows air near the edge of the embouchure hole to produce sound. Flute playing actions can be broadly divided into two categories: fingering and sound production. Fingering requires combining more than 25 different fingerings using nearly all of the fingers, and mastering it demands extensive practice.

3.1.2 Automatic key fingering by servo motor. We developed a handheld robotic flute with automatic fingering, as shown in Figure 3. The automatic fingering mechanism is equipped with a microcontroller and a motor driver, and each mechanism is fitted with its own servo motor. Musical Instrument Digital Interface (MIDI) messages [23] are used as musical control input. The servo

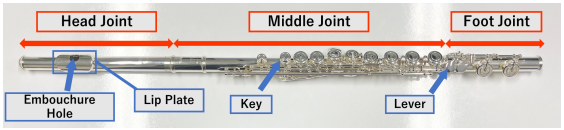


Figure 2: The flute's sections: the head joint, middle joint, and foot joint.

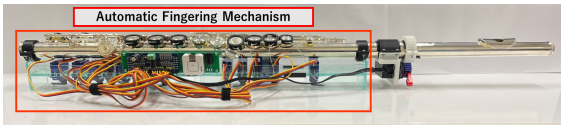


Figure 3: Developed robo-flute.

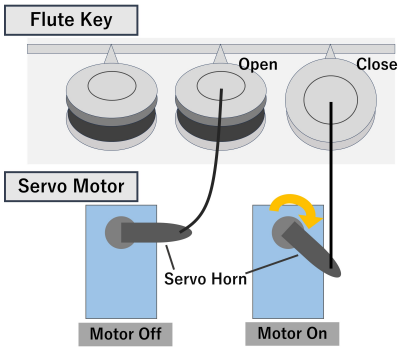


Figure 4: Mechanism for pressing keys using wires. A wire is attached to the key, and the other end is connected to the servo horn. When the servo motor rotates, the wire is pulled, causing the key to be pressed down.

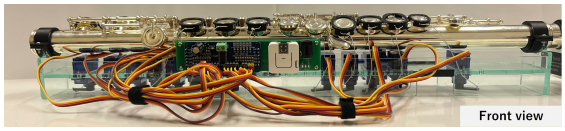
motors are driven according to each note, thereby automating key opening and closing and allowing the performer to play by simply blowing into the flute. In the following, we describe the design principles and provide an overview of these mechanisms.

If the keys were pressed directly from above, as in human playing, the mechanism would likely become bulky. Therefore, this study adopts a wire-based key-pressing method, as shown in Figure 4. In this method, the wire attached to each key is pulled by a motor rotation to press the key.

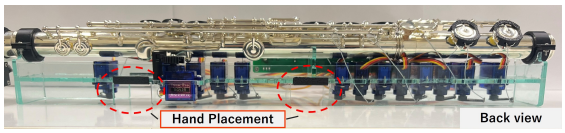
To keep the mechanism non-destructive and easily detachable, the wire is not fixed directly to the key but is attached in a removable structure. To ensure proper closure of the tone hole, a detachable key cover is designed to apply force stably to the key surface.

3.1.3 Mechanism Overview. Figure 5(a) shows a flute equipped with the automatic fingering mechanism. The flute is mounted on an acrylic body on which fourteen SG92R servo motors (Tower Pro) are arranged, each driving one key.

The motor control board is installed on the side of the body. A dedicated 3D-printed fixture is used to secure the flute, allowing it to be attached and removed with bolts and nuts. A silicone sheet is placed between the fixture and the flute to prevent slipping. A dedicated hand rest, positioned near the usual flute-supporting point, allows the player to hold the mechanism stably while maintaining a natural playing posture (Figure 5(b)). The mechanism uses two key/lever pressing methods: *a wire-based method* and *a rack-and-pinion method.*



(a)



(b)

Figure 5: Overview of the automatic fingering mechanism. (a) Front view. A circuit board for motor control is mounted on the side of the body. (b) Rear view. The body shape is designed to be easily supported by hand, enabling the reproduction of a normal playing posture.

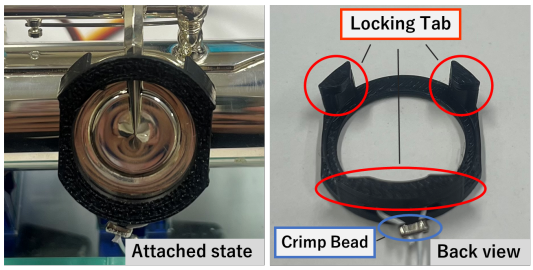


Figure 6: Removable key cover

(i) Wire-Based Method. In this method, each key is fitted with a removable key cover (Figure 6(a)). The wire is attached to the lower part of the key cover, with the other end attached to the servo horn. Motor rotation pulls the wire and presses the key to close the tone hole (for levers, pressing opens the corresponding key). A crimp bead is used to fix the wire at the desired position (Figure 6(b)). The underside of the key cover is shaped to hook securely onto the key and prevent detachment during performance (Figure 6(b)).

On the other hand, due to their structure, levers make it difficult to use key covers. Since levers do not serve to close tone holes, wires are attached directly, and the levers are pressed down by the rotation of the servo motors. Additionally, the wire is given adequate slack to allow for removal.

(ii) Rack-and-Pinion Method. For keys mounted on the side, attaching a key cover is difficult. Since these keys need to reliably close the tone holes, a wire-based method, as used for the levers, is unlikely to provide stable operation. Therefore, a rack-and-pinion method, as shown in Figure 7(a), is adopted to convert the rotational motion of the motor into linear motion. By rotating the pinion, the rack is driven back and forth, and the key is pressed directly by the end of the rack, thereby enabling the actuation of the side key (Figure 7(b)).

3.2 Servo Control System

Figure 8 shows the configuration of the circuit board and its surrounding components. The board is equipped with a power connector, a microcontroller, and a motor driver. In this study, the microcontroller is the AtomS3-Lite from M5Stack, and the motor driver is the PCA9685 from NXP. Servo motors are driven

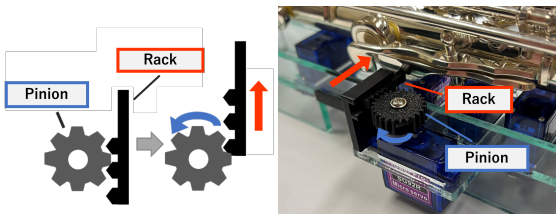


Figure 7: Rack & pinion system for Briccialdi key

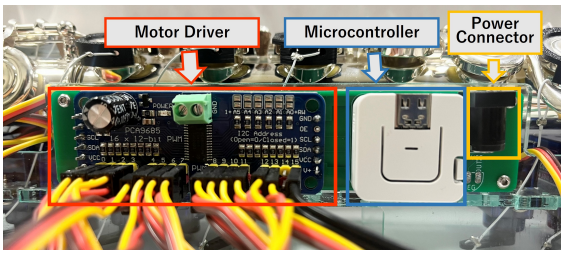


Figure 8: Configuration of the circuit board and its surrounding components. From right to left, the power connector, microcontroller, and motor driver are mounted. The servo motors are controlled by these components.

using Pulse Width Modulation (PWM) control. PWM control varies the effective output voltage by changing the duration of the high level in a pulse train with a fixed period.

3.3 Software: Gesture Cue Detection System

Gesture cue detection aims to interpret a performer's preparatory movements as expressions of musical intention. A head nod performed immediately before the start of a piece is regarded as a cue that anticipates the intended onset timing. To detect such gesture cues in real time, we implemented a camera-based motion tracking system in Python.

The system uses a MediaPipe Face Landmarker [25] to track the performer's facial region from video captured at 30 fps. A velocity curve is extracted from the vertical motion of the facial landmarks. From this curve, the maximum peak (maximum upward velocity) and the subsequent minimum peak (minimum downward velocity) are detected sequentially. Following Choi et al. [5], the temporal interval between these two peaks is defined as the duration of the gesture cue. Based on this duration, the intended onset timing is predicted using a flutist-specific gesture cue–onset ratio model.

The predicted onset time is computed as follows:

$$t_{\text{onset}} = t_{\max} + d \times R, \tag{1}$$

where $t_{\max}$ denotes the time of the maximum velocity peak, d is the gesture cue duration, and R is the cue–onset ratio. In this system, we adopt the ratio $R = 2.28$ proposed by Choi et al. [5].

At the predicted onset time, the Python system sends a MIDI signal to the robotic flute, triggering key actuation while the human performer produces sound at the same moment. To account for the robot's mechanical latency of approximately 50 ms, the trigger timing is compensated such that the actual key actuation aligns naturally with the human performer's breath onset (Figure 9).

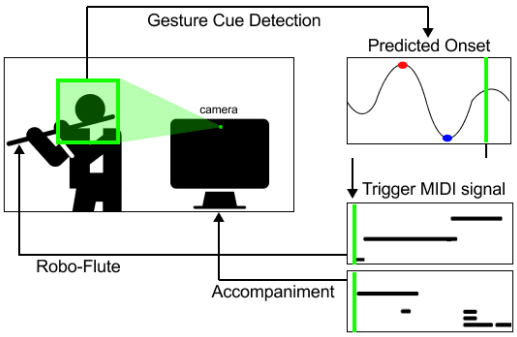


Figure 9: Gesture cue detection and triggering system.

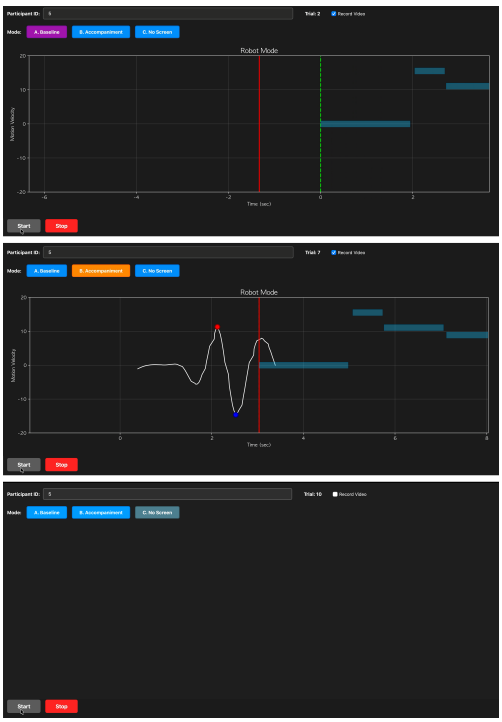


Figure 10: Overview of the three interface conditions. A: Timer-based onset with piano roll. B: Gesture cue–based onset with visual feedback. C: Gesture cue–based onset without visual feedback.

3.4 Interface Design for Human–Robot Ensemble

To investigate the effects of gesture cue–based control on the performer's ensemble experience, we designed three interface conditions that vary along two factors: the method of initiating performance (timer-based vs. gesture-based) and the presence or absence of visual feedback (Figure 10).

3.4.1 Condition A: Timer-based Onset. Gesture cue detection was not used. The system provided a 3-second countdown and then automatically started MIDI playback. The performer produced sound in synchrony with the countdown, while a piano roll and playhead provided visual feedback on playback progress. This condition represents a typical semi-automated setup in which the robot determines the onset timing and the performer adapts accordingly.

3.4.2 Condition B: Gesture Cue with Visual Feedback. The gesture cue detection system (Section 3.3) is activated. A head nod cue is analyzed in real time to predict onset timing and trigger MIDI playback. The interface displays the motion velocity curve, detected peaks, the predicted onset marker, and the piano roll playhead. In this condition, onset control is led by the performer, with the robot responding to the interpreted gesture.

3.4.3 Condition C: Gesture Cue without Visual Feedback. This condition used the same gesture cue detection and onset prediction as Condition B, but without visual feedback. Synchronization is perceived solely through auditory cues, such as key actuation and accompaniment. This allows us to isolate the effect of the visual interface through comparison with Condition B.

Across all conditions, MIDI playback by the robotic flute is synchronized with a piano accompaniment, while the performer's motion data are captured via camera and both video and audio recordings are collected simultaneously.

4 Experimental Design

4.1 Experimental Procedure

This experiment examined whether a gesture-cue-based robotic flute system could be perceived as an ensemble partner rather than a playback device, and how it influenced perceptions of partnership, agency, and leadership at performance onset. We manipulated (1) the presence of gesture cues and (2) the presence of a gesture-linked visual interface.

We used a within-subject design with three conditions: A (timer-based onset with piano roll), B (gesture-cue-based onset with visual feedback), and C (gesture-cue-based onset without visual feedback). The piece was *Moon River* (Henry Mancini), in which the piano and flute parts begin simultaneously. Participants completed 18 trials (two 9-trial sequences, each including all conditions three times), with order counterbalanced across participants, and up to 15 practice trials.

Onset alignment was evaluated as the temporal difference between robotic key actuation and human audio onset (asynchrony = $onset_{\text{robot}} - onset_{\text{human}}$). Seven trials were excluded due to operational issues. After each trial, participants rated their experience on a 1–7 Likert scale; after all trials, they completed a post-experiment questionnaire and a semi-structured interview. Quantitative measures included onset asynchrony, and qualitative measures assessed partnership, agency, leadership, and trust.

4.2 Participants

Eight flutists participated in the user study. Four were professional-level performers, and the remaining four were advanced amateurs. All participants had at least five years of ensemble experience, and none had previously performed with the robotic flute.

4.3 Questionnaires

4.3.1 Trial-level Questionnaire. Immediately after each trial, participants rated a single representative item for each experiential dimension (1 = strongly disagree, 7 = strongly agree), as summarized in Table 1.

4.3.2 Post-Experiment Questionnaire. The post-experiment questionnaire captured participants' overall judgments across conditions, as summarized in Table 2.

Item	Construct	Statement
Q1	Partnership	The robot felt like a partner playing with me.
Q2	Agency	The robot understood my musical intention and responded accordingly.
Q3	Leadership / Control	I felt in control at the performance onset.
Q4	Trust	I could trust the robot's onset timing.
Q5	Timing experience	The robot's onset timing felt natural.
Q6-1	Gesture responsiveness	The robot responded at an appropriate moment to my gesture.
Q6-2	Visual interface	The visual interface supported timing judgment.
Q7	Usability	The system was easy to use.

Table 1: Trial-level questionnaire items.

Item	Description	Mean	SD
Post Q1	Gesture cue increased trust	4.75	1.83
Post Q2	Gesture cue increased togetherness	6.50	0.53
Post Q3	Interface increased predictability	4.00	1.77
Post Q4	Interface increased confidence	4.38	1.60
Post Q5	No interface increased anxiety (reverse-coded)	4.25	2.43
Post Q6	Musically meaningful	6.62	0.74
Post Q7	Value for real ensemble use	5.38	1.41

Table 2: Post-experiment questionnaire results (Q1–Q7). Mean and SD are reported ($N = 8$). Q5 was reverse-coded.

Gesture cue effect (B/C vs A) Post Q1. With gesture cues, I could trust the robot more at onset. Post Q2. With gesture cues, the robot felt more like an ensemble partner.

Visual interface effect (B vs C) Post Q3. With the interface, onset timing was easier to predict. Post Q4. With the interface, I could use gesture cues more confidently. Post Q5. Without the interface, I felt more anxious or hesitant at onset (reverse-coded).

Overall evaluation Post Q6. The performance was musically meaningful. Post Q7. The system felt valuable for real ensemble settings.

5 Results

5.1 Onset Timing Asynchrony

Figure 11 shows condition-specific histograms of asynchrony. The timer-based condition A showed the smallest standard deviation. Between the gesture-cue conditions, C exhibited a smaller standard deviation than B. In terms of mean asynchrony, A was positive, indicating that participants tended to wait for the robot before playing. In contrast, B and C included negative values, showing that in some trials the human onset occurred before the robot. There was no substantial difference between the amateur and professional groups (Professional flutists: 0.036 ± 0.163 s, amateurs: 0.031 ± 0.205 s).

5.2 Trial-Level Questionnaire

Q1–Q3 were highest in condition C, followed by B, and lowest in A (Figure 12). In contrast, Q4–Q5 were highest in A, lower

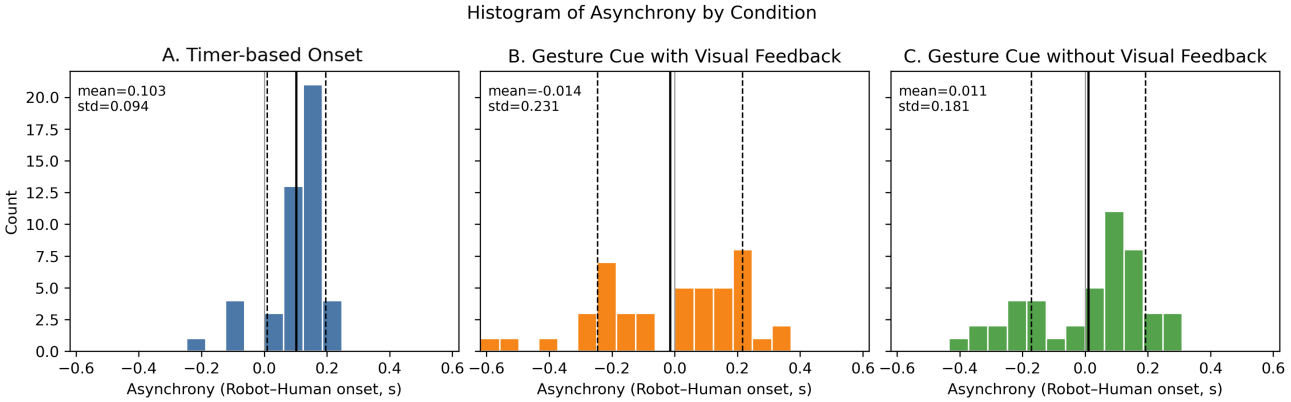


Figure 11: Condition-specific histograms of asynchrony. A) Timer-based onset, B) Gesture cue with visual feedback, C) Gesture cue without visual feedback. The solid vertical line marks the mean; dashed lines mark ±1 SD.

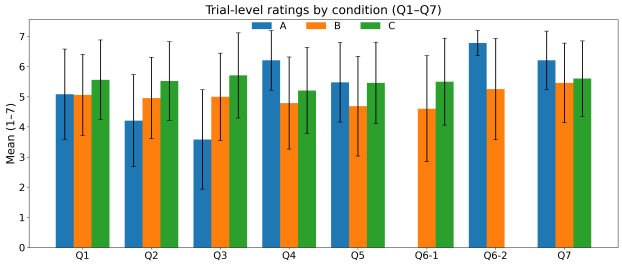


Figure 12: Trial-level mean ratings by condition (Q1–Q7). Error bars indicate ±1 SD.

in B, and intermediate in C (Table 1, Figure 12). Usability (Q7) was also highest in A. For condition-specific items, Q6-1 was higher in C than B, while Q6-2 was higher in A than B (Figure 12). Between-group comparison (Mann–Whitney U) showed a significant difference only for Q3 (leadership/control; p = .042), with professionals reporting higher ratings than amateurs (Pro: M = 5.31, SD = 0.27; Amateur: M = 4.22, SD = 1.01).

5.3 Post-Experiment Questionnaire

The post-experiment results showed high ratings for Post Q2 (gesture cues increased togetherness), Q6 (musical meaningfulness), and Q7 (value for real ensemble use). The interface-related items were near neutral: Post Q3 and Post Q4 were around the midpoint, and reverse-coded Post Q5 was also near neutral (M=4.25, SD=2.43), suggesting no consistent increase in anxiety when the interface was absent (Table 2).

5.4 User Interview Analysis

After the questionnaires, semi-structured interviews were conducted. The responses were organized into thematic categories aligned with the study questions.

Partnership and ensemble feeling. Many participants reported that conditions B/C produced a stronger sense of "playing together." In particular, condition C was often described as more partner-like, whereas condition A was often perceived as a playback device.

Agency and leadership. Condition A was described as robot-led, with little sense of personal control. Condition C was associated with stronger agency and a greater sense of initiating on one's own timing. In condition B, the visual feedback sometimes reduced participants' sense of leadership by splitting attention between bodily movement and the screen.

Trust and predictability. Early trials in conditions B and C were often described as uncertain, but trust increased over repeated trials. Several participants noted that, once they adapted, the timing felt more reliable unless an error occurred.

Gesture recognition and usability. Participants noted that small gestures were not always recognized consistently, leading them to exaggerate their movements. Several comments suggested a need for personalized calibration of nod depth and clearer criteria for gesture peaks.

Mixed effects of the visual interface (B). Some participants found the interface helpful for prediction, whereas others reported that it interfered with natural timing.

Timing strategies and adaptation. Participants described different onset strategies across conditions: Condition A relied on pre-timed breathing, while B/C relied on gesture cues followed by a short internal count. Adaptation improved performance over time, though errors could reintroduce uncertainty.

Suggested improvements. Common requests included support for tempo changes, better alignment with breath timing, and customization of gesture sensitivity.

6 Conclusion

This study presents a semi-automated robotic flute system that detects a human performer's gesture cues in real time to coordinate onset timing. Quantitative and qualitative results suggest a trade-off between temporal stability and experiential agency: the timer-based condition yielded greater timing stability, whereas gesture-cue-based conditions were associated with higher ratings of partnership, agency, and leadership. Participants also more often described the robot as an ensemble partner rather than a playback device when onset was initiated through their own gestures. Differences between the two gesture conditions were modest, and no consistent advantage of visual feedback was observed. These findings suggest that semi-automated musical robots can support more collaborative human–robot ensemble interaction. Future work should examine whether gesture-based coordination can extend beyond onset timing to tempo modulation, dynamics, and expressive interaction.

7 Ethical Standards

This study was approved by the Institutional Review Board (IRB). All participants provided informed consent for video recording and data sharing. Each participant session lasted one hour, including the interview. Participants received appropriate compensation. This research was supported by government funding. The authors declare no conflict of interest.

8 Acknowledgments

This work has been supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) under Grant RS-2023-NR077289 and Grant RS-2024-00358448.

References

[1] Laura Bishop, Carlos Cancino-Chacón, and Werner Goebl. 2019. Moving to communicate, moving to interact: Patterns of body motion in musical duo performance. *Music Perception: An Interdisciplinary Journal* 37, 1 (2019), 1–25.

[2] Laura Bishop and Werner Goebl. 2015. When they listen and when they watch: Pianists' use of nonverbal audio and visual cues during duet performance. *Musicae Scientiae* 19, 1 (2015), 84–110.

[3] Laura Bishop and Werner Goebl. 2018. Beating time: How ensemble musicians' cueing gestures communicate beat position and tempo. *Psychology of Music* 46, 1 (2018), 84–106.

[4] D. Chin and G. Xia. 2022. A Computer-Aided Multimodal Music Learning System with Curriculum: A Pilot Study. In *Proceedings of the International Conference on New Interfaces for Musical Expression*. https://doi.org/10.21428/92fbeb44.c6910363 Article 39.

[5] Jaeran Choi, Taegyun Kwon, and Juhan Nam. 2025. Predicting Flutist Onset Timing in Duet Performance: A Multimodal Analysis of Gesture and Breath Cues. In *Ismir 2025 Hybrid Conference*.

[6] T. Fei, X. Chen, and L. Zhou. 2019. Neural Network Based Online Anthropomorphic Performance Decision-Making Approach for Dual-Arm Dulcimer Playing Robot. *Journal of Advanced Computational Intelligence and Intelligent Informatics* 23 (2019), 838–846. https://doi.org/10.20965/jaciii.2019.p0838

[7] F. Heller, I. M. Cheung Ruiz, and J. Borchers. 2017. An Augmented Flute for Beginners. In *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME 2017)*. 34–37. https://doi.org/10.5281/zenodo.1176161

[8] Guy Hoffman and Gil Weinberg. 2011. Interactive Improvisation with a Robotic Marimba Player. *Autonomous Robots* 31 (2011), 133–153. https://doi.org/10.1007/s10514-011-9237-0

[9] I. Indreica, M. D. Doloiu, I.-A. Spulber, G. Măceşanu, B. Sibişan, and T.-T. Cociaş. 2025. Design and Control of a 32-DoF Robot for Music Performance Using AI and Motion Planning. *Engineering Proceedings* 113 (2025), 53. https://doi.org/10.3390/engproc2025113053

[10] W. Jo, H. Park, B. Lee, and D. Kim. 2015. A Study on Improving Sound Quality of Violin Playing Robot. In *Proceedings of the 6th International Conference on Automation, Robotics and Applications (ICARA)*. 185–191. https://doi.org/10.1109/ICARA.2015.7081145

[11] Ajay Kapur. 2005. A History of Robotic Musical Instruments. In *Proceedings of the International Computer Music Conference*. https://www.mistic.ece.uvic.ca/publications/2005_icmc_robot.pdf

[12] T. Kato, T. Ashikari, C. Matoba, A. Mawatari, and P. Thumwarin. 2021. Fabrication of a Breathing Assist Device for Saxophone Players with Breathing Problems. *Journal of Drive and Control* 18 (2021), 72–76. https://doi.org/10.7839/ksfc.2021.18.4.72

[13] T. Kato, K. Shimazaki, K. Higashijima, R. Tokunaga, P. T. Na Ayuthaya, and P. Thumwarin. 2018. Refinement of a Thai Flute-Playing Robot in Thai Style. In *Proceedings of the International Conference on Engineering, Applied Sciences, and Technology (ICEAST)*. 1–4. https://doi.org/10.1109/ICEAST.2018.8434466

[14] T. Kato, K. Shimazaki, S. Nundrakwang, P. Chuprasert, and P. Thumwarin. 2019. Proposal of the Concept of a Breathing Assist System for Saxophone Players with Breathing Problems. In *Proceedings of the 5th International Conference on Engineering, Applied Sciences and Technology (ICEAST)*. 1–4. https://doi.org/10.1109/ICEAST.2019.8802568

[15] Satoshi Kawase. 2014. Gazing behavior and coordination during piano duo performance. *Attention, Perception, & Psychophysics* 76 (2014), 527–540.

[16] Gou Koutaki and Masahiro Hamanaka. 2025. Automatic Fingering Saxophone Quartet System. In *Entertainment Computing – ICEC 2025*. 597–601. https://doi.org/10.1007/978-3-032-02555-5_52

[17] J. Kuroda and Gou Koutaki. 2022. Sensing Control Parameters of Flute from Microphone Sound Based on Machine Learning from Robotic Performer. *Sensors* 22 (2022), 2074. https://doi.org/10.3390/s22052074

[18] Y. Kusuda. 2008. Toyota's Violin-Playing Robot. *Industrial Robot: An International Journal* 35 (2008), 504–506. https://doi.org/10.1108/01439910810909493

[19] M. Lau, J. Anderson, and J. Baltes. 2025. Integrating Humanoid Robots with Human Musicians for Synchronized Musical Performances. *PeerJ Computer Science* 11 (2025), e2632. https://doi.org/10.7717/peerj-cs.2632

[20] Jiayin Li, T. Hu, S. Zhang, and H. Mi. 2019. Designing a Musical Robot for Chinese Bamboo Flute Performance. In *Proceedings of the Seventh International Symposium of Chinese CHI*. 117–120. https://doi.org/10.1145/3332169.3332264

[21] J.-Y. Lin, M. Kawai, Y. Nishio, S. Cosentino, and Atsuo Takanishi. 2019. Development of Performance System with Musical Dynamics Expression on Humanoid Saxophonist Robot. *IEEE Robotics and Automation Letters* 4 (2019), 1684–1690. https://doi.org/10.1109/LRA.2019.2897372

[22] Akira Maezawa and Kazuhiko Yamamoto. 2017. MuEns: A multimodal human-machine music ensemble for live concert performance. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. 4290–4301.

[23] MIDI Manufacturers Association. 1996. MIDI 1.0 Detailed Specification, Version 4.2. Available online. https://midi.org/midi-1-0-core-specifications Accessed: 21 November 2025.

[24] J. Murphy, J. McVay, A. Kapur, and D. Carnegie. 2013. Designing and Building Expressive Robotic Guitars. In *Proceedings of the International Conference on New Interfaces for Musical Expression*. 557–562. https://doi.org/10.5281/zenodo.1178618

[25] Google Research. 2025. MediaPipe Face Landmarker. https://ai.google.dev/edge/mediapipe/solutions/vision/face_landmarker.

[26] Eric Singer, Jeff Feddersen, Chad Redmon, and Bil Bowen. 2004. LEMUR's Musical Robots. In *Proceedings of the International Conference on New Interfaces for Musical Expression*. https://nagasm.org/NIME/NIME04/paper/NIME04_3D01.pdf

[27] Jorge Solis, Y. Sugita, K. 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 and Autonomous Systems* 86 (2016), 174–183. https://doi.org/10.1016/j.robot.2016.08.024

[28] Jorge Solis, K. Taniguchi, T. Ninomiya, T. Yamamoto, and Atsuo Takanishi. 2008. Development of Waseda Flutist Robot WF-4RIV: Implementation of Auditory Feedback System. In *Proceedings of the 2008 IEEE International Conference on Robotics and Automation*. 3654–3659. https://doi.org/10.1109/ROBOT.2008.4543771

[29] Chia-Jung Tsay. 2013. Sight over sound in the judgment of music performance. *Proceedings of the National Academy of Sciences* 110, 36 (2013), 14580–14585.

[30] K. Tsurumi, R. Marutsuka, and Gou Koutaki. 2024. Semi-Automatic Performance Support Robot That Can Attach and Detach Guitars. In *Proceedings of the IEEE 13th Global Conference on Consumer Electronics (GCCE)*. 232–235. https://doi.org/10.1109/GCCE62371.2024.10760755

[31] H. Wang, X. Zhang, and Fumiya Iida. 2024. Human–Robot Cooperative Piano Playing With Learning-Based Real-Time Music Accompaniment. *IEEE Transactions on Robotics* 40 (2024), 4650–4669. https://doi.org/10.1109/TRO.2024.3484633

[32] Q. Xu, D. Yang, M. Li, X. Ren, X. Yuan, L. Tang, X. Wang, S. Liu, M. Yang, Y. Liu, and M. Yang. 2024. Design and Verification of Piano Playing Assisted Hand Exoskeleton Robot. *Biomimetics* 9 (2024), 385. https://doi.org/10.3390/biomimetics9070385

[33] X. Zhao, Y. Zhang, Y. Zhang, P. Zhang, J. Yu, and S. Yuan. 2025. Development and Evaluation of a Novel Upper-Limb Rehabilitation Device Integrating Piano Playing for Enhanced Motor Recovery. *Biomimetics* 10 (2025), 200. https://doi.org/10.3390/biomimetics10040200