INTERACTIVE MELODY GENERATION AND ROBOTIC SAXOPHONE PERFORMANCE: THE ROBOSAX MELODY SLOT MACHINE

Masatoshi HAMANAKA[1] and **Gou KOUTAKI**[2]

[1]**RIKEN**, Kyoto, Japan
[2]**Kumamoto University**, Japan

ABSTRACT

This paper presents RoboSax Melody Slot Machine, a system that combines Melody Slot Machine, a dial-based interface for exploring melodically coherent variations, with RoboSax, a robotic saxophone that automates fingering with servo motors. Selected melodies are transmitted as MIDI to RoboSax, while the performer supplies breath and articulation. Because the melodic variations are generated by a GTTM-based morphing method, users can explore related phrases without losing structural coherence. The system therefore supports both accessible melody creation and embodied wind-instrument performance within a single interactive loop.

1. INTRODUCTION

Sound and music computing research has increasingly explored ways to support non-experts in musical creation and performance. However, many forms of composition and instrumental performance still require specialized musical knowledge, notation reading, and physical technique.

As one approach to lowering these specific barriers, we developed *RoboSax Melody Slot Machine*, which combines interactive melody generation with robotic saxophone performance (Fig. 1). Users select melodic variants on a tablet, and the system realizes them as physical performances through a hybrid human–robot instrument.

The main contributions of this work are as follows:

- **An integrated platform for melody generation and robotic wind-instrument performance.** *RoboSax Melody Slot Machine* links melody morphing, MIDI-based control, and robotic saxophone fingering within one system.
- **A semi-automatic human–robot performance model.** RoboSax automates fingering while leaving breath and articulation to the performer, reducing technical barriers while preserving embodied wind performance.
- **Real-time and ensemble deployment.** The same framework supports live reshaping of melodies and coordinated performance with multiple RoboSax instruments.

Copyright: © 2026. This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 Unported License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

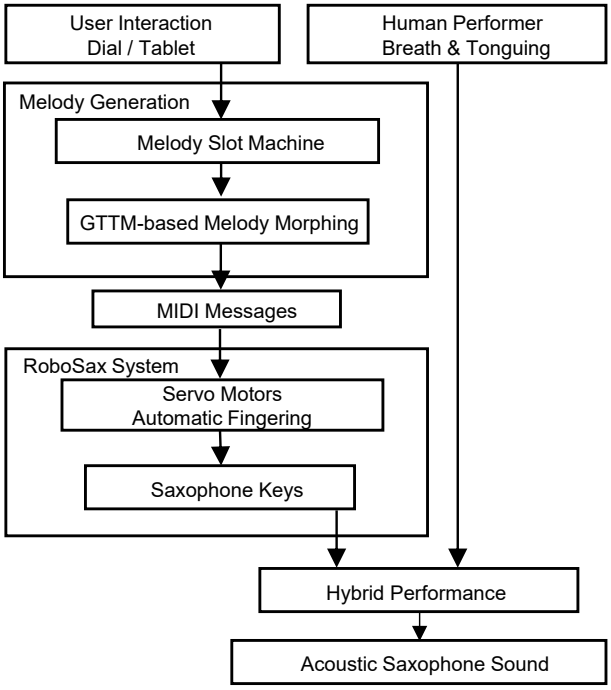


Figure 1. Overview of the RoboSax Melody Slot Machine.

2. SYSTEM OVERVIEW

The system consists of three layers:

1. Melody generation layer (Melody Slot Machine)
2. Musical communication layer (MIDI)
3. Physical performance layer (RoboSax)

Users choose a melody variant in Melody Slot Machine. The selected phrase is converted into MIDI and sent to RoboSax, which actuates the saxophone keys via servo motors. The performer supplies breath and articulation, producing a hybrid human–robot performance. The same pipeline can distribute coordinated parts to multiple RoboSax instruments in ensemble settings.

3. MELODY SLOT MACHINE

The Melody Slot Machine generates melodic variations using a GTTM-based melody morphing method (Fig. 2). Rather than switching among independent melodies, it interpolates between phrases while preserving their structural relationships.

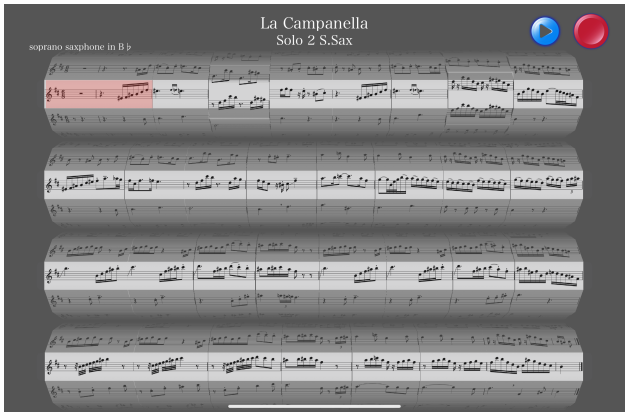


Figure 2. Melody Slot Machine interface on an iPad. Scrollable segments show related phrases, and the current segment is highlighted in red.

Rather than treating each melody as a flat sequence of notes, the method represents it with a GTTM time-span tree [1–3], which hierarchically expresses the relative structural importance of notes. This hierarchical representation allows melody morphing to operate on musical structure rather than only on surface note sequences. During the morphing process, reductions are first applied to lower- level branches of the time-span tree, so structurally important notes are preserved while less structurally important notes are gradually removed. This process enables the generation of intermediate melodies that retain the structural characteristics of the source melodies. As shown in Fig. 3, melody morphing first extracts the common structure of two melodies from their time-span trees, then partially reduces the notes in the branches that differ between them, and finally combines the reduced structures to generate intermediate melodic forms.

The melodic candidates are generated in advance using the melody morphing method. During performance, the user does not edit individual notes or generate new candidates in real time; instead, the user rotates the dial-like interface to browse the precomputed set of structurally coherent melodic variations and selects one of them for playback. This design makes the control process simple while keeping the musical output constrained to melodies that preserve the GTTM-based structural relationships.

4. ROBOSAX

RoboSax is a semi-automatic robotic saxophone that automates fingering while leaving breath and articulation to the performer.

As shown in Fig. 4, a custom metal frame is mounted on a standard saxophone body. The frame holds the actuation mechanism, control electronics, and servo motors. Because saxophone pitch is controlled by opening and closing keys, RoboSax reproduces fingering by mechanically actuating the corresponding keys through a wire-driven mechanism.

The current prototype uses 19 AGFRC A20BHS high-voltage brushless digital servomotors, which provide sufficient torque and response speed for real-time saxophone key actuation while maintaining stable continuous operation. Each servomotor is connected to a saxophone key by a thin wire; rotating the servo pulls the wire and actuates the corresponding key. This mechanism allows the robot to

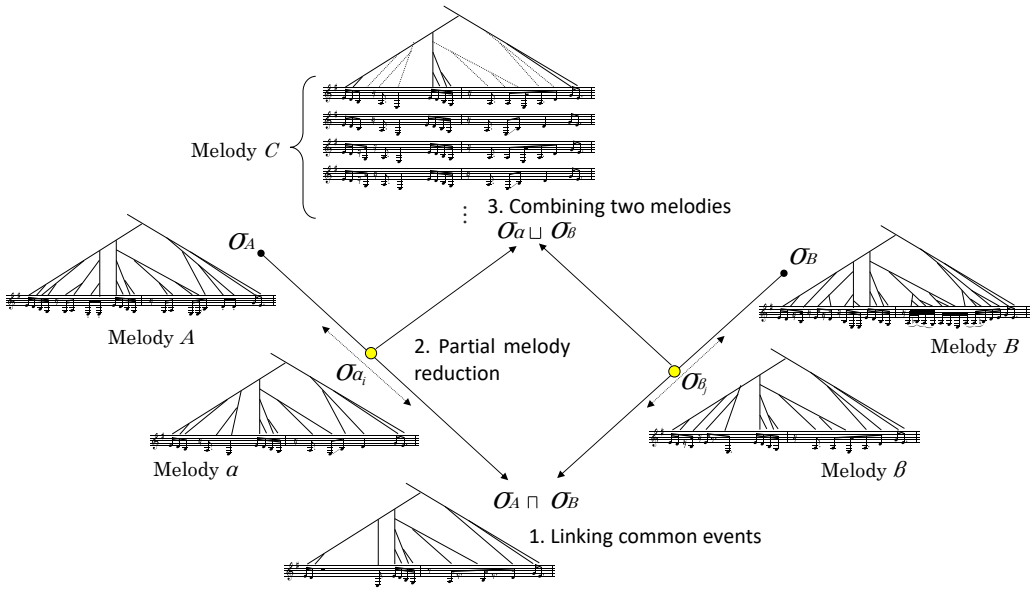


Figure 3. Melody morphing based on GTTM time-span trees: the method extracts common structure, partially reduces differing branches, and combines the resulting structures to generate intermediate phrases.

reproduce the fingering required to play approximately 32 semitones, corresponding to about two and a half octaves.

The control unit is implemented using an M5Stack microcontroller together with an M5Stack Servo2 module. The M5Stack receives MIDI messages from DAWs or custom software and communicates with the Servo2 module via I^2C. The Servo2 module generates PWM control signals for the servomotors.

The A20BHS servomotors support an operating voltage of 4.8–8.4 V. Although higher supply voltages provide faster response, they also increase heat generation. Therefore, the RoboSax system is typically operated at approximately 7.2 V to balance response speed and thermal load during continuous performance.

In the proposed system, RoboSax receives MIDI messages generated by the Melody Slot Machine and executes the corresponding fingering automatically, while the performer produces sound through breath and articulation.

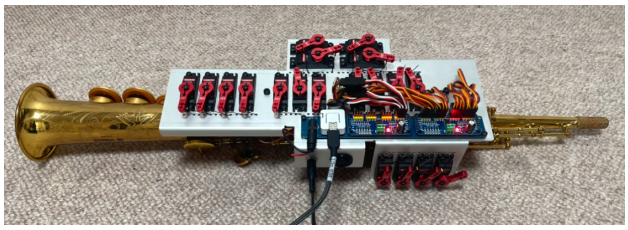


Figure 4. RoboSax hardware. Servo motors on a custom frame control the saxophone keys while the performer provides breath and articulation.

We also developed soprano, alto, tenor, and baritone RoboSax instruments [4], forming an SATB robotic ensemble (Fig. 5).

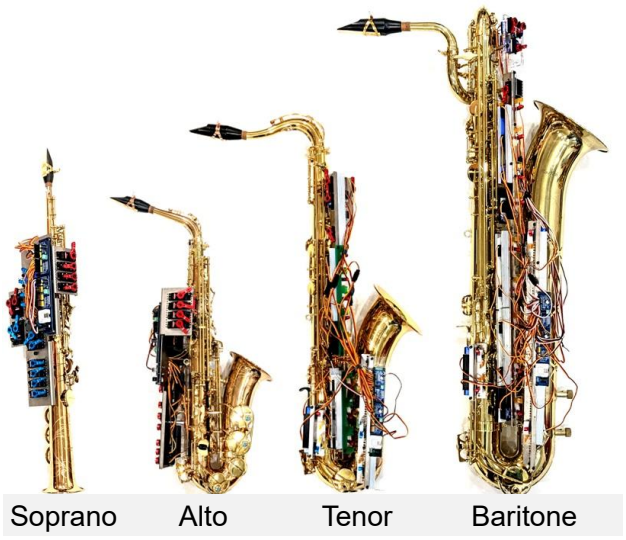


Figure 5. SATB RoboSax ensemble.

Using the same control pipeline, melodies generated by the morphing system can be distributed across the ensemble, and users can reshape the musical structure in real time from the tablet interface.

5. INTERACTION DESIGN

RoboSax Melody Slot Machine is organized as a shared interaction loop. A user selects melodic variants with the Melody Slot Machine interface, the selected phrase is translated into MIDI, RoboSax executes the fingering, and the performer supplies breath and articulation. This division lowers the technical burden of fingering while preserving the embodied aspects of wind-instrument performance.

Because melody selection continues during playback, the system supports real-time co-composition rather than fixed-score reproduction. In ensemble settings, one participant can shape melodic structure while one or more performers realize it physically.

6. RELATED WORK

Research related to the proposed system can be discussed from four perspectives: algorithmic melody generation, robotic musical instruments, new musical interfaces for non-experts, and embodied human–machine musical interaction.

6.1 Algorithmic Melody Generation and Interactive Composition

Algorithmic composition has long explored methods for generating musical structures using rule-based, probabilistic, and corpus-based approaches [5–9]. Early examples such as Mozart's musical dice game [10] demonstrated that pre-structured musical fragments can be recombined to produce musically valid outcomes. In sound and music computing, later studies expanded this idea toward interactive systems in which users can explore musical variation spaces rather than compose note by note.

From a computational perspective, these approaches can be broadly categorized into rule-based, probabilistic, and neural methods. Rule-based approaches explicitly encode musical knowledge using deterministic constraints. Probabilistic approaches extend these rules by assigning stochastic weights, allowing multiple possible interpretations of musical structure. Neural approaches further improve generative performance by learning complex musical patterns from large-scale datasets.

Within this landscape, the GTTM Analyzer is a collective term for a family of computational analysis methods based on the Generative Theory of Tonal Music (GTTM) [1]. It includes ATTA, a rule-based method for automatic time-span tree construction [11–13]; σGTTM, a probabilistic extension that handles multiple interpretations of musical structure [14–17]; and deepGTTM, a neural-network-based method for estimating hierarchical musical structures [18–21]. Together, these methods span rule-based, probabilistic, and neural approaches to computational GTTM analysis.

Their common objective is to acquire hierarchical time-span trees from musical input. The resulting time-span trees serve as the structural representation on which

GTTM-based melody morphing operates, forming the computational foundation of the Melody Slot Machine [2, 3].

Melody Slot Machine extends this line of work by turning structure-preserving melody generation into an interactive selection interface [22,23]. Rather than requiring users to compose explicitly, the system allows them to navigate a precomputed space of GTTM-based melodic variations through a simple dial or tablet-based interaction. The generated melody is then directly coupled to robotic instrumental performance via RoboSax.

6.2 Robotic Musical Instruments

Robotic musical instruments have been studied as platforms for musical expression, novel performance practice, and human–machine collaboration [24,25]. Prior work has explored robotic percussion, keyboard, string, and wind instruments, often focusing on precision, extended technique, or forms of performance that exceed human physical limitations [26, 27].

A useful distinction is between *fully automated* and *human-in-the-loop* instruments. The former spans mechanical music boxes and automata [28–30] to robots for violin, piano, guitar, trumpet, flute, and saxophone [27, 31–38], whereas the latter retain part of the performance process to preserve embodiment and agency [25]. Semi-automatic instruments in this category assist human players through devices such as fingering or embouchure support [39–42].

RoboSax belongs to the latter category. It automates fingering while preserving human breath and articulation, and extends this model to ensemble performance.

6.3 Accessible Musical Interfaces for Non-Experts

Many musical interface studies have aimed to lower the barrier to musical participation for non-experts [43–45]. In the NIME and SMC communities, this has led to the development of systems that simplify gesture acquisition, reduce the need for instrumental training, or translate intuitive actions into musically meaningful outcomes [46–48].

The proposed system shares this goal, but supports both *musical decision making* and *instrumental embodiment*. Melody Slot Machine lets users choose structurally valid melody variations, while RoboSax lets them produce wind-instrument sounds without mastering conventional fingering. The result is a single interface that links accessible composition and accessible performance.

6.4 Embodied Human–Machine Musical Interaction

Embodied interaction has become an important perspective in music technology research [45, 49]. Musical meaning is not only represented symbolically but is also shaped through bodily action, instrument handling, and performer feedback. Accordingly, many interactive music systems aim to preserve or augment the bodily aspects of performance rather than replace them entirely [50, 51].

Hybrid human–machine systems are particularly relevant in this context because they distribute musical agency across computational processes, mechanical systems, and human performers [25]. RoboSax Melody Slot Machine adopts this approach by assigning melody generation to computation and user interaction, fingering to the robot, and breath and articulation to the performer.

7. EXPERIMENTS

We evaluated RoboSax through measurements of operating speed, acoustic comparison, and use studies.

7.1 Evaluation of operating speed

We measured the time required for the prototype robo-alto saxophone to close each key using a 240 Hz high-speed camera. For comparison, the same movements were measured for an experienced saxophonist.

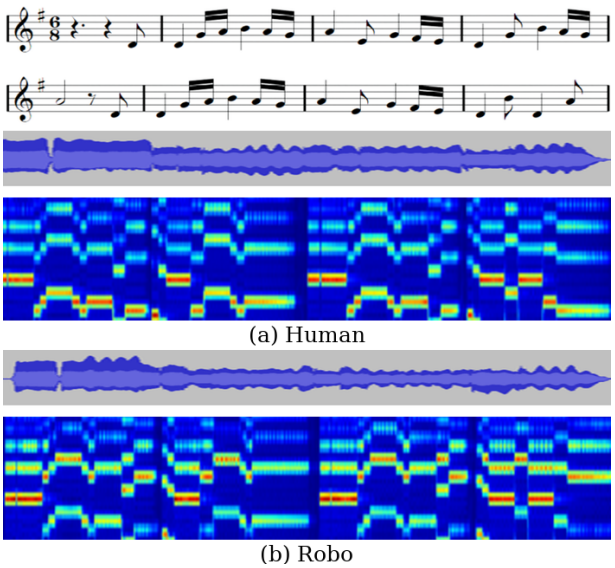
Table 1 shows that the robot closed every key faster than the human performer. The average closing time was 32 ms for the robot and 47 ms for the human; the G key was slowest for the human because it is operated by the left-hand ring finger.

Table 1. Operating speed of Robo-Alto Sax [ms].

	Oct	B	Bb	A	G	F	E	D	Ave.
Robo	29.2	25.0	33.3	29.2	37.5	33.3	33.3	37.5	32.3
Human	29.2	45.8	50.0	50.0	58.3	45.8	50.0	45.8	46.9

7.2 Acoustic analysis

We compared robotic and human performances of the piece shown in Fig. 6. Both were recorded while following click tracks, and the resulting waveforms and Constant-Q transforms were broadly similar. The robot showed stronger overtones, possibly due to acoustic changes caused by the metal structure attached to the instrument.



(a) Human

(b) Robo

Figure 6. Comparison of human and RoboSax performances: (a) human performance and (b) RoboSax performance.

We also conducted a listening test. Because servo noise could reveal the robotic take, similar mechanical noise was added to the human recording. Table 2 summarizes the results. Among 10 respondents, four identified the human performance correctly, three chose the robot, and three reported no difference. Seven respondents preferred the robot's performance in this small-scale listening test. While this result is not statistically generalizable, it suggests a possible tendency that reduced fingering demands may support stable tone production under some conditions.

Table 2. Results of instrumental listening test

Response	Human	Robot	No Difference
Correct identification	4	3	3
Better performance	2	7	1

7.3 Use Studies

This subsection presents three case studies involving beginners, ensemble players, and a professional saxophonist.

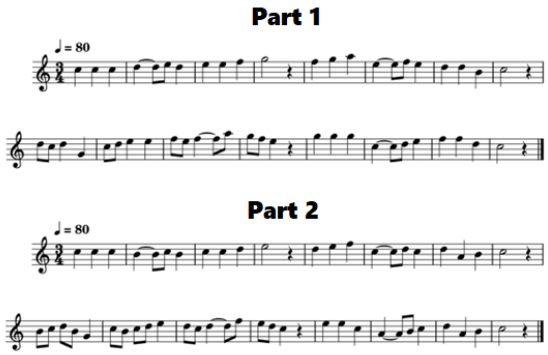


Figure 7. Example musical score used in the beginner RoboSax performance study.

7.3.1 Beginner performance

Two students with no saxophone experience and no ability to read staff notation performed on the robot alto saxophone after a 30-minute introduction to reed sound production. Using the score shown in Fig. 7, they were able to play a simple tune without mastering conventional fingering. This suggests that RoboSax lowers the entry barrier to wind-instrument performance by separating breath production from fingering.

7.3.2 Ensemble study

With seven members of a university wind orchestra club, we conducted monthly ensemble sessions from March 2024 to March 2025 using an SATB RoboSax quartet, with occasional AATB reconfiguration. A public mini-concert was held in November 2024. The long-term study revealed three main issues: breath timing relative to key motion, difficulty using the GUI during ensemble playing, and performer fatigue caused by instrument weight.



(a) RoboSax practice



(b) RoboSax mini concert

Figure 8. RoboSax ensemble session with wind orchestra members.

We addressed these issues by compensating for robot–human delay, providing printed notation, adding stands, and using single-ear click monitoring. After these refinements, the quartet performed publicly without major problems. Players reported reduced fingering difficulty and expressive possibilities such as vibrato, although the system was more fatiguing than conventional saxophone performance. These results suggest that RoboSax can support intermediate players as well as beginners, but still requires system-specific practice.

7.3.3 Professional performance with melody morphing

After refining the system through ensemble use, we conducted an experiment with a professional saxophonist using a combined RoboSax and melody-morphing system. A tablet-based interface allowed another participant to modify the melody in real time while the saxophonist performed it on a soprano RoboSax. This study demonstrates that the system can support shared performance in which one participant shapes melodic structure and another realizes it physically.

8. DISCUSSION

Across the experiments, a consistent design implication is that RoboSax is most effective when it is treated not as a fully autonomous instrument but as a role-sharing performance system. Automating fingering removes a major technical barrier, while leaving breath and articulation to the performer preserves bodily engagement and expressive control. This balance helps explain why the system was usable by beginners, adaptable to ensemble playing, and still musically meaningful for a professional saxophonist.

These findings should be interpreted as evidence for a role-sharing performance model rather than as a claim

that robotic fingering generally improves saxophone performance. In the proposed system, computational processing provides structurally coherent melodic candidates, RoboSax executes the discrete fingering required for pitch production, and the human performer remains responsible for breath, articulation, timing, and musical expression. This division of roles preserves important embodied aspects of wind-instrument performance while reducing the technical barrier associated with fingering. The contribution of the system therefore lies not only in automating saxophone key operation, but also in connecting three usually separate activities—melodic decision making, instrumental actuation, and acoustic sound production—within a single interactive performance loop.

A limitation of the current interaction design is that users cannot directly manipulate individual notes during performance. Instead, they select from a precomputed set of GTTM-based melodic variations. This constraint may limit fine-grained control, but it is also consistent with the primary objective of the system: to broaden access to two complementary musical experiences, namely melodic creation through the Melody Slot Machine and embodied performance through RoboSax. The design therefore prioritizes accessibility and exploratory interaction over low-level real-time editing.

Similarly, separating breath production from fingering should not be regarded as a replacement for conventional saxophone training. In traditional instrumental education, coordinating breath, articulation, and fingering is an essential learning goal. RoboSax is instead intended as an interface that allows non-expert users to experience wind instrument performance without first mastering conventional fingering. At the same time, isolating breath control may offer new learning perspectives, such as focusing on tone stability and articulation timing. Whether this separation can support a stepwise learning process toward conventional saxophone performance remains an important topic for future work.

The evaluations reported in this paper should also be regarded as preliminary. The operating-speed comparison involved a single experienced saxophonist, and the listening test was conducted with 10 respondents. Therefore, the results do not provide statistically generalizable evidence of performance quality or listener preference. Instead, they serve as initial observations indicating that RoboSax can operate at practical speeds and that its acoustic output can be perceived as comparable to human performance under the tested conditions. Future work will include larger-scale experiments with more performers, more listeners, and more diverse musical materials.

At the same time, the use studies highlight practical constraints for future development. Successful performance depended on timing compensation between robot and performer, alternative score or cue support during ensemble use, and physical support to reduce fatigue caused by the added mechanism. These observations suggest that future robotic wind instruments should be designed not only for mechanical accuracy but also for rehearsal workflow, ergonomics, and shared control in performance settings.

9. CONCLUSION

This paper presented RoboSax Melody Slot Machine, a system that integrates GTTM-based melody morphing with robotic saxophone fingering. By assigning melody selection to the interface, fingering to the robot, and breath and articulation to the performer, the system supports a hybrid form of wind-instrument performance in which composition and physical sound production occur in the same interaction loop.

The experiments and use studies provide preliminary evidence for the feasibility of this design from multiple perspectives. The operating speed measurements showed that RoboSax can execute key motions at practical speeds, while the acoustic comparison indicated that robotic and human performances can be perceptually similar under the tested conditions. The use studies further suggest that the system can support beginners who lack conventional fingering skills, ensemble players who must coordinate with other performers, and professional musicians engaged in real-time collaborative performance.

Taken together, these results suggest that RoboSax Melody Slot Machine can broaden access to musical experiences by connecting computational melody generation, embodied instrumental interaction, and shared performance practice. The extension to ensemble performance also points toward applications in hybrid concerts, educational settings, and interactive installations.

In this sense, the system also provides an experimental platform for examining how algorithmic musical structure, robotic actuation, and human breath can be coordinated within a single performance practice. Such a perspective is important for clarifying which musical actions should remain embodied and which can be delegated to computational or robotic processes.

Future work includes exploring richer interaction techniques for controlling musical structure, investigating expressive control of robotic wind instruments, and developing larger robotic ensembles that support more complex musical textures. Longitudinal studies will also be needed to examine how repeated use changes performers' coordination strategies and sense of musical agency.

10. REFERENCES

[1] F. Lerdahl and R. Jackendoff, *A Generative Theory of Tonal Music*. MIT Press, 1983.

[2] M. Hamanaka, K. Hirata, and S. Tojo, "Melody morphing method based on gttm," in *Proceedings of the International Computer Music Conference*, 2008, pp. 155–158.

[3] M. Hamanaka, K. Hirata, and S. Tojo, "Implementation of melodic morphing based on generative theory of tonal music," *J. New Music Res.*, vol. 51, no. 1, pp. 86–102, 2022.

[4] G. Koutaki and M. Hamanaka, "Robo-sax ensemble system with melody morphing," in *Extended Abstracts for the Late-Breaking Demo Session of the International Society for Music Information Retrieval Conference*, 2025.

[5] L. Hiller and L. Isaacson, *Experimental Music: Composition with an Electronic Computer*. McGraw-Hill, 1959.

[6] D. Cope, *Computer Models of Musical Creativity*. MIT Press, 2005.

[7] G. Papadopoulos and G. A. Wiggins, "Ai methods for algorithmic composition: A survey, a critical view and future prospects," in *Proceedings of the AISB'99 Symposium on Musical Creativity*, 1999, pp. 110–117.

[8] J. D. Fernandez and F. Vico, "Ai methods in algorithmic composition: A comprehensive survey," *J. Artif. Intell. Res.*, vol. 48, pp. 513–582, 2013.

[9] G. Nierhaus, *Algorithmic Composition: Paradigms of Automated Music Generation*. Springer, 2009.

[10] S. A. Hedges, "Dice music in the eighteenth century," *Music and Letters*, vol. 59, no. 2, pp. 180–187, 1978.

[11] M. Hamanaka, K. Hirata, and S. Tojo, "Atta: Automatic time-span tree analyzer based on extended gttm," in *Proceedings of the 6th International Conference on Music Information Retrieval*, 2005, pp. 358–365.

[12] M. Hamanaka, K. Hirata, and S. Tojo, "Implementing generative theory of tonal music," *Journal of New Music Research*, vol. 35, no. 4, pp. 249–277, 2006.

[13] M. Hamanaka and K. Hirata, "Implementing methods for analyzing music based on lerdahl and jackendoff's generative theory of tonal music," in *Computational Music Analysis*. Springer, 2016, pp. 221–249.

[14] Y. Miura, M. Hamanaka, K. Hirata, and S. Tojo, "Use of decision tree to detect gttm group boundaries," in *Proceedings of the International Computer Music Conference*, 2009, pp. 125–128.

[15] K. Kanamori, M. Hamanaka, and J. Hoshino, "Method to detect gttm local grouping boundaries based on clustering and statistical learning," in *Proceedings of the International Computer Music Conference / Sound and Music Computing Conference*, 2014.

[16] M. Hamanaka, K. Hirata, and S. Tojo, "σgttm iii: Learning-based time-span tree generator based on pcfg," in *Proceedings of the International Symposium on Computer Music Multidisciplinary Research*, 2015, pp. 303–317.

[17] E. Nakamura, M. Hamanaka, K. Hirata, and K. Yoshii, "Tree-structured probabilistic model of monophonic written music based on the generative theory of tonal music," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, 2016, pp. 276–280.

[18] M. Hamanaka, K. Hirata, and S. Tojo, "deepgttm-i: Local boundaries analyzer based on deep learning technique," in *Proceedings of the International Symposium on Computer Music Multidisciplinary Research*, 2016, pp. 8–20.

[19] M. Hamanaka, K. Hirata, and S. Tojo, "deepgttm-ii: Automatic generation of metrical structure based on deep learning technique," in *Proceedings of the Sound and Music Computing Conference*, 2016, pp. 221–249.

[20] M. Hamanaka, K. Hirata, and S. Tojo, "deepgttm-iii: Multi-task learning with grouping and metrical structures," in *Proceedings of the International Symposium on Computer Music Multidisciplinary Research*, 2017, pp. 238–251.

[21] M. Hamanaka, K. Hirata, and S. Tojo, "Time-span tree leveled by duration of time-span," in *Proceedings of the 15th International Symposium on Computer Music Multidisciplinary Research*, 2021, pp. 155–164.

[22] M. Hamanaka, "Melody slot machine: A controllable holographic virtual performer," in *Proceedings of the 27th ACM International Conference on Multimedia*, 2019.

[23] M. Hamanaka, T. Nakatsuka, and S. Morishima, "Melody slot machine," in *ACM SIGGRAPH 2019 Emerging Technologies*, 2019.

[24] E. Singer, K. Larke, and D. Bianciardi, "LEMUR GuitarBot: MIDI robotic string instrument," in *Proceedings of the 2003 Conference on New Interfaces for Musical Expression*, 2003.

[25] G. Weinberg, S. Driscoll, and R. M. Parry, "HAILE - an interactive robotic percussionist," in *Proceedings of the International Computer Music Conference*, 2005, pp. 622–625.

[26] E. Singer, J. Feddersen, C. Redmon, and B. Bowen, "Lemur's musical robots," in *Proceedings of the 2004 Conference on New Interfaces for Musical Expression*, 2004, pp. 181–184.

[27] J. Solis, K. Taniguchi, T. Ninomiya, T. Yamamoto, and A. Takanishi, "Development of waseda flutist robot wf-4riv: Implementation of auditory feedback system," in *Proceedings of the 2008 IEEE International Conference on Robotics and Automation*, 2008, pp. 3654–3659.

[28] N. H. Fletcher and T. D. Rossing, *The Physics of Musical Instruments*, 2nd ed. Springer, 1998.

[29] A. W. J. G. Ord-Hume, *Clockwork Music: An Illustrated History of Mechanical Musical Instruments from the Musical Box to the Pianola, from Automaton Lady Virginal Players to Orchestrion*. Allen and Unwin, 1973.

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

[31] M. Kajitani, "Development of musician robots," *J. Robot. Mechatron.*, vol. 1, no. 3, pp. 254–255, 1989.

[32] G. Byambatsogt, L. Choimaa, and G. Koutaki, "Data generation from robotic performer for chord recognition," *IEEJ Trans. Electron. Inf. Syst.*, vol. 141, no. 2, pp. 205–213, 2021.

[33] K. Kodama and G. Koutaki, "Development of guitar playing robot by pwm control of solenoid," in *Proceedings of the 2019 IEEE 8th Global Conference on Consumer Electronics*, 2019, pp. 291–293.

[34] A. Kapur, "A history of robotic musical instruments," in *Proceedings of the International Computer Music Conference*, 2005.

[35] S. R. Ness, S. Trail, P. F. Driessen, W. A. Schloss, and G. Tzanetakis, "Music information robotics: Coping strategies for musically challenged robots," in *Proceedings of the 12th International Society for Music Information Retrieval Conference*, 2011, pp. 567–572.

[36] J. Lin, K. Mao, N. Yuya, C. Sarah, and A. Takanishi, "Development of performance system with musical dynamics expression on humanoid saxophonist robot," *IEEE Robot. Autom. Lett.*, vol. 4, no. 2, pp. 1684–1690, 2019.

[37] Y. Kusuda, "Toyota's violin-playing robot," *Ind. Robot*, vol. 35, no. 6, pp. 504–506, 2008.

[38] A. Kapur and E. Singer, "A retrieval approach for human/robotic musical performance," in *Proceedings of the 7th International Conference on Music Information Retrieval*, 2006, pp. 363–364.

[39] Y. Kurosawa and K. Suzuki, "Robot-assisted playing with fingering support for a saxophone," in *Proceedings of the International Computer Music Conference*, 2010.

[40] J. Solis, Y. Sugita, K. Petersen, and A. Takanishi, "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.*, vol. 86, pp. 174–183, 2016.

[41] J. Solis, M. Bergamasco, K. Chida, S. Isoda, and A. Takanishi, "The anthropomorphic flutist robot wf-4riv teaching flute playing to beginner students," in *Proceedings of the 2004 IEEE International Conference on Robotics and Automation*, 2004, pp. 146–151 Vol.1.

[42] F. Heller, I. M. C. Ruiz, and J. Borchers, "An augmented flute for beginners," *Proceedings of the International Conference on New Interfaces for Musical Expression*, pp. 15–19, 2017.

[43] S. Jordà, "Digital instruments and players: Part i – efficiency and apprenticeship," in *Proceedings of the 2004 Conference on New Interfaces for Musical Expression*, 2004, pp. 59–63.

[44] R. Fiebrink and P. R. Cook, "The wekinator: A system for real-time, interactive machine learning in music," in *Proceedings of the 11th International Society for Music Information Retrieval Conference*, 2010.

[45] M. Leman and P.-J. Maes, "The role of embodiment in the perception of music," *Empir. Musicol. Rev.*, vol. 9, no. 3–4, pp. 236–246, 2015.

[46] P. Cook, "Principles for designing computer music controllers," in *Proceedings of the 2001 Conference on New Interfaces for Musical Expression*, 2001.

[47] D. Wessel and M. Wright, "Problems and prospects for intimate musical control of computers," *Comput. Music J.*, vol. 26, no. 3, pp. 11–22, 2002.

[48] E. Miranda and M. Wanderley, *New Digital Musical Instruments: Control and Interaction Beyond the Keyboard*. A-R Editions, 2006.

[49] M. Leman, *Embodied Music Cognition and Mediation Technology*. MIT Press, 2007.

[50] M. M. Wanderley and N. Orio, "Evaluation of input devices for musical expression: Borrowing tools from hci," *Comput. Music J.*, vol. 26, no. 3, pp. 62–76, 2002.

[51] G. Paine, "Interaction as material: The techno-somatic dimension," *Organised Sound*, vol. 20, no. 1, pp. 82–89, 2015.