Melody Slot Machine with RoboSax

MASATOSHI HAMANAKA, RIKEN, Japan
GOU KOUTAKI, Kumamoto University, Japan

Additional Key Words and Phrases: Melody Slot Machine, RoboSax, GTTM, Melody morphing, AI music generation, Saxophone robotics

1 Program Notes

Melody Slot Machine with RoboSax is our AI-based system that combines an interactive melody variation interface with a robotic saxophone performer (Fig. 1). Melody Slot Machine is a dial with the staves of music displayed on an iPad, which can be rotated to change the melody variations. The melody variations are generated on the basis of the AI-based melody-morphing method (Fig. 2) and can be partially switched to another variation without any significant change in the overall melody structure and with no musical breakdown. The microcomputer in the RoboSax receives the MIDI note-on signal from Melody Slot Machine and moves the servomotor so that the fingering corresponds to the note number.

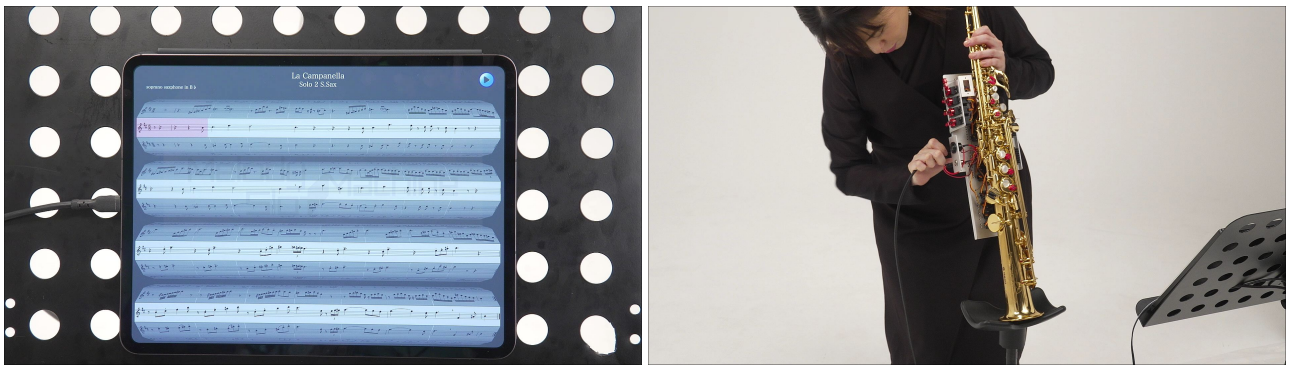

Fig. 1. Melody Slot Machine and RoboSax

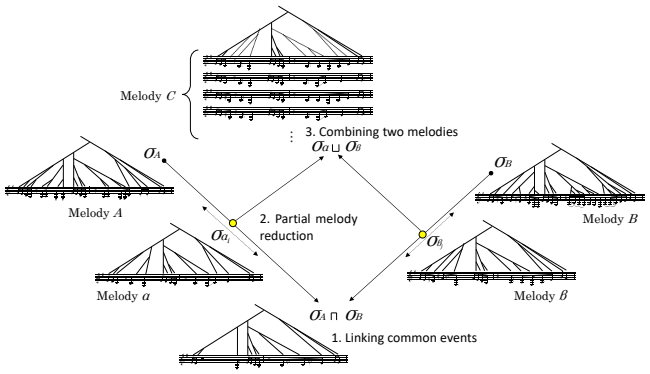

Fig. 2. Melody-morphing method

Authors' Contact Information: Masatoshi Hamanaka, RIKEN, Tokyo, Japan; Gou Koutaki, Kumamoto University, Kumamoto, Japan.

This work is licensed under a Creative Commons Attribution 4.0 International License.
NIME '26, June 23–26, 2026, London, UK
© 2026 Copyright held by the owner/author(s).

2 Project Description

Our project combines an interactive melody generation system based on the Generative Theory of Tonal Music (GTTM) [10] with a robotic saxophone performer. The GTTM is used to analyze musical structure and obtain a time-span tree, which is then employed to generate meaningful melody variations. The melody variations in Melody Slot Machine can be partially switched to another variation without any significant change in the overall structure of the time-span tree and with no musical breakdown.

Previous time-span-tree analyzers [1, 8] exhibited low performance because they analyzed in a bottom-up manner using only local information. We have thus constructed deepGTTM systems, which introduce AI into GTTM analysis and dramatically improve the performance of time-span-tree analysis. We developed separate deep-learning-based analyzers: deepGTTM-I for local boundary detection and deepGTTM-II for metrical structure generation [3, 4], which were integrated as deepGTTM-I&II [5]. Subsequently, deepGTTM-III extended this approach to simultaneously learn grouping and metrical structures [6], enabling more sophisticated structural understanding. Most recently, deepGTTM-IV introduces adaptive melody morphing capabilities [2, 7] specifically designed for interactive music systems like Melody Slot Machine. The GTTM is a compilation of musical knowledge gained from composers' experience. Thus, we have shown that a composer's musical knowledge can be learned using deep neural networks across multiple generations of deep learning models.

RoboSax uses a microcomputer and servomotors (Fig. 3) attached to a saxophone to automatically execute the fingering of melodies with various variations. The microcomputer receives the MIDI note-on signal output from Melody Slot Machine and moves the servomotor so that the fingering corresponds to the note number. The keys of the saxophone and servomotors are connected with wires. When the servomotor is moved and a wire is pulled, the key moves and the tone hole closes. RoboSax can finger faster and more accurately than even the most experienced saxophonist.

Robot musical instruments can be classified into two categories: fully autonomous robots and semi-autonomous robots. Fully autonomous robots perform music entirely without human intervention, composing and executing performances autonomously [11, 13]. In contrast, semi-autonomous robots like RoboSax [9, 12] automate specific aspects of musical performance, such as fingering, while allowing human musicians to focus on expressive elements. This hybrid approach enhances musical artistry because the human performer can concentrate on the most expressive and interpretive aspects of the performance—in RoboSax's case, breath control, tonguing articulation, and dynamic shaping. The human performer's ability to shape tone color, phrasing, and emotional nuance becomes paramount when the mechanistic constraints of fingering are removed. RoboSax thus represents a paradigm where automation amplifies rather than diminishes human artistic expression, creating a new form of human-robot musical collaboration.

Melody Slot Machine with RoboSax challenges human composers and performers through AI and robotics technology. The inclusion of humans also emphasizes their importance. For example, humans selecting the dials of Melody Slot Machine add an element of amusement and improvisation to a piece of music. RoboSax's servomotors are fast and precise, but it is the performer's blowing and tonguing that give the tone its expression. We hope that the Melody Slot Machine with RoboSax will give you an idea of how AI and robot technologies and humans can better communicate with each other.

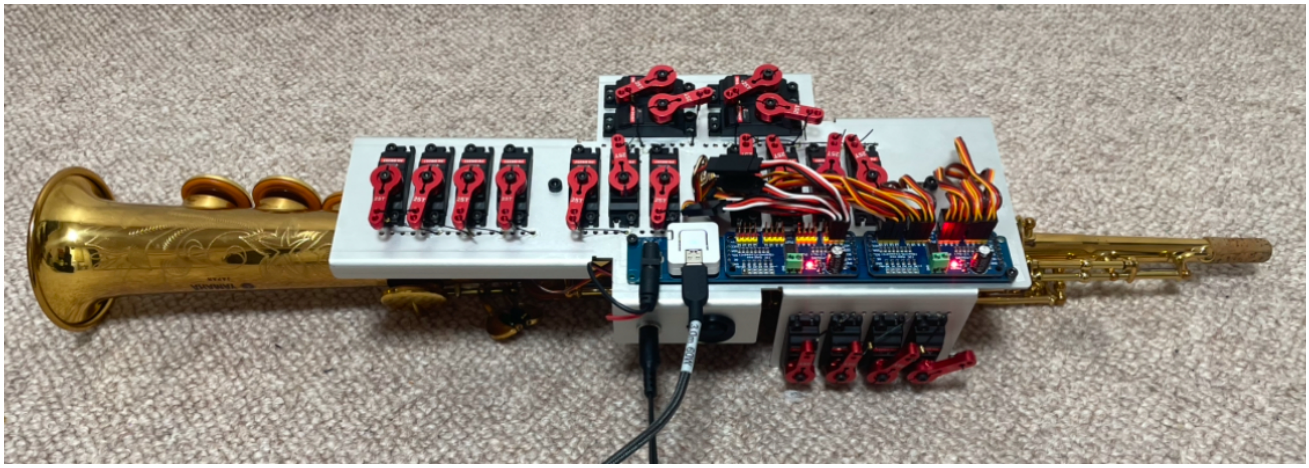

Fig. 3. Servomotors

3 Technical Notes

Title of the piece: Melody Slot Machine with RoboSax
Length: The minimum playing time is 1 minute and 50 seconds but can be extended to 5 minutes and 30 seconds by adding loops. The saxophone melody changes with each loop.
Technical Requirements and Setup Specifications:

- Projector and screen: 1 (We will send a single HDMI feed to the venue projector. The feed combines the iPad screen and the GoPro video of the auto-fingering saxophone using a four-screen HDMI video compositor.)
- Speakers: 2 (Stereo output required for playing the accompaniment part.)
- Tables: Not required.
- Power: 100–240V, 50/60Hz, 5A
- Space: Minimum space is 2 square meters.

We bring an iPad, a GoPro, a mobile display, a four-screen HDMI video compositor, an HDMI adapter for the iPad, a power supply, and a 19-inch 3U rack case.

We also bring all required cables and adapters: a 3-m USB-C cable from the iPad to the HDMI adapter, a 2-m HDMI cable from the iPad HDMI adapter to the four-screen HDMI video compositor, an HDMI cable from the GoPro output to the four-screen HDMI video compositor, a 2-m HDMI cable from the compositor to the mobile display, and a 10-m HDMI cable from the compositor to the venue projector.
Setup time: 30 minutes.

4 Media Links

Project page and video: https://aip.riken.jp/news/robosax/?lang=en

5 Ethical Standards

This music submission is aligned with NIME's ethical standards. This work involved no human-subject study or animal experimentation. No conflicts of interest, financial or non-financial, are declared.

Acknowledgments

This work was supported by JSPS KAKENHI Grant number 24H00748.

References

[1] Masatoshi Hamanaka, Keiji Hirata, and Satoshi Tojo. 2006. Implementing "Generative Theory of Tonal Music". *Journal of New Music Research* 35, 4 (2006), 249–277. https://doi.org/10.1080/09298210701563238 arXiv:https://doi.org/10.1080/09298210701563238

[2] Masatoshi Hamanaka, Keiji Hirata, and Satoshi Tojo. 2008. Melody Morphing Method Based on GTTM. In *Proceedings of the International Computer Music Conference*. Belfast, Northern Ireland, 1–8.

[3] Masatoshi Hamanaka, Keiji Hirata, and Satoshi Tojo. 2016. deepGTTM-I: Local Boundaries Analyzer Based on Deep Learning Technique. In *Proceedings of the International Symposium on Computer Music Multidisciplinary Research*. Marseille, France, 8–20.

[4] Masatoshi Hamanaka, Keiji Hirata, and Satoshi Tojo. 2016. deepGTTM-II: Automatic Generation of Metrical Structure Based on Deep Learning Technique. In *Proceedings of the Sound and Music Computing Conference*. Hamburg, Germany, 221–249.

[5] Masatoshi Hamanaka, Keiji Hirata, and Satoshi Tojo. 2016. deepGTTM-I&II: Local Boundary and Metrical Structure Analyzer Based on Deep Learning Technique. In *Proceedings of the International Symposium on Computer Music Multidisciplinary Research*. Marseille, France, 1–15.

[6] Masatoshi Hamanaka, Keiji Hirata, and Satoshi Tojo. 2017. deepGTTM-III: Multi-task Learning with Grouping and Metrical Structures. In *Proceedings of the International Symposium on Computer Music Multidisciplinary Research*. Matosinhos, Portugal, 212–230.

[7] Masatoshi Hamanaka, Keiji Hirata, and Satoshi Tojo. 2023. deepGTTM-IV: Deep Learning Based Time-Span Tree Analyzer of GTTM. In *Proceedings of the International Symposium on Computer Music Multidisciplinary Research*. Tokyo, Japan, 262–281.

[8] Masatoshi Hamanaka and Satoshi Tojo. 2009. Interactive GTTM Analyzer. In *Proceedings of the 10th International Society for Music Information Retrieval Conference*. Kobe, Japan, 523–526.

[9] Ajay Kapur, Michael Darling, James Murphy, Jordan Hochenbaum, Olin Vallis, and Courtney Bahn. 2011. The KarmetiK NotomotoN: A New Breed of Musical Robot for Teaching and Performance. In *Proceedings of the International Conference on New Interfaces for Musical Expression*. Oslo, Norway, 228–231.

[10] Fred Lerdahl and Ray Jackendoff. 1983. *A generative theory of tonal music*. The MIT Press, Cambridge. MA.

[11] Jorge Solis, Karen Petersen, Tsutomu Ninomiya, et al. 2009. Development of anthropomorphic musical performance robots: From understanding the nature of music performance to its application to entertainment robotics. In *Proceedings of the 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*. St. Louis, MO, 2519–2524.

[12] Reo Ueta and Masatoshi Hamanaka. 2025. Performance Design in Semi-Autonomous Robot Musical Instruments. In *Proceedings of NIME*.

[13] Gil Weinberg and Scott Driscoll. 2006. Toward robotic musicianship. *Computer Music Journal* 30, 4 (2006), 14–20.