Repurposing a Rhythm Accompaniment System for Pipe Organ Performance

Nicholas Evans
Music Technology Group
Universitat Pompeu Fabra
Barcelona, Spain
nicholas.evans@upf.edu

Behzad Haki
Music Technology Group
Universitat Pompeu Fabra
Barcelona, Spain
behzad.haki@upf.edu

Sergi Jordà
Music Technology Group
Universitat Pompeu Fabra
Barcelona, Spain
sergi.jorda@upf.edu

Figure 1: A performer-level image of the performance at *Palau Güell*

Abstract

This paper presents an overview of a human-machine collaborative musical performance by *Raül Refree* utilizing multiple MIDI-enabled pipe organs at *Palau Güell*, as part of the Organic concert series. Our earlier collaboration focused on live performances using drum generation systems, where generative models captured rhythmic transient structures while ignoring harmonic information. For the organ performance, we required a system capable of generating harmonic sequences in real-time, conditioned on *Refree*'s performance. Instead of developing a comprehensive state-of-the-art model, we integrated a more traditional generative method to convert our pitch-agnostic rhythmic patterns into harmonic sequences. This paper details the development process, the creative and technical considerations behind the final performance, and a reflection on the efficacy and adaptability of the chosen methodology.

Keywords

Real-time accompaniment generation, live performance, symbolic music

NIME '25, June 24–27, 2025, Canberra, Australia

1 Introduction

Palau Güell, a historic architectural landmark in Barcelona designed by *Antoni Gaudí*, is renowned for its distinctive design and acoustics. The *Organic*[1] concert series, hosted at this venue, featured a series of performances aimed at exploring the dynamic interplay between instruments and acoustics of spaces. As such,

[1] https://inici.palauguell.cat/en/organic/

the main aim of the series was to create immersive sonic experiences for audiences by utilizing the venue and its distinct acoustics as an integral part of the performance.

A key feature of *Palau Güell* is its permanent organ, crafted by *Albert Blancafort*[2], a renowned local organ maker. This MIDI-enabled instrument was custom-designed according to the acoustics of the space. Moreover, two portable organs, also built by *Blancafort*, were available for use in the performance. While the main organ remained fixed in the venue, the smaller ones were intended to be positioned based on the spatial and artistic needs of each performance.

Over the past two years, we have collaborated with *Raül Refree* on a series of live performances focusing on human-machine collaborative improvisation. For these performances, we developed a real-time drum accompaniment generation system called *GrooveTransformer*. Invited to participate in the Organic series, *Refree* proposed using this opportunity to adapt our previous work to a new context.

GrooveTransformer is a transformer-based model [15] trained for real-time drum accompaniment generation [4, 5]. Unlike our previous work, which focused solely on rhythm generation, this organ performance required adapting our system to incorporate pitch. An additional constraint of the performance was *Refree*'s insistence to not rely on any pre-composed material for both training or performance[3], wanting the accompaniment to solely reflect the improvised moment.

While we considered pairing *GrooveTransformer* with an existing secondary real-time system, we identified two main drawbacks. First, deep neural models inevitably carry stylistic biases inherited from their training data, constraining how the generated output could reflect the real-time improvised performance. Second, we had only a short window of time for experimentation with very limited rehearsal opportunities. Given these constraints, we sought to investigate how a state-of-the-art model could be expanded and recontextualized by integrating a comparatively simple yet effective traditional generative model—assigning pitch to *GrooveTransformer*'s rhythmic sequences via a Markov-based system trained in real-time. This approach allowed us to inject minimal prior bias and implement direct controls to enhance the expressive affordances of the system in live performance.

2 Related Work

Symbolic-to-symbolic generative systems, designed as musical agents, have the potential to function both autonomously and semi-autonomously, crafting music in response to external cues while acting as dynamic collaborators in live performances. Over time, the techniques used have evolved to address various musical goals and technical challenges.

Early systems, such as *The Continuator* [10], often used Markov models to learn from and respond to a musician's real-time input. Likewise, evolutionary algorithms have been explored for certain types of accompaniment generation [3, 9].

A more recent wave of systems leverages neural networks for greater modeling capacity. Google's *AI-Duet*[4] enables a call-and-response style of piano improvisation by interpreting a user's input in real time. *SongDriver* [16], *BachDuet* [2], *RL-Duet* [6], *ReaLchords* [17], and *ReaLJam* [12]—introduce musical agents capable of shaping melodic or harmonic structures in response to a human performer's actions. These systems are either trained on very specific repertoires, or are aimed at a very specific type of accompaniment generation.

Notochord [13], however, is trained on a diverse set of recordings that support a variety of generative interactions. Its design enables instrument designers to create custom digital musical instruments around the model, without limiting the deployment context. Several projects have employed Notochord in diverse contexts, where human performers collaborate with Notochord in real-time [1, 7].

3 Venue and Instruments

The venue comprised three floors, as illustrated in Figure 2. The audience was seated on the first floor, while the main organ—a fixed, non-portable instrument—was positioned on the third floor. This organ supported remote control via MIDI, enabling the performer to remain at the audience level. However, to minimize visual distractions and encourage active listening while emphasizing the spatial characteristics of the venue, the performer ultimately played the organ directly from the third floor, remaining completely out of sight from the audience.

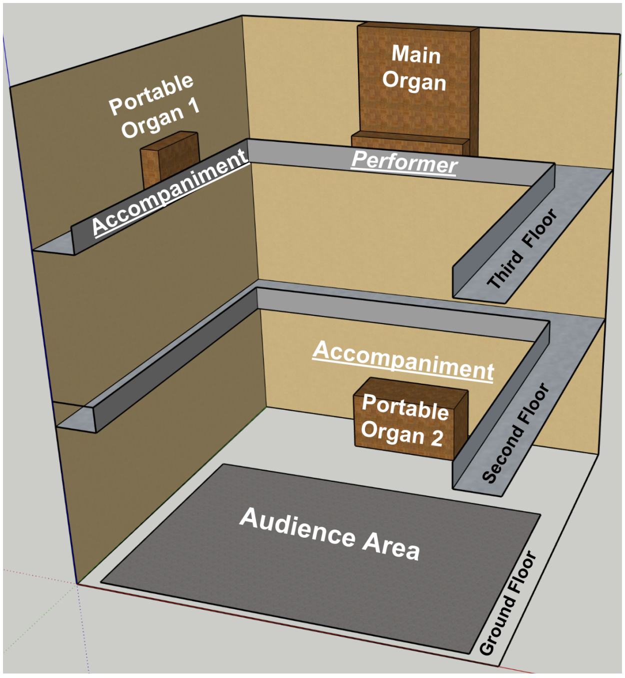

Figure 2: Overview of the venue and instruments

Initially, the portable MIDI-enabled organs, intended for playback of the generated accompaniments, were to be placed at the audience level. However, given the significant distance between the audience and the performer, concerns arose about the audibility of the generations near the main organ, where *Refree* would be located. Consequently, one of the accompaniment organs was positioned on the third floor as a monitor for the performer, while the other remained on the first floor.

Notably, despite *Palau Güell*'s reflective surfaces and large volume, the space exhibits relatively low reverberation, unlike

[2] https://www.orguesblancafort.com/en/blancafort-om/
[3] Had we been able to generate the accompaniments according to pre-composed material, we could have considered offline state-of-the-art generative methods such as [14]. However, the real-time requirement demanded a lightweight approach that could dynamically respond to an unplanned improvisational context.
[4] https://experiments.withgoogle.com/ai-duet

traditional organ venues that are typically highly reverberant. This unique characteristic allows for playback of shorter transient sequences that are not masked in the reverberation of the space. As detailed in the next section, these acoustic and spatial considerations played a crucial role in shaping the performance design.

4 Performance Development

In addition to considering the venue's layout, minimal rehearsal time, and the performance requirements, our intention was to design a system that could facilitate a completely improvised performance without inherited stylistic biases from a pre-composed or existing dataset. This meant creating a framework that could dynamically respond to the performer's input in real-time, without relying on prior training or pre-composed material. To accomplish this, we implemented a Markov-based model that integrates with *GrooveTransformer* and actively listens to the live performance to inject harmonic content and note durations to the generated rhythms. However, rather than functioning as a fully autonomous accompaniment generator, the system was designed to be actively supervised by a secondary performer using a set of utilities for pre- and post-generation processing.

4.1 Overall Structure of the Accompaniment System

The system consists of four main components: (1) a real-time performance on the main organ, (2) a rhythm accompaniment generator (*GrooveTransformer*), (3) a custom Markov-based Max [11] device that generates harmonic content and note durations for the accompaniment rhythm, and (4) two accompaniment organs that receive the generated sequences.

Figure 3 gives an overview of the communication between each component of the system. First, we will consider the Max device that contains separate Markov models for pitch and note duration. These models are trained in real-time by receiving and processing all MIDI notes played on the main organ. In parallel, all MIDI notes played on the main organ are received and processed by *GrooveTransformer*. In contrast to the Markov models, which are trained on pitch and note duration, *GrooveTransformer* is conditioned on rhythmic content—specifically, the onsets and velocities of the MIDI notes. From this input, *GrooveTransformer* generates, in real-time, an accompanying 9-voice rhythmic pattern with expressive timing and velocity that would typically be mapped to a traditional drum kit. In this case, however, each onset of the generated rhythm is used to trigger the Markov model to output the next accompaniment note. The result is a multi-voice pattern that complements the main organ performance both rhythmically and harmonically. Finally, this pattern is sent from the Max device and played by the accompaniment organs.

4.2 Max Device and Controls

The Max device functions as a sub-system, giving the performer control in four main areas: (1) incoming MIDI, (2) Markov model type and behavior, (3) multi-channel output algorithms, and (4) per channel pitch transposition and note-duration modification.

Incoming MIDI is received by the device on two channels. The first channel is used to train the Markov models, with an optional pitch range filter that restricts training to notes within the specified range. The second channel serves as the rhythm source, triggering the Markov model to generate the next note for playback. Although *GrooveTransformer* was the rhythmic source for this performance, the playback of the generations can be triggered by the onsets of any external MIDI input. Lastly, a velocity threshold filter allows further control over rhythm density by excluding incoming notes below a specified velocity.

The device also includes a mono/poly mode for handling rhythm input. In mono mode, incoming notes that overlap within a set duration are ignored, making it ideal for generating melodies. In poly mode, all incoming notes trigger the Markov model, enabling chord generation.

The Markov model settings allow the performer to choose between first- and second-order models for both pitch and duration. Additionally, the performer can specify whether the next generated note should be conditioned on the most recent note from the main performance or the last note generated by the model itself.

After generation, notes are routed to designated output channels according to one of four algorithms, creating two distinct note streams:

- **Alternate:** Sends each note to Channel 1 or Channel 2 in an alternating sequence.
- **Random Channel:** Assigns each note probabilistically to either Channel 1 or Channel 2.
- **Pitch Assign:** Maps specific pitch values to a fixed channel.
- **Range Assign:** Groups pitch values into intervals based on the "Pitch Range" setting and alternates their assignment between the two channels.

Alternatively, a performer may also select no algorithm and allow every note to be output on both channels simultaneously. Lastly, the performer can apply an octave transposition (-4 to +4 octaves) and a note duration multiplier (0.25 to 16) independently to each channel.

These aforementioned settings and transformations allow the performer a high degree of control in shaping the musical output while remaining coherent with the main performance. For example, certain settings can induce slow, sustained bass notes, while others create rapid flutters of high-pitched notes or dense harmonic textures.

5 Reflections and Conclusions

The eventual performance configuration featured *Refree* as the primary performer on the main organ, while the first author of this paper acted as a secondary performer to supervise the accompaniment system and maintain a synchronized tempo with *Refree*. Designed to be supportive rather than dominant, supervision of the accompaniment system ensured that the focus remained on the primary performance. Since the organs are sonically similar to each other, overly dense or harmonically similar accompaniment notes could easily overcrowd the primary performance. To prevent this, the secondary performer adjusted the parameters of the accompaniment system in real time, controlling note density, pitch transposition, and note duration to maintain clear harmonic and rhythmic distinction between the accompaniment and main performance.

We also expected that using multi-channel outputs to send different note streams to the accompaniment organs—positioned at various levels of the venue—would create a less crowded sound and provide clearer separation from the main organ. Unfortunately, it was very difficult for the performers to hear and monitor

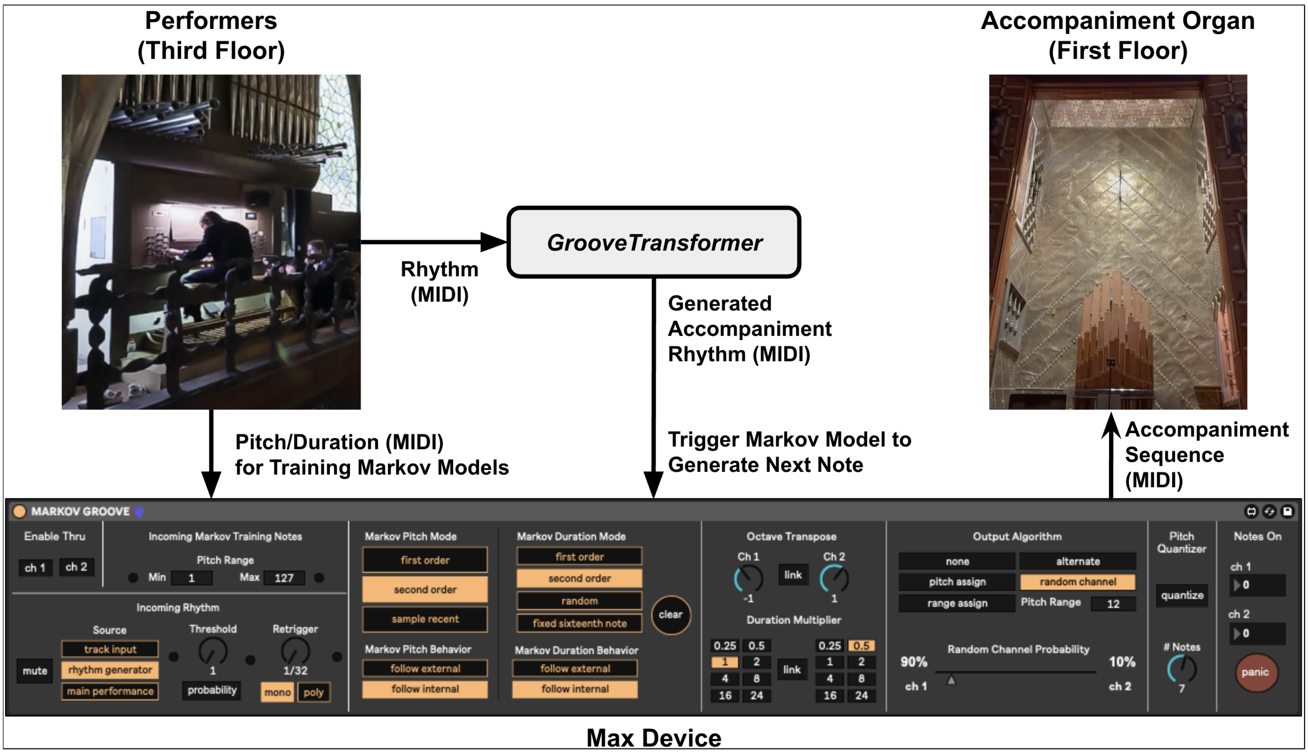

Figure 3: Full overview of accompaniment system.

the organ positioned at the first floor. Therefore, instead of utilizing the multi-channel output algorithms developed for the accompaniment system, both organs received the same stream of generated notes so that the third floor accompaniment organ could serve as a monitor for the organ located on the first floor.

Although the design of this accompaniment system was dependent on the constraints and context of this specific performance, we see this system as being very adaptable, and potentially more effective in other contexts. In an environment in which each instrument is sonically distinct and positioned to optimize room acoustics, the accompaniment system could fully leverage the multi-channel output algorithms and further enhance the spatial separation and dynamic interplay between different musical elements.

Ultimately, this work underscores the entanglement of artistic constraints and technical design decisions, illustrating how model selection is inherently shaped by the specificities of the performance context. Rather than defaulting to the most recent or computationally advanced models, we opted for a method that prioritized responsiveness and sensitivity to live interaction. In doing so, we demonstrate that in co-creative, improvisational settings, the most appropriate model is not always the most sophisticated, but rather the one best aligned with the situated demands of the performance environment.

Lastly, a collection of recordings, demos, and source code is available at

https://NIME2025OrganPerformance.github.io

6 ACKNOWLEDGMENTS

This research was funded by the Secretaría de Estado de Digitalización e Inteligencia Artificial, and the European Union-Next Generation EU, under the program Cátedras ENIA 2022. "IA y Música: Cátedra en Inteligencia Artificial y Música" (Reference: TSI-100929-2023-1). This work is also supported by IMPA project PID2023-152250OB-I00 funded by MCIU/AEI/10.13039/501100011033/FEDER, UE.

7 Ethical Standards

Our research strictly follows the ethical guidelines and standards outlined in the NIME Principles & Code of Practice on Ethical Research [8]. Throughout this project, we have taken measures to prevent any situations that might result in a conflict of interest.

In our commitment to accessibility and transparency, we have openly shared all developed tools and resources with the public.

References

[1] Jack Armitage and Victor Shepardson. 2024. Augmenting the Expressivity of the Notochord Generative MIDI Model for Arca's "The Light Comes in the Name of the Voice" Magnetic Resonator Piano Installation. In *Proceedings of the 5th Conference on AI Music Creativity.*

[2] Christodoulos Benetatos, Joseph VanderStel, and Zhiyao Duan. 2020. BachDuet: A deep learning system for human-machine counterpoint improvisation. In *Proceedings of the 20th International Conference on New Interfaces for Musical Expression.* NIME.

[3] John A. Biles. 1994. GenJam: A Genetic Algorithm for Generating Jazz Solos. In *Proceedings of the 1994 International Computer Music Conference.* Michigan Publishing.

[4] Nicholas Evans, Behzad Haki, and Sergi Jorda. 2024. GrooveTransformer: A Generative Drum Sequencer Eurorack Module. In *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME) 2024.* NIME.

[5] Behzad Haki. 2025. *Design, development, and deployment of real-time drum accompaniment systems.* Ph. D. Dissertation. Universitat Pompeu Fabra.

[6] Nan Jiang, Sheng Jin, Zhiyao Duan, and Changshui Zhang. 2020. RL-Duet: Online music accompaniment generation using deep reinforcement Learning. In *Proceedings of the The Thirty-Fourth AAAI Conference on Artificial Intelligence.* AAAI.

[7] Karl Johannsson, Thor Magnusson, and Victor Shepardson. 2023. Intimate Musical Collaboration with a Probabilistic Model. In *Proceedings of the 4th Conference on AI Music Creativity.* AIMC.

[8] Fabio Morreale, Nicolas Gold, Cécile Chevalier, and Raul Masu. 2023. NIME Principles & Code of Practice on Ethical Research.

[9] Fabian Ostermann, Igor Vatolkin, and Günter Rudolph. 2017. Evaluation Rules for Evolutionary Generation of Drum Patterns in Jazz Solos. In *Proceedings of the 6th International Conference on Computational Intelligence in Music, Sound, Art and Design (Lecture Notes in Computer Science)*.

[10] François Pachet. 2003. The Continuator: Musical Interaction With Style. *Journal of New Music Research* 32, 3 (2003), 333–341.

[11] Miller Puckette, David Zicarelli, et al. 1990. Max/msp. *Cycling* 74 (1990), 1990–2006. https://cycling74.com/products/max

[12] Alexander Scarlatos, Yusong Wu, Ian Simon, Adam Roberts, Tim Cooijmans, Natasha Jaques, Cassie Tarakajian, and Cheng-Zhi Anna Huang. 2025. ReaLJam: Real-Time Human-AI Music Jamming with Reinforcement Learning-Tuned Transformers. *CoRR* (2025).

[13] Victor Shepardson, Jack Armitage, and Thor Magnusson. 2022. Notochord: a Flexible Probabilistic Model for Embodied MIDI Performance. In *Proceedings of the 3rd Conference on AI Music Creativity*. AIMC.

[14] John Thickstun, David Leo Wright Hall, Chris Donahue, and Percy Liang. 2024. Anticipatory Music Transformer. *Trans. Mach. Learn. Res.* 2024 (2024). https://openreview.net/forum?id=EBNJ33Fcrl

[15] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is All you Need. In *Proceedings of Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems (NeurIPS 2017), Long Beach, CA, USA*.

[16] Zihao Wang, Kejun Zhang, Yuxing Wang, Chen Zhang, Qihao Liang, Pengfei Yu, Yongsheng Feng, Wenbo Liu, Yikai Wang, Yuntai Bao, and Yiheng Yang. 2022. SongDriver: Real-time Music Accompaniment Generation without Logical Latency nor Exposure Bias. In *Proceedings of MM '22: The 30th ACM International Conference on Multimedia*. ACM.

[17] Yusong Wu, Tim Cooijmans, Kyle Kastner, Adam Roberts, Ian Simon, Alexander Scarlatos, Chris Donahue, Cassie Tarakajian, Shayegan Omidshafiei, Aaron C. Courville, Pablo Samuel Castro, Natasha Jaques, and Cheng-Zhi Anna Huang. 2024. Adaptive accompaniment with ReaLchords. In *Proceedings of the Forty-first International Conference on Machine Learning*. ICML.