Real-time Musical Collaboration with a Probabilistic Model

Karl Johannsson
Department of Computer Science
Reykjavik University
Reykjavik, Iceland
karljohann@gmail.com

Victor Shepardson
Intelligent Instruments Lab
University of Iceland
Reykjavik, Iceland
victorshepardson@hi.is

Enrique Hurtado
Arte y Tecnologıa, Bellas Artes
University of the Basque Country
Leioa, Spain
enrique.hurtado@ehu.eus

Thor Magnusson
Intelligent Instruments Lab
University of Iceland
Reykjavik, Iceland
thormagnusson@hi.is

Hannes Högni Vilhjálmsson
Department of Computer Science
Reykjavik University
Reykjavik, Iceland
hannes@ru.is

***Abstract*—This paper describes the development of a system to play an ancient Basque collaborative percussion instrument (the txalaparta) with a human performer. A deep probabilistic model was trained on two large datasets, and transfer learning used to train the model on high-quality data abstracted from txalaparta performances. The aim is to create a system that effectively responds to the human player in a performance environment, where the model's response must feel instantaneous, organic, and sensible in response to the user's input. The paper outlines the design and implementation of the system, including data collection, model training, and performance visualization. The study contributes to the field of interactive music systems and demonstrates the potential of deep learning in creating intelligent musical systems that can collaborate with human performers.**

***Index Terms*—Music Generation, Deep Learning, Collaborative Instruments, Txalaparta**

I. Introduction

In recent years, the field of music generation has grown significantly, driven by the development of advanced deep learning models capable of learning complex patterns and structures in large music datasets, and the increasing availability of those datasets.

This evolution has brought with it tools with the ability to create convincingly humanesque music.

A myriad of systems have been developed to generate music, some to make it sound human (e.g., [5], [7]), others to discover new sounds (e.g., [2]), while still others create (or assist in creating) scores for human musicians to perform (e.g., [1]). These systems more commonly perform in an offline setting, where the whole process takes place before the work is heard. Systems for real-time (or ”live”) collaborative musical performance, however, are less prevalent.

For such a system to work, it must listen to the collaborating musicians and continuously calculate an appropriate response based on the processed input, much like a human player. That is a resource-heavy task, and the latency must be low for a system to respond in real-time.

Fig. 1. Members of the Intelligent Instruments Lab play a txalaparta consisting of four planks. The performers hold a baton in each hand, hitting the planks a varying amount of times, splitting each bar in a consecutive manner. Source: [9]

This study demonstrates an attempt to blend deep learning and traditional musical collaboration, shedding light on the challenges and potential of AI in interactive music systems.

For the study, a system was developed to play along with a human player in a convincing manner, using a txalaparta, a collaborative Basque percussion instrument, and Notochord [8], a probabilistic neural network model. Data was gathered from performances of expert txalaparta players, that was used to train the neural network.

The paper describes the development of the system, gathering the necessary training data, analyzing the system's output, attempts at implementing the system for a real-world setting, and finally, the difficulties the system must overcome to be a useful collaborator.

II. Background

The txalaparta consists of one or more wooden planks hit with special batons (see Fig. 1). It is played simultaneously by two players, improvising in a call-and-response fashion. Playing the txalaparta is an intimate process; it is a dynamic conversation between two players, where one feeds upon the expression of the other and integrates it into their own expression, with the shared purpose of creating music. As uncomplicated as it may seem to hit wooden planks with batons, the instrument's beauty lies in the collaboration of the two players, playing together expressively in harmonious coordination. The players must work together to create the music of the txalaparta. In fact, it is one of few instruments that can not be played by a single performer.

Instruments are almost exclusively designed for individual performers to create music solo or in ensembles. Instruments designed explicitly for multiplayer (or collaborative) performances are rare. Collaborative instruments require multiple players to work together to create a cohesive musical experience. This is a different level of interdependence from playing a solo instrument in an ensemble, as the performance of the instrument itself stands or falls with the collaboration of more than one player. Out of necessity, collaborative instruments will provide and require a more immersive musical experience where the players can influence and affect each other on a much more fragile level.

Of these collaborative instruments, some are collaborative by necessity, as they might physically require more than one person to perform the instrument. The txalaparta, however, is a collaborative instrument by tradition only. Its interface is a shared one that requires two players to create a meaningful experience.

By design, there is more intimacy in performing a collaborative instrument than playing individual instruments in an ensemble. The goal is the same, to create music, but some degree of control is always entrusted to the other players.

The txalaparta tradition is, in its nature, an intimate creative experience shared by two players, with all the nuances and subtleties of human communication, and, thus, an inherently human one. Musical rhythms generated by machines are, on the other hand, non-human (though not necessarily lacking human qualities).

In addition, with its traditionally fluid timing, the txalaparta does not lend itself well to a grid. It is played with expression, and the tempo is a matter of negotiation between the two players. This is an important aspect of the instrument and a challenge for the non-human player.

The txalaparta has been the subject of digitization before. Hurtado developed the digital txalaparta [3], an interactive system that formalizes the practices of the txalaparta and uses a Markov model to predict the next hit pattern. There has, however, been no previous research on using deep learning techniques to enable a computer to play the txalaparta collaboratively with a human player.

The use of deep learning models for rhythm generation has several advantages over more traditional approaches. Deep learning models have the potential to model richer distributions and incorporate more context. This gives them more capacity to generate novel and unexpected rhythms, which can add to the depth of the performance, as well as adapt and respond to the input of other musicians with more flexibility, enabling more immersive collaborative music creation, which can allow the performer to explore new creative possibilities and new styles of playing.

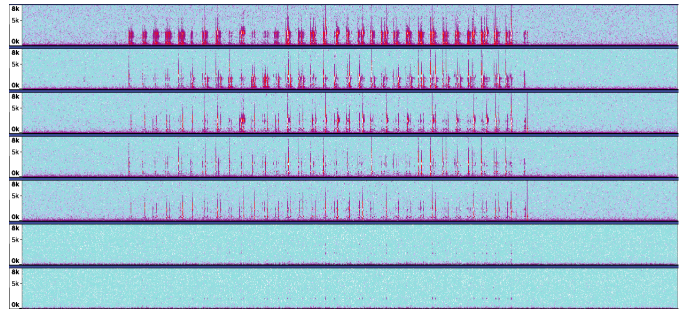

Fig. 2. Spectrogram of recording. Top is audio channel, next four batons, and bottom two are planks.

III. Methodology

This section describes the project's methodology, including deep learning model, data collection, and system implementation.

A. The Probabilistic Model

Notochord is "a deep probabilistic model for sequences of structured events" [8, p 1] recently developed by Shepardson et al. [8] at the IIL. Intended for performance settings, Notochord has a response latency of under 10 ms, making it appealing for collaborative performances. Especially for the txalaparta, where play can become lively.

Notochord expects a MIDI-like format, a single event contains an instrument, a pitch, velocity (*amplitude*), and clock (*timedelta*; elapsed time). The output of the model is a single event, a representation used to synthesize a single txalaparta hit.

B. The Data

To construct a convincing txalaparta performance, the following four attributes were recorded: (1) which baton performed on (2) which plank with (3) what amplitude, and (4) what relevant time.

The temporal dependencies for the txalaparta are minimal because the system continuously responds to the input of the human player, who steers the structure of the performance. The harmonies and melodies are reduced to three notes (planks), which are not ordered in any obvious hierarchy. Finally, the instruments (batons) are consistently four, in groups of two, with the groups further interdependent within themselves.

Notochord comes pre-trained on the Lakh MIDI dataset (LMD) [6]. To maintain adaptability, Notochord uses over 100,000 MIDI files from LMD, containing a diverse range of songs of assorted quality.

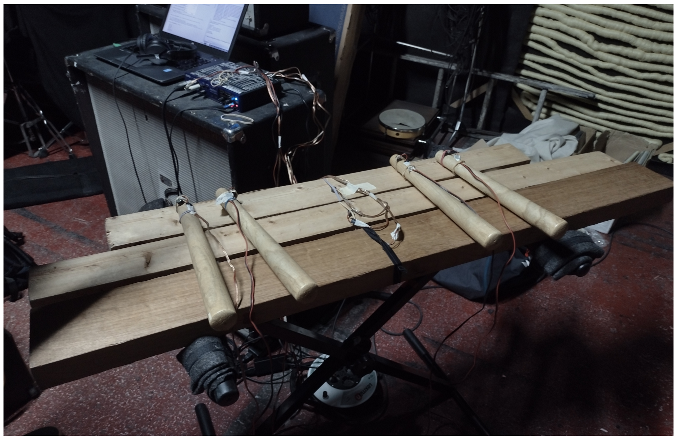

Fig. 3. Recording setup in Bilbao.

Txalaparta data was collected by reading digitized audio representations of events in time – in this case, musical rhythm from the txalaparta – into a specifically developed system that processes and records the representations.

Onsets from sensors (contact microphones) on the batons were detected and registered as elapsed time from the first event (see Fig. 2). The salient amplitude of plank sensors (also contact microphones) at the time of onset was registered as the hit plank and velocity was registered as a normalized value of the baton amplitude.

Recordings were performed by txalaparta performers in Bilbao (see Fig. 3). Two pairs of players recorded a total of five performance sessions, resulting in approximately four hours of recordings, ranging from 6:26 to 45:14 minutes each. The high-quality txalaparta data was then used to train the model

C. Training

The following settings and hyperparameters were ultimately used for training:

- *Optimizer*: AdamW
- *Batch size*: 32
- *Batch length*: 2048
- *Hidden layers*: 1024
- *Embedding size*: 512
- *RNN layers*: 2
- *MLP layers*: 2
- *Dropout*: p = 0.1 (*for MLP layers only*)
- *Learning rate*: 1e-4

A single training run consisted of 1200 epochs.

A dataset was created from approximately nine hours of the high-quality txalaparta recordings (see section III-B). Because of the scarcity of data, the train-validation split for the txalaparta datasets was set to .1 (default is .03). A random sampler was used with replacement in validation, so each sampled unit is placed back to be eligible for sampling again, where data augmentation will make it different each time. The batch length was also set as high as the smallest data file would allow. The batch size starts at 32 but increases to 64 after 10 billion events and again to 128 at 15 billion. In addition, the model uses a gradient clipping (L2 norm) set to a maximum of 1.0.

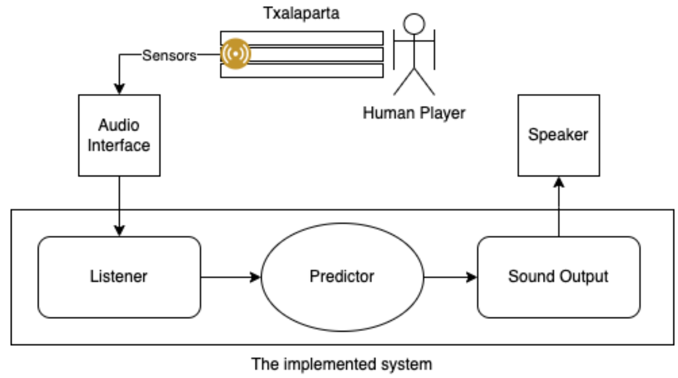

Fig. 4. Diagram of system. Sensors on batons and planks send data to the *listener* through an audio interface. The listener continuously detects hits in the received data and sends a message to the *predictor* on each hit. The predictor uses this information to predict the next hit and sends that information to the *sound output*, which transforms the information into sound through a speaker.

D. System Implementation

The final system, a combination of three modular components (see Fig. 4), is described as follows:

(1) A **listener** that processes and analyzes sound input (detects onsets) from the txalaparta and sends representations of the baton, plank, timing, and amplitude as MIDI messages to a designated bus for the predictor.

(2) The **predictor** is a model that learns patterns in the representations. It also has an input-output interface that interprets incoming data and projects the model's response. Notochord comes equipped with an "improviser" Python script. In short, an interface for the querying logic of the Notochord model, with a MIDI server that relays each event to Notochord and a client that sends messages to a user-specified bus. This script was adapted to account for the limited pitches and instruments in the txalaparta. Additional rules specific to the txalaparta were programmed to constrain the probabilistic model. For example, it could only predict up to X sequential events for each player before switching to the other, in line with the turn-taking style of txalaparta playing. Probabilistic steering controls of Notochord allowed sampling from, for example, the lower 80% of probability mass when predicting delta-time, to encourage the subtly accelerating quality of many txalaparta performances

(3) Finally, a **sound output** component accepts symbolic data and transforms it into sound using samples of the instrument or digital filters. Since Notochord sends MIDI messages to any user-specified bus and channel, a DAW was used to listen for messages on those buses and channels and play txalaparta samples.

The second component is the main focus of the study, while the first and third need to be available for it to be of any use.

1) Visualization: An integral part of collaborative music improvisation is communication and feedback. The experience

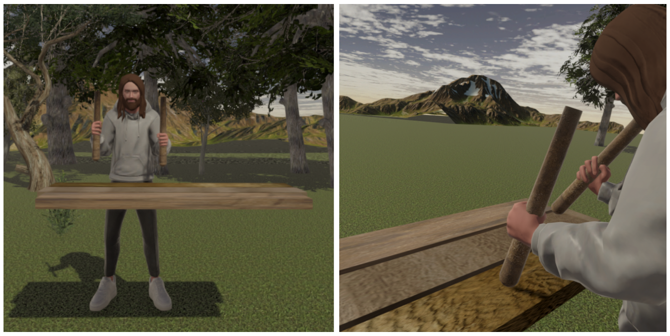

Fig. 5. The virtual txalaparta player from front and side.

of music, as well as the experience of musical collaboration, is more than simply the sounds perceived [4] and it can be assumed that sensing the movements and gestures of a collaborating performer would be beneficial to the understanding of the current musical situation.

That is especially true for the txalaparta, where the players effectively play as one person. In relation to the development of the digital txalaparta, Hurtado et al. [3, p 137] found that "visual feedback [...] proved crucial in txalaparta performance" and was a large part of player interaction.

A virtual txalaparta player was implemented in Unity (see Fig. 5). Due to the overall system computing requirements, the lag of response, especially when the playing was lively, was sometimes confusing and development of the avatar was postponed.

IV. Conclusion

This paper described an implementation of a system to perform collaboratively with a human player. The study successfully demonstrates the *potential* of deep learning models to generate rhythmic patterns that complement and interact with human musicians in real-time.

Music is not a problem to solve. It does not have rules that suggest quantifiable gain or explicit reward. Music is an open-ended expression of human emotion and has no intrinsic value other than its effect on human emotion. Though the system is a work in progress, it does seem entirely possible, with more research and more data, that a properly tuned and well-fed model could convincingly play along with a human player in an enjoyable musical exchange.

AI music generation is at a stage in its development where it is beginning to become useful beyond experiments. This study contributes to the broader field of interactive music systems by shedding light on the challenges and opportunities in AI-driven collaborative music. It is the hope of the authors that the intelligent txalaparta system will be used for people's enjoyment, as well as used to study the experience of intimately collaborating with a machine. Entrusting, even a part of, one's creative expression with a machine will most certainly become commonplace with the growing usefulness of AI.

Acknowledgment

The Intelligent Instruments project (INTENT) is funded by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant agreement No. 101001848).

References

[1] G. Assayag. Computer assisted composition today. In *1st symposium on music and computers, Corfu*, 1998.

[2] J. Engel, C. Resnick, A. Roberts, S. Dieleman, D. Eck, K. Simonyan, and M. Norouzi. Neural audio synthesis of musical notes with wavenet autoencoders, 2017.

[3] E. Hurtado, T. Magnusson, and J. Rekalde. Digitizing the txalaparta: Computer-based study of a traditional practice. *Comput. Music J.*, 43(2–3):125–141, jun 2020.

[4] A. Jensenius. *Methods for Studying Music-Related Body Motion*, pages 805–818. Springer, 01 2018.

[5] F. Liang. Bachbot: Automatic composition in the style of bach chorales. *University of Cambridge*, 8:19–48, 2016.

[6] C. Raffel. The lakh MIDI dataset v0.1, 2016.

[7] Y. Ren, J. He, X. Tan, T. Qin, Z. Zhao, and T.-Y. Liu. Popmag: Pop music accompaniment generation, 2020.

[8] V. Shepardson, J. Armitage, and T. Magnusson. Notochord: a flexible probabilistic model for real-time MIDI performance. In *Proceedings of the 3rd Conference on AI Music Creativity. The 3rd Conference on AI Music Creativity (AIMC 2022)*, 2022.

[9] E. Thorvalds. Personal communication. Available online: https://iil.is/news/txalaparta_dynjandi, March 2023. Accessed: May 7, 2024.