AIMC 2023

Intimate Musical Collaboration with a Probabilistic Model

Karl Johannsson Thor Magnusson Victor Shepardson

Published on: Aug 29, 2023

URL: https://aimc2023.pubpub.org/pub/v44bkgoy

License: Creative Commons Attribution 4.0 International License (CC-BY 4.0)

ABSTRACT

Recent advancements in deep learning have created many opportunities in the field of music. This research explores training a deep learning model to play a percussion instrument collaboratively with a human player. The aim is to create a system that convincingly responds to the human player's performance in real-time. Music generation is more commonly handled offline than in a performance environment, where the model's response must feel instantaneous, organic, and sensible in response to the user's input. A probabilistic model was trained on two large datasets, and transfer learning used to train the model on high quality data abstracted from a collaborative percussion instrument, the txalaparta. The paper outlines the design and implementation of the system, including data collection, model training, and performance visualization. The research 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.

Introduction

With the recent flood of research in specialized artificial intelligence, tools have surfaced with the ability to create convincingly humanesque music of all genres. A myriad of systems have been developed to generate music, some to make it sound human (e.g., [1][2]), others to discover new sounds (e.g., [3]), while still others create (or assist in creating) scores for human musicians to perform (e.g., [4]). 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.

Collaborating with an AI in an intimate manner, where the human player and the AI system are required to work together for a successful performance, has rarely (if at all) been attempted in a real-world setting. AI is frequently applied as decoration for human-produced music, or in artistic endeavors, where technical accuracy is less important than the outcome of the experiment itself.

For this research, a system was developed to play along with a human player, using a collaborative Basque percussion instrument, named txalaparta, and a probabilistic neural network model, named Notochord [5].

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.

Background

The txalaparta consists of one or more wooden planks hit with special batons. 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. Symons [6] notes that for an instrument to be considered collaborative (or "entangled"), either the instruments interface requires players to cooperate to create meaningful music or the input and output of individual controllers are coupled in such a way that sounds created by the actions of one performer directly affect the input and output of another.

Some collaborative instruments 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. There is, however, no actual necessity for two players. A single person could easily play the instrument independently, but that is not txalaparta in the same way that hitting a violin's back with its bow is, traditionally, at least, not playing the violin.

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.

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 since there is no requirement to adhere to a grid – as would be the case for most other instruments (e.g., real-time drum accompaniment [7][8]) – things such as adjustments to events on a grid are replaced by the continuous fluidity of time.

The txalaparta has been the subject of digitization before. Hurtado developed the digital txalaparta [9], 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 can generate novel and unexpected rhythms, which can add to the depth of the performance. They can also 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.

The Probabilistic Model

Notochord is "a deep probabilistic model for sequences of structured events" [5] recently developed by Shepardson et al. [5] at the IIL. Intended for performance settings, it offers, among other things, steerability, real-time harmonizing, and machine improvisation. Notochord has a response latency of under 10 ms, which makes it appealing for collaborative performances, especially for the txalaparta, where play can become lively.

Notochord expects a MIDI-like format,[1] a single event contains an instrument (*baton*; anonymous instrument (drums), integer $265 - 268$), a pitch (*plank*; tom-drum pitches, integer $41, 43, 45$), velocity (*amplitude*; constrained range, integer $60 - 127$), and clock (*timedelta*; elapsed time, float). Clock is converted to time difference in data processing. The output of the model is a single event with an array of possible predictions for each element coupled with the probability of each prediction.

The Data

The following four attributes were deemed sufficient to construct a convincing txalaparta performance and 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 responds to the dynamics 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) [10] for up to 50 billion events, which is available in the form of a model checkpoint.[2] To maintain adaptability, Notochord uses a diverse range of songs of assorted quality contained in over 100,000 MIDI files from LMD.

As well as LMD, another dataset also used for pre-training, due to its inherent rhythmic focus, was the Expanded Groove MIDI Dataset (E-GMD) [11]. The dataset is a collection of MIDI drum patterns that aims to provide a comprehensive and diverse dataset for research and development in the field of music generation. It

consists of 444 hours worth of human drumming in MIDI and audio format, covering a wide range of musical styles, genres, and tempos.

Notochord predicts MIDI note-off events instead of note lengths. That is, it predicts two separate events, the note-on, and the note-off, instead of predicting a single event with a time length. The txalaparta is an idiophone, and all notes are of similar length. It was therefore decided to ignore note-offs in the data. The model, however, needed to be adjusted to account for single events. That was done by performing pre-training on the E-GMD dataset with note-offs removed from the data.

Txalaparta Data

As part of ongoing research, a txalaparta was built at the Intelligent Instruments Lab (IIL) at the Iceland University of the Arts, where the research is hosted.[3] Sensors were placed on the instrument to abstract the human playing. The goal was to "teach" a machine to play the instrument collaboratively with a human player by analyzing their playing pattern and responding to it.

This was done 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.

Recordings were performed by collaborators in Bilbao. 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.

Onsets from sensors (contact microphones) on the batons were detected and registered as elapsed time from the start of the recording session. 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.

Notochord expects data to be in a MIDI-like format, so for training, baton identities were mapped to the indices of four anonymous drums, $265 - 268$. Plank identities were similarly mapped to the numbers of five MIDI pitches, 41, 43, 45, 47, and 48. These numbers were chosen because they represent tom drums in the MIDI drum program.

The amplitudes measured from the baton at onset time were normalized and rescaled to integers in the range of $60 - 127$, corresponding to the MIDI velocity range of $0 - 127$ (the txalaparta has a relatively small dynamic range).

System Implementation

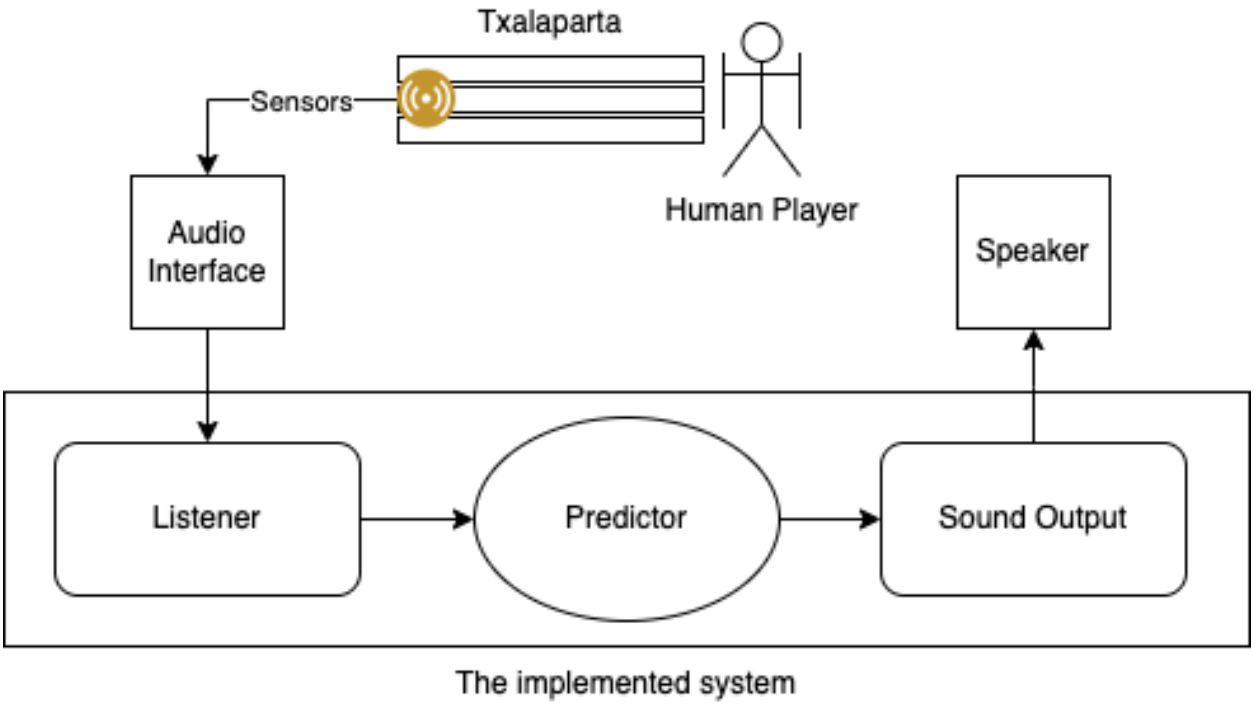


Figure 1. Diagram of system. Sensors on batons and planks send data to the *listener* through an audio interface. The listener continuously detects hits in the received data and sends a message to the *predictor* on each hit. The predictor uses this information to predict the next hit and sends that information to the *sound output*, which transforms the information into sound through a speaker.

This section describes the development of a system that plays the txalaparta collaboratively with a human player. The collaborative framework for human-machine interaction includes an interface for the system as a whole, the models specifically, and the output, as well as representations of it. The final system, a combination of three modular components (see Figure 1), 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. "Steering" was also applied for some

additional parameters, such as the number of responses before the predictor should send another, truncating time to the density of predictions, and stopping when the user stops.

(3) Finally, a **sound output** component accepts symbolic data and transforms it into sound using samples of the instrument. 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 research, while the first and third need to be available for it to be of any use. In addition to the components required for the system to work, a **visualizer** component was implemented to visually represent the system's output.

Visualization

An integral part of collaborative music improvisation is communication and feedback. The experience of music, as well as the experience of musical collaboration, is more than simply the sounds perceived [12] 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 continually communicate throughout the entire performance. In relation to the development of the digital txalaparta, Hurtado et al. [9] found that "visual feedback [...] proved crucial in txalaparta performance" and was a large part of player interaction.

The sonic feedback of a baton hitting a plank is a good indicator of what has happened, but when the player perceives that indicator, the event has already occurred, giving the human player less time to prepare for their next hit.

A virtual txalaparta player was implemented to provide the user with a graphical aid that visually represents the hit predictions. The virtual player was implemented with Unity and consists of a player (avatar) holding a baton in each hand, a txalaparta, and a (Basque-esque) mountain scene (see Figure 2).

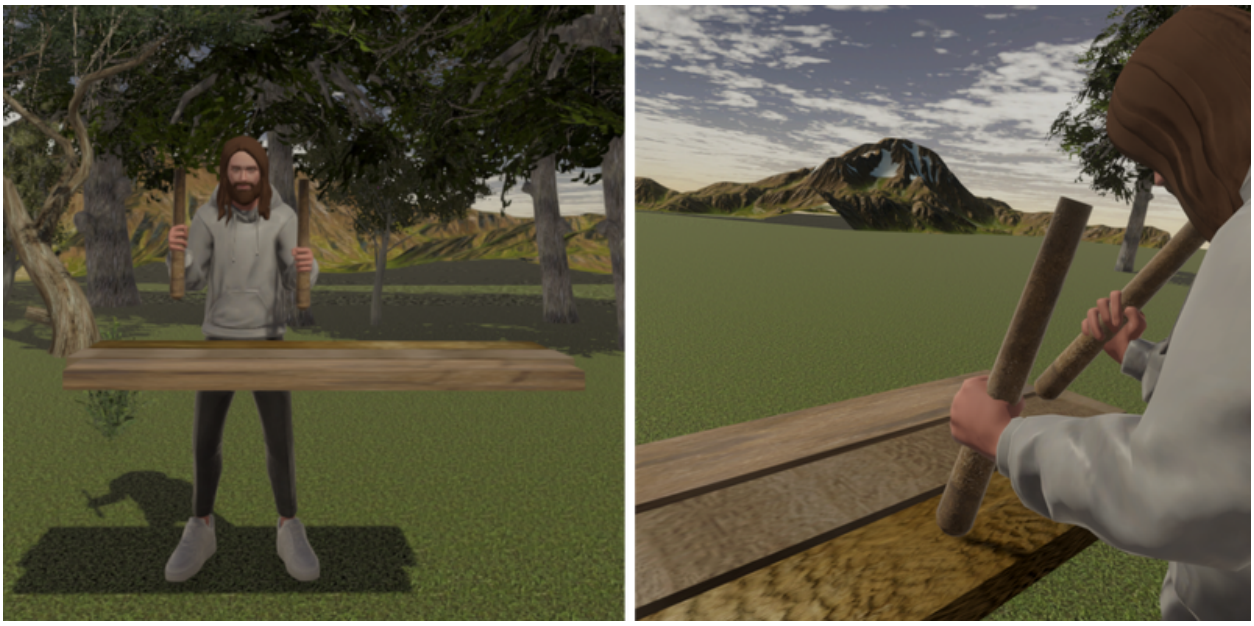


Figure 2. The virtual txalaparta player from front and side.

Notochord sends its predictions to a frontend logic which calculates the speed of the hit according to how far in advance it is. If the human playing is fast, then the predictions are closer in time than if the playing were slow, so the avatar's playing also appears fast.

Conclusion

This paper described an implementation of a system to perform collaboratively with a human player. The research successfully demonstrates the *potential* of deep learning models to generate rhythmic patterns that complement and interact with human musicians in real-time. But music is not a problem to solve. It is an open-ended expression of human emotion and has no intrinsic value other than its effect on human emotion. Even though the system does not show many signs of intelligence yet, it seems entirely possible, with more research and 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. There is no doubt that humans and machines will collaborate more and more intimately in the coming years. Whether in live performances or offline music generation, the possibilities are endless and just beginning to be explored.

Acknowledgments and Disclosure of Funding

Thanks to Dr Enrike Hurtado for his collaboration, involving the recording of performance data, system design, user testing and general guidance.

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).

Footnotes

1. MIDI with the addition of anonymous instruments. Also, MIDI uses tracks and program change events, whereas Notochord sets instrument ID on every event ↩

2. See https://github.com/Intelligent-Instruments-Lab/iil-python-tools/tree/master/notochord ↩

3. https://iil.is/research/txalaparta ↩

References

1. Liang, F. (2016). Bachbot: Automatic composition in the style of bach chorales. *University of Cambridge*, *8*, 19–48. ↩

2. Ren, Y., He, J., Tan, X., Qin, T., Zhao, Z., & Liu, T.-Y. (2020). *PopMAG: Pop music accompaniment generation*. https://arxiv.org/abs/2008.07703. Retrieved from https://arxiv.org/abs/2008.07703 ↩

3. Engel, J., Resnick, C., Roberts, A., Dieleman, S., Eck, D., Simonyan, K., & Norouzi, M. (2017). *Neural audio synthesis of musical notes with WaveNet autoencoders*. https://arxiv.org/abs/1704.01279. Retrieved from https://arxiv.org/abs/1704.01279 ↩

4. Assayag, G. (1998). Computer assisted composition today. *1st Symposium on Music and Computers, Corfu*. https://www.music.mcgill.ca/~gary/306/week11/Assayag_ComputerAssisted_1998.pdf. Retrieved from https://www.music.mcgill.ca/~gary/306/week11/Assayag_ComputerAssisted_1998.pdf ↩

5. Shepardson, V., Armitage, J., & Magnusson, T. (2022). Notochord: A flexible probabilistic model for real-time MIDI performance. *Proceedings of the 3rd Conference on AI Music Creativity. The 3rd Conference on AI Music Creativity (AIMC 2022)*. ↩

6. Symons, S. (2022). Entangled instruments: Exploring participatory digital music making environments. *International Conference on New Interfaces for Musical Expression*. Retrieved from https://nime.pubpub.org/pub/fo51e2v0 ↩

7. Haki, B., Nieto, M., Pelinski, T., & Jordà, S. (2022). Real-Time Drum Accompaniment Using Transformer Architecture. *Proceedings of the 3rd Conference on AI Music Creativity*. AIMC. ↩

8. Gillick, J., Roberts, A., Engel, J. H., Eck, D., & Bamman, D. (2019). Learning to groove with inverse sequence transformations. *CoRR, abs/1905.06118*. Retrieved from http://arxiv.org/abs/1905.06118 ↩

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

10. Raffel, C. (2016). *The lakh MIDI dataset v0.1*. https://colinraffel.com/projects/lmd/. Retrieved from https://colinraffel.com/projects/lmd/ ↩

11. Callender, L., Hawthorne, C., & Engel, J. (2020). *Improving perceptual quality of drum transcription with the expanded groove MIDI dataset*. https://arxiv.org/abs/2004.00188. Retrieved from https://arxiv.org/abs/2004.00188 ↩

12. Jensenius, A. (2018). *Methods for studying music-related body motion*. Springer. ↩