Spirio Sessions: Experiments in Human-Machine Improvisation with a Digital Player Piano

Sebastian Trump[1,2], Ismael Agchar[2], Ilja Baumann[2], Franziska Braun[2], Korbinian Riedhammer[2], Lea Siemandel[2], and Martin Ullrich[1]

[1] Nuremberg University of Music
[2] Technische Hochschule Nürnberg
sebastian.trump@hfm-nuernberg.de

Abstract. This paper presents an ongoing interdisciplinary research project that deals with free improvisation and human-machine interaction, involving a digital player piano and other musical instruments. Various technical concepts are developed by student participants in the project and continuously evaluated in artistic performances. Our goal is to explore methods for co-creative collaborations with artificial intelligences embodied in the player piano, enabling it to act as an equal improvisation partner for human musicians.

Keywords: Human-Machine Improvisation, Co-creativity, Player Piano

1 Introduction

Many attempts have been made in the last decades to develop interfaces which allow computers to become actors in musical performances, ranging from direct control as in digital musical instruments (cf. e.g. Miranda, Wanderley, & Kirk, 2006) to virtual autonomous players with a higher degree of creative agency (Gifford et al., 2018).

The *Sprio Sessions* project aims to explore concepts of free improvisation among humans and machines in different research directions by prototype development, different combinations of software modules, and artistic evaluation. To give the computer-generated musical material in this human-machine collaboration scenario a physical presence comparable to that of other traditional musical instruments the machine player here acts in an embodied form of a digital player piano[3] (cf. similar approaches e.g. in Brown, 2018, or the marimba-playing robot improvisor Shimon by Hoffman & Weinberg, 2010) instead of using loudspeakers for the actual sonic realization. Within this framing of a duo setting consisting of the player piano controlled by an AI system and a human musician, we are aiming at the exploration of various computational approaches for the interactive generation of musical material.

[3] We use a Steinway & Sons digitally-enabled Spirio-R grand piano, which gave the project its initial working title that has been retained since then.

The theoretical-epistemic foundations of our project refer to the concept of "musical cyborgs", which assembles the diverse configurations of human-machine co-creativity in the context of musical performance within the framework of critical posthumanities (Braidotti, 2016). From this perspective, the setting examined here can be described as one possible variation of more-than-human sonic collaborations (Ullrich & Trump, in press). Therefore, the objective is not to simulate human pianism—even if distinct building blocks and processes involving machine learning seem to point in this direction—but to establish a relational aesthetics that encourages genuine machine artifacts and at the same time minimises human preselection.

This paper will briefly outline our research design and general methodological approach, then go into more detail on each of the current research directions, and give an outlook towards future work.

2 Research Design

The *Spirio Sessions* project is designed around questions of interactivity in free musical improvisation with computational systems following Rowe's (1993) player paradigm and its constitutive criterion of creative agency (Bown & McCormack, 2011). The improvisational setting around the player piano forms a conceptual framework within which a wide-ranging spectrum of technical approaches—music information retrieval (MIR), rule-based AI, statistical modeling, and neural networks—is to be prototypically explored. The interdisciplinary research group involved here brings together scholars from interdisciplinary music research and computer science, as well as graduate students from computer science, media computer science, jazz performance and music pedagogy.

2.1 Methods

All newly developed software elements in this project are modular in design and intended to enable the most flexible combinations between each other. A Max/MSP patch serves as a hub for the individual modules, which are integrated via virtual MIDI ports and Open Sound Control (OSC). The project uses an experimental approach and therefore asks for artistic potentials and creative capacities of different technical concepts within the given setting rather than looking for an ideal solution. Many of the AI techniques studied so far have already been used in other computational music generation projects, but often not in interactive scenarios. Hence, the artistic research (Klein, 2018), carried out by the participating music students, is a crucial methodological component for the evaluation of modified software prototypes. Such elements of subjective assessment commonly used in research on computational systems for music improvisation (Gifford et al., 2018, 25) are applied here in systematically recorded sessions[4] after each major development step and in defined parameter configurations.

[4] Demo videos of performances using software from the following research directions are available in Trump (2021).

2.2 Research Directions

First prototype: Markov Chains, Adaptive Attention, and Arpeggios During the preparatory phase of the research project, a first prototype operating with simple Markov chains as elementary building blocks of machine learning was created. This early experiment was implemented entirely in Max/MSP and used the extension *ml.lib* (Smith & Garnett, 2012) to embed 2nd order Markov chains for pitch progressions modeling. A continuous measure of note density influences the degree of attention to new input material from a pitch-tracked audio signal and the addition of randomly selected arpeggios of symmetrical interval structures (cf. Fig. 1).

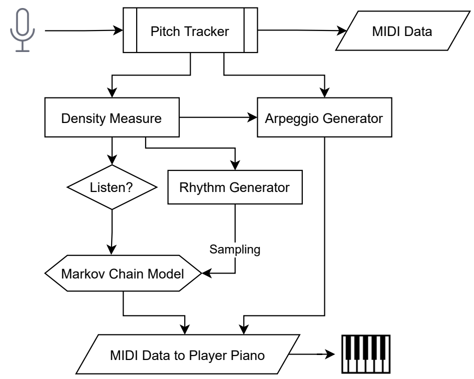

Fig. 1: Process flow of first prototype Max/MSP patch

HMM-based Improvisation Building on the statistical approach of the first prototype, the focus of this subproject is on the investigation of automatically generated musical improvisation using Hidden Markov Models (HMM). Extending regular Markov Models the HMM topology is defined by the number of hidden states, the arrangement of the state transitions, and a set of possible observable emissions (Jurafsky & Martin, 2020). For musical improvisation, states and emissions can be assigned different meanings (Marom, 1997), such as notes, note durations, velocity, intervals, or chords (Simon, Morris, & Basu, 2008). In HMM training, a distinction can be made whether training is event-triggered, e.g., after a note is played, or time-triggered, e.g., after each quarter beat. Investigated parameters affecting the training process itself are the window size, the transition and emission probabilities, as well as the weighting for retraining. The probabilities can be pre-trained on MIDI data or initialized with specific distributions such as Gaussian, Discrete, or randomly. It can make a difference whether the training is performed with a flat start or if retraining algorithms like Viterbi and EM are applied (Jurafsky & Martin, 2020). For music generation, it is possible to sample from the HMM or to make a prediction based on an observed sequence. The number of generated samples and the sample rate will affect the resulting melody.

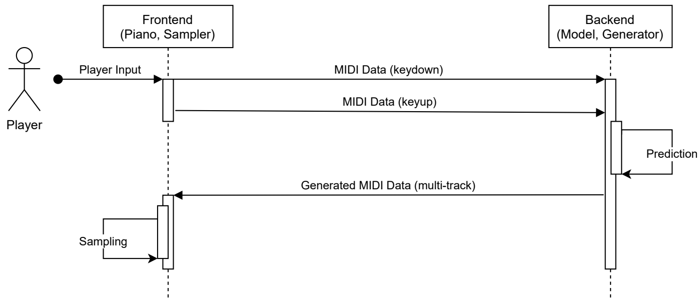

Fig. 2: Client-server architecture for neural network experiments

Neural Networks for Interactive Multitrack Music Generation As part of a master's thesis, neural network approaches for the generation of an interactive multi-instrument accompaniment with the lowest possible latency have been tested. The system is implemented using a client-server architecture (cf. Fig. 2) with a web interface for symbolic MIDI input. For the generation, existing models, e.g. from the Google Magenta project, were examined and adapted where necessary. In particular, generative deep learning models such as Variational Autoencoders (VAE) (Roberts, Engel, Raffel, Hawthorne, & Eck, 2019), Generative Adversarial Networks (GANs) (Dong, Hsiao, Yang, & Yang, 2017) and Transformers (Huang et al., 2018) were considered. The MusicVAE model has turned out to be the most suitable for this purpose and was adopted as the basis for the implementation. The server provides a REST API so that other front-end systems can be connected. After the server has generated the accompaniment, the data is sent to the front-end and rendered time-synchronously into MIDI data.

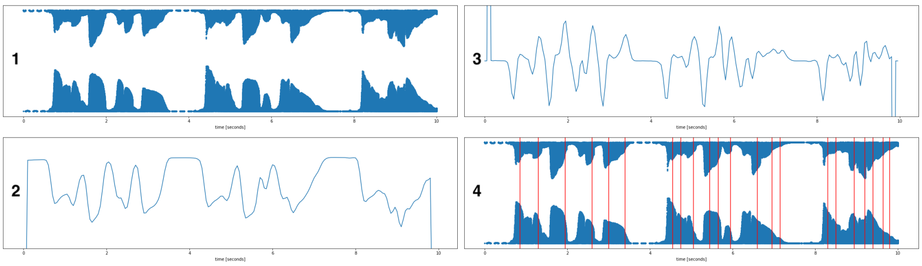

Fig. 4: Steps of the rhythm detector following Bello et al. (2005): **(1)** Wave-form signal, **(2)** Smoothed by convoluting it with a Hann-window $w(n) = sin^2 \frac{\pi n}{N}$ with frame index n and window length N, **(3)** Detection signal with characteristics of rhythmically important (energetically high, i.e. a dominant peak in the amplitude envelop) beats: calculated as the discrete derivative $s'(n) = s(n) - s(n-1)$ of (2), **(4)** Maxima in (3) exceeding a certain threshold ϵ marked as found beats.

Rhythm Detector In order to improve rhythmic synchronization and entrainment (Clayton, Sager, & Will, 2005), a custom implementation for a beat detector (cf. Fig. 4) was developed as an additional subproject. We found the window length N of the Hann-window and the peak threshold ϵ most crucial to achieving good detection results, heavily depending on the input instrument. To enable for realtime processing here, the input audio signal must be processed in suitable chunks and analyzed in parallel on multiple threads. The detected beats are then sent out as OSC messages.

Sequence-to-Sequence Neural Networks The goal of this upcoming subproject is to use sequence-to-sequence neural networks (S2SNN) to model one part of the interacting musical duo. As a typical example throughout epochs, it will focus on a duo of a melody (woodwind) and accompanying instrument (piano), comparing the very structured baroque *basso continuo* setup with free improvisation. The central question is: Provided symbolic input/output (eg. MIDI), can a S2SNN generate an accompaniment for a melody and vice versa? Related questions are: how much context is needed, how can the model anticipate the other player, how to achieve rhythmic synchronization? The models will be trained on prerecorded duo performances, potentially leveraging the full context. The test scenario however will be stream-based, i.e. the model may store history but can't look ahead. This work will focus on symbolic data (eg. MIDI), which can easily be discretized (input) or synthesized (output) via the player piano.

3 Future Work

We can already see that the creative potential for our systems lies less in isolated software elements than in their intelligent combination and the choice of appropriate parameters. In this sense, each new Spirio Sessions subproject expands the field of possibilities in several new directions. One followup research during the next phase of the project will further dive into the idea of rhythm-like, rule-driven music generation techniques. For this, probabilistic and transformational grammars will be explored in a linguistic appraoch to the process of music generation (Keller & Morrison, 2007 and Putman & Keller, 2015). Another promising approach will address the specifics of the piano pedal and examine modeling techniques for this purpose. In addition, the comparison of the different conceptual designs should also contribute to the desideratum of clearly specified evaluation methods (Gifford et al., 2018, 32) for such systems. The artistic research in music will extend from the dyadic interaction between human soloist and machine to more complex collective settings.

4 Acknowledgements

This work was supported by LEONARDO – Center for Creativity and Innovation, which is funded by the German Federal and States initiative "Innovative Hochschule".

References

Bello, J., Daudet, L., Abdullah, S., Duxbury, C., Davies, M., & Sandler, M. (2005). A tutorial on onset detection in music signals. *IEEE Transactions on Speech and Audio Processing*, *13*(5).

Bown, O., & McCormack, J. (2011). Creative Agency: A Clearer Goal for Artificial Life in the Arts. In G. Kampis, I. Karsai, & E. Szathmáry (Eds.), *Advances in Artificial Life. Darwin Meets von Neumann* (pp. 254–261). Berlin, Heidelberg: Springer.

Braidotti, R. (2016). The Critical Posthumanities; or, Is Medianatures to Naturecultures as Zoe Is to Bios? *Cultural Politics*, *12*(3), 380–390.

Brown, A. R. (2018). Creative improvisation with a reflexive musical bot. *Digital Creativity*, *29*(1), 5–18.

Clayton, M., Sager, R., & Will, U. (2005). In time with the music: The concept of entrainment and its significance for ethnomusicology. In *European Meetings in Ethnomusicology* (Vol. 11, pp. 1–82). Romanian Society for Ethnomusicology.

Dong, H.-W., Hsiao, W.-Y., Yang, L.-C., & Yang, Y.-H. (2017). Musegan: Multi-track sequential generative adversarial networks for symbolic music generation and accompaniment. *arXiv:1709.06298*.

Gifford, T., Knotts, S., McCormack, J., Kalonaris, S., Yee-King, M., & d'Inverno, M. (2018). Computational systems for music improvisation. *Digital Creativity*, *29*, 19–36.

Hoffman, G., & Weinberg, G. (2010). Shimon: An interactive improvisational robotic marimba player. In *CHI '10 Extended Abstracts on Human Factors in Computing Systems* (pp. 3097–3102). New York, NY, USA: Association for Computing Machinery.

Huang, C.-Z. A., Vaswani, A., Uszkoreit, J., Shazeer, N., Simon, I., Hawthorne, C., … Eck, D. (2018). Music Transformer. *arXiv:1809.04281*.

Jurafsky, D., & Martin, J. H. (2020). Speech and language processing. Prentice Hall.

Keller, R., & Morrison, D. (2007). A grammatical approach to automatic improvisation. *Proceedings of the 4th Sound and Music Computing Conference, SMC 2007*.

Klein, J. (2018). The Mode is the Method - or How Research Can Become Artistic. *Artistic Research - Is There Some Method?*. Retrieved from https://www.academia.edu/36239994/The_Mode_is_the_Method_-_or_How_Research_Can_Become_Artistic

Marom, Y. (1997). *Improvising Jazz with Markov Chains* (Unpublished doctoral dissertation). The University of Western Australia.

Miranda, E. R., Wanderley, M. M., & Kirk, R. (2006). *New digital musical instruments: Control and interaction beyond the keyboard*. A-R Editions.

Putman, A., & Keller, R. (2015). A transformational grammar framework for improvisation. In *First International Conference on New Music Concepts.*

Roberts, A., Engel, J., Raffel, C., Hawthorne, C., & Eck, D. (2019). A hierarchical latent vector model for learning long-term structure in music. *arXiv:1803.05428*.

Rowe, R. (1993). *Interactive music systems: Machine listening and composing.* Cambridge, Mass.: MIT Press.

Simon, I., Morris, D., & Basu, S. (2008). MySong: Automatic accompaniment generation for vocal melodies. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 725–734). New York, NY, USA: Association for Computing Machinery.

Smith, B. D., & Garnett, G. E. (2012). Unsupervised play: Machine learning toolkit for max. In G. Essl, R. B. Gillespie, M. Gurevich, & S. O'Modhrain (Eds.), *12th international conference on new interfaces for musical expression, NIME 2012, ann arbor, michigan.* nime.org.

Trump, S. (2021). *Spirio Sessions Demo Videos.* Zenodo. Retrieved from `https://doi.org/10.5281/zenodo.4635617` doi: 10.5281/zenodo.4635617

Ullrich, M., & Trump, S. (in press). Sonic Collaborations between Humans, Nonhuman Animals and Artificial Intelligences: Contemporary and Future Aesthetics in More-Than-Human Worlds. *Organised Sound*, *28*(1).