

NONOTO: A Model-agnostic Web Interface for Interactive Music Composition by Inpainting

Théis Bazin

Sony CSL
Paris, France

theis.bazin@outlook.com

Gaëtan Hadjeres

Sony CSL
Paris, France

gaetan.hadjeres@sony.com

Abstract

Inpainting-based generative modeling allows for stimulating human-machine interactions by letting users perform stylistically coherent local editions to an object using a statistical model. We present *NONOTO*, a new interface for interactive music generation based on inpainting models. It is aimed both at researchers, by offering a simple and flexible API allowing them to connect their own models with the interface, and at musicians by providing industry-standard features such as audio playback, real-time MIDI output and straightforward synchronization with DAWs using Ableton Link.

Keywords interfaces, generative models, inpainting, interactive music generation, web technologies, open-source software

Introduction

We present a web-based interface that allows users to compose symbolic music in an interactive way using generative models for music. We strongly believe that such models only reveal their potential when actually used by artists and creators. While generative models for music have been around for a while (Boulanger-Lewandowski, Bengio, and Vincent 2012; Hadjeres, Pachet, and Nielsen 2017; Roberts et al. 2018), the conception of A.I.-based interactive interfaces designed for music creators is still burgeoning. We contribute to this emerging area by providing a general web interface for many music generation models so that researchers in the domain can easily test and promote their works in actual music production and performance settings. This desire follows from the seminal work by Theis, van den Oord, and Bethge, in which the authors advocate that quantitative evaluation of generative models in an unambiguous way is hard and that "generative models need to be evaluated directly with respect to the application(s) they were intended for" (Theis, van den Oord, and Bethge 2015). Lastly, we hope that the present work will contribute in making A.I.-assisted composition accessible to a wider audience, from non musicians to professional musicians, helping bridge the gap between these communities.

Drawing inspiration from recent advances in interactive interfaces for image restoration and editing (Isola et al. 2016; Jo and Park 2019; Yu et al. 2018), we focus on provid-

ing an interface for "inpainting" models for symbolic music, which are models that are able to recompose *a portion of a score* given all the other portions. The reason is that such models are more suited for an *interactive* use (compared to models generating a full score all at once) and let users play an active part in the compositional process. As an outcome, users can feel that the composition is the result of their work and not just something created by the machine. Furthermore, allowing quick exploration of musical ideas in a playful setting can enhance creativity and provide accessibility: the "technical part" of the composition is taken care of by the generative model which allows musicians as well as non experts in music to express themselves more freely.

Contributions: The key elements of novelty are: (a) easy-to-use and intuitive interface for users, (b) easy-to-plug interface for researchers allowing them to explore the potential of their music generation algorithms, (c) web-based and model-agnostic framework, (d) integration of existing music inpainting algorithms, (e) novel paradigms for A.I.-assisted music composition and live performance, (f) integration in professional music production environments.

The code for the interface is distributed under a GNU GPL license and available, along with ready-to-use packaged standalone applications and video demonstrations of the interface, on our GitHub¹.

Existing approaches

The proposed system is akin to the FlowComposer system (Papadopoulos, Roy, and Pachet 2016) which offers to generate sheets of music by performing local updates (using Markov Models in their case). However, this interface does not exhibit the same level of interactivity as ours since no real-time audio nor MIDI playback is available, which limits the tool to solely studio usage and makes for a less spontaneous and reactive user experience.

The recent tools proposed by the Google Magenta team as part of their Magenta Studio effort (Adam Roberts 2019) are more aligned with our aims in this project: they offer a selection of Ableton Live plugins (using Max for Live) that make use of various symbolic music generation models for rhythm as well as melody (Roberts et al. 2018;

¹<https://github.com/SonyCSLParis/NONOTO>

Huang et al. 2018). Similarly, the StyleMachine, developed by Metacreative Technologies ², is a proprietary tool that allows to generate midi-tracks into Ableton Live in various dance music styles using a statistical model trained on different stylistic corpora of electronic music. Yet these tools differ from ours in the generation paradigm used: they offer either continuation-based (the model generates the end of a sequence given the beginning) or complete generation (the model generates a full sequence, possibly given a template), thus breaking the flow of music on new generations. We believe that this limits their level of interactivity as opposed to local, inpainting-based models as ours, as mentioned previously. In particular, it hinders their usage in live, performance contexts.

Suggested mode of usage

The interface displays a musical score which loops forever as shown in Figure 1. Users can then modify the score by regenerating any region only by clicking/touching it. The displayed musical score is updated instantly without interrupting the music playback. Other means of control are available depending on the specificity of the training dataset: we implemented, for instance, the positioning of fermatas in the context of Bach chorales generation or the control of the chord progression when generating Jazz leadsheets or Folk songs. These *metadata* are sent, along with the sheet, to the generative models when performing re-generations. The scores can be seamlessly integrated in a DAW so that the user (or even other users) can shape the sounds, add effects, play the drums or create live mixes. This creates a new jam-like experience in which the user controlling the A.I. can be seen as just one of the multiple instrument players. This interface thus has the potential to create a new environment for collaborative music production and performance.

Since our approach is flexible, our tool can be used in conjunction with other A.I.-assisted musical tools like Magenta Studio (Adam Roberts 2019) or the StyleMachine (from Metacreative Technologies).

Technology

Framework

Our framework relies on two elements: an interactive web interface and a music inpainting server. This decoupling is strict so that researchers can easily plug-in new inpainting models with little overhead: it suffices to implement how the music inpainting model should function given a particular user input. We make heavy use of modern web browser technologies, making for a modular and hackable code-base for artists and researchers, allowing e.g. to edit the interface to allow for some particular means of interaction or to add support for some new metadata specific to a given corpus.

Interface The interface is based on OpenSheetMusicDisplay ³, a TypeScript library aimed at rendering MusicXML

²<https://metacreativetech.com/products/stylemachine-lite/>

³<https://github.com/opensheetmusicdisplay/opensheetmusicdisplay/>

sheets with vector graphics. Using Tone.js (Mann 2015), a JavaScript library for real-time audio synthesis, we augmented OSMD with real-time audio playback capacities, allowing users to preview the generated music in real-time from within the interface. Furthermore, the audio playback is uninterrupted by re-generations, enabling a truly interactive experience.

Generation back-end and communication For better interoperability, we rely on the MusicXML standard to communicate scores between the interface and the server. The HTTP-based communication API then just consists in two commands that are required server-side:

- A `/generate` command which expects the generation model to return a fresh new sheet of music in the MusicXML format to initialize a session,
- A `/timerange-change` command which takes as parameter the position of the interval to re-generate. The server is then expected to return an updated sheet with the chosen portion regenerated by the model using the current musical context.

DAW integration In order for NONOTO to be readily usable in traditional music production and performance contexts, we implemented the possibility of integrating the generated scores in any DAW in real time. To this end, we provide the user with the option of either rendering the generated sheet to audio in real-time from within the web interface using Tone.js or of routing it via MIDI to any virtual MIDI port on the host machine, using the JavaScript bindings to the Web MIDI API, WebMidi.js ⁴. We also integrated support for Ableton Link ^{5,6}, an open-source technology developed by Ableton for easy synchronization of musical hosts on a local network, allowing to synchronize the interface with e.g. Ableton Live. Adding support for these technologies does not represent novel advances on our side *per se*, yet, paired with the support of arbitrary generation back-ends, they allow to quickly test new generation models in a standard music production environment with minimal overwork and make for a beneficial tool for researchers – and the first of its kind to our knowledge.

Conclusion

We have introduced NONOTO, an interactive, open-source and hackable interface for music generation using inpainting models. We invite researchers and artists alike to make it their own by developing new models or means of interacting with those. This high level of hackability is to a large extent permitted by the wide range of technologies now offered in a very convenient fashion by modern web browsers, from which we draw heavily. Ultimately, we hope that providing tools such as ours with a strong focus on usability, accessibility, affordance and hackability will help shift the general perspective on machine learning for music creation, transitioning from the current and somewhat negative view

⁴<https://github.com/djipco/webmidi>

⁵<https://github.com/Ableton/link/>

⁶<https://github.com/2bbb/node-abletonlink>

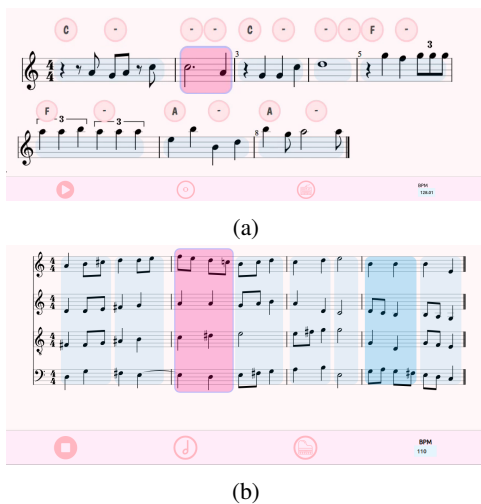


Figure 1: Our web interface used on different datasets: 1a melody and symbolic chords format, 1b four-part chorale music.

of "robot music", replacing musicians, towards a more realistic and humbler view of it as *A.I.-assisted* music.

Acknowledgements

The authors would like to thank the anonymous reviewers for their helpful comments.

References

- Adam Roberts, Yotam Mann, J. E. C. R. 2019. Magenta studio. <https://magenta.tensorflow.org/studio-announce>.
- Boulanger-Lewandowski, N.; Bengio, Y.; and Vincent, P. 2012. Modeling Temporal Dependencies in High-Dimensional Sequences: Application to Polyphonic Music Generation and Transcription. *arXiv e-prints* arXiv:1206.6392.
- Hadjeres, G.; Pachet, F.; and Nielsen, F. 2017. DeepBach: a steerable model for Bach chorales generation. In *Proc. of the 34th International Conference on Machine Learning (ICML)*, 1362–1371.
- Huang, C. A.; Vaswani, A.; Uszkoreit, J.; Shazeer, N.; Hawthorne, C.; Dai, A. M.; Hoffman, M. D.; and Eck, D. 2018. An improved relative self-attention mechanism for transformer with application to music generation. *CoRR* abs/1809.04281.
- Isola, P.; Zhu, J.-Y.; Zhou, T.; and Efros, A. A. 2016. Image-to-image translation with conditional adversarial networks. *ArXiv*.
- Jo, Y., and Park, J. 2019. SC-FEGAN: Face Editing Generative Adversarial Network with User's Sketch and Color. *arXiv e-prints* arXiv:1902.06838.
- Mann, Y. 2015. Interactive music with Tone.js. In *Proceedings of the 1st Web Audio Conference*.
- Papadopoulos, A.; Roy, P.; and Pachet, F. 2016. Assisted lead sheet composition using FlowComposer. In Rueher, M.,

ed., *Principles and Practice of Constraint Programming*, 769–785. Cham: Springer International Publishing.

Roberts, A.; Engel, J.; Raffel, C.; Hawthorne, C.; and Eck, D. 2018. A hierarchical latent vector model for learning long-term structure in music. In *Proceedings of the 35th International Conference on Machine Learning*, 4364–4373.

Theis, L.; van den Oord, A.; and Bethge, M. 2015. A note on the evaluation of generative models. *arXiv e-prints* arXiv:1511.01844.

Yu, J.; Lin, Z.; Yang, J.; Shen, X.; Lu, X.; and Huang, T. S. 2018. Free-form image inpainting with gated convolution. *CoRR* abs/1806.03589.