VISUAL SUPPORT FOR HUMAN-AI CO-COMPOSITION

Simeon Rau* **Frank Heyen*** **Michael Sedlmair**
*Equal contributions VISUS, University of Stuttgart, Germany
simeonrau@gmx.de, {frank.heyen, michael.sedlmair}@visus.uni-stuttgart.de

ABSTRACT

We propose a visual approach for AI-assisted music composition, where the user interactively generates, selects, and adapts short melodies. Based on a short input melody, we automatically generate multiple continuation samples. Repeating this step and in turn generating continuations for these samples results in a tree or graph structure, where nodes represent melodies and links connect nodes that are adjacent in time. We visualize this structure with two visualizations, where nodes display the piano roll of the corresponding sample. By interacting with these visualizations, the user can quickly listen to, choose, and adapt melodies, to iteratively create a composition. A third visualization provides an overview over larger numbers of samples, allowing for insights into the AI's predictions and its sample space.

1. INTRODUCTION

Music composition is difficult for two reasons: First, the composer needs to know at least some music theory, including rules and patterns that allow to create the intended sound and feeling. Second, lack of inspiration makes it hard to come up with new patterns and melodies.

Music composition AI [1] allows users to generate entire compositions as either notes [2–5] or audio [6]. These finished results are useful, for instance, as video background music, but lack personality and likely will not completely replace human-composed pieces. Instead, the user would like to be in control of an AI's configuration, but also be able to choose from different suggestions and adapt them to their taste [7].

To this end, we propose a user-centered approach, where the user queries the AI for multiple melody suggestions at once and then chooses between them. Visualization eases this choice, as it allows the user to focus on or listen to only the most interesting-looking samples.

In summary, we contribute an AI- and visualization-driven approach that helps hobby musicians or composers with need for inspiration compose melodies interactively.

2. RELATED WORK

A survey by Huang et al. [7] on usage and requirements of AI for music co-creation found that users desire control, authorship, and creative freedom and often create multiple samples for a part to choose from. Our approach addresses these requirements and facilitates sample choice by visualizing multiple samples and second-level continuations at the same time. Control and creative freedom are additionally supported through adaptions such as fill-in replacements or editing single notes.

Bach Doodle [8] allows users to harmonize an input melody in the style of Bach using *CoCoNet* [9]. Louie et al. [10] later improved the usability for novice users. Our approach instead supports composing a melody in the first place, that could then be harmonized through CoCoNet or similar models as well. *SketchNet* [11] gives users more control in music generation by letting them sketch melodies and filling in. In contrast, our approach allows users to choose and edit suggestions before combining them into a composition. SketchNet could be integrated to make these suggestions more relevant.

Park et al. [12] propose a recursive metric for symbolic melodic similarity that could be applied for music information retrieval. Gotham et al. [13] follow a similar approach using edit distances. Instead of comparing whole pieces, we use similarity metrics to sort or visually layout short melodies by similarity to form summaries.

Abstractly, our approach follows the idea of visual parameter analysis [14], which has not been explored systematically for semi-automatic music composition yet.

3. DESIGN

We now explain our intended target users, followed by a description of our main visualizations.

3.1 Users, Tasks, and Workflow

We target two user roles, that can be shared by the same person: The *composer* and the *AI analyst*. Composers would use our approach when lacking experience or inspiration. To have maximum control over the process, they want to create a composition step by step. For each step, they need to choose between different possible continuations that they may want to adapt or later replace.

AI analysts want to see how a model works and what kinds of melodies it predicts. They therefore want to see the influence of different starting melodies and hyperparameters, such as temperature, on produced samples.

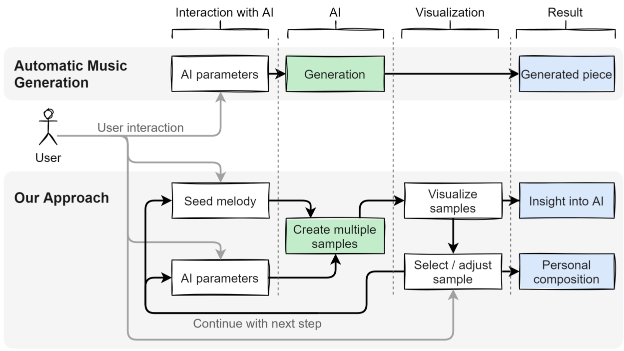

Figure 1. Our approach gives users control and artistic freedom through iterative choice and adaption of AI-suggested melodies. Users gain insight into the AI's behavior.

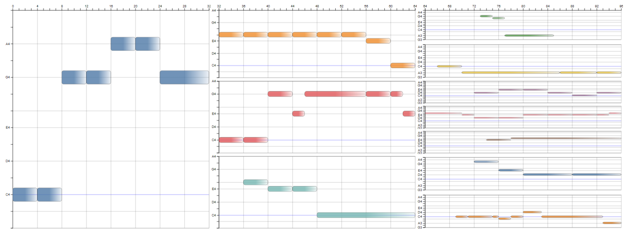

Figure 2. We visualize a tree or graph of melody samples similar to an icicle plot with piano rolls in all nodes.

Figure 1 shows our user workflow in comparison to automatic music generation. Our users start with some seed melody, by entering notes through mouse clicks or a MIDI device. Based on this seed, we generate multiple continuation samples with Magenta's *basicRNN* [1] using different hyperparameter settings. Repeating this step and in turn generating continuations for these samples results in a tree of melodies. When replacing an existing section of the composition using fill-in, the tree turns into a directed acyclic graph (see Figure 3). Each level of the graph corresponds to one temporal section of the composition.

3.2 Melody Graphs

We visualize our melody graph with an icicle plot, where nodes display the piano rolls of the corresponding samples (Figure 2). These piano rolls share a common time axis from left to right, but each has its own pitch axis from bottom to top. We allow interaction with the nodes, such as adding their notes to the current composition or listening to a single melody sample or the whole path.

To support sample selection, we allow sorting samples in each graph level. For example, the user could sort by the variance of intervals between notes to focus on more vivid melodies. Since connections between nodes would be lost in an icicle plot when sorting, we extended it by adding links between nodes (Figure 3). These links connect nodes that could follow each other in the composition and encode values of the chosen sorting metric in the link's thickness.

[1] https://github.com/magenta/magenta/tree/main/magenta/models/melody_rnn

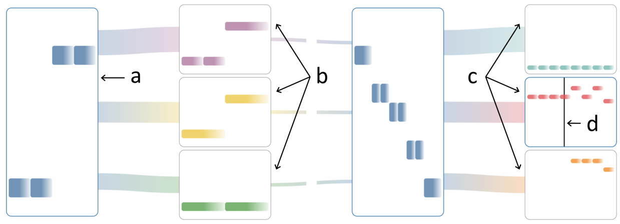

Figure 3. A Sankey-based visualization of a melody sample graph. Nodes represent samples and display them through piano rolls, links connect related nodes and encode the current sorting metric's values in their width. **a)** The full-height nodes are part of the current composition. **b)** Nodes in the middle show options for a fill-in. **c)** The right-most nodes show possible continuations. **d)** A time cursor shows that the center right node is currently played back as audio.

3.3 Melody Similarity

In order to better understand the AI model and its suggestions, the user can generate tens or hundreds of melody samples with different model hyperparameters. Since our previous visualizations do not scale to this amount of data, we chose a different design for this use case.

We want to show an overview of all samples, by visually grouping based on similarity. Our similarity metric considers the interval between notes as well as the rhythm, with a user-defined weighting between both. We apply dimensionality reduction to the pair-wise similarities to compute a 2D layout, which we then display as a scatterplot. Here, we draw melody samples as circles that encode information about hyperparameters or melody statistics in their color and radius. Using a resizable circular brush, the user can select a neighborhood of samples for which we display aggregated visualizations such as a piano roll density estimation (see also our supplemental video).

4. CONCLUSION

We propose a novel, user-centered approach for human-AI co-composition based on visualization of multiple generated suggestions. Our current implementation is limited to monophonic melodies, but could by easily extended to polyphony when using other AI models and similarity metrics. In the future, we want to test different models [3, 4] to either directly generate polyphonic music or harmonize melodies as a second step. We further plan to explore aggregations and glyphs for more compactly representing melodies. Future work also needs to empirically validate our ideas with users.

5. ACKNOWLEDGEMENTS

This work was funded by the Cyber Valley Research Fund – Project InstruData.

6. REFERENCES

[1] J. D. Fernández and F. Vico, "AI methods in algorithmic composition: A comprehensive survey," *Journal of Artificial Intelligence Research (JAIR)*, vol. 48, pp. 513–582, 2013.

[2] H. Dong, W. Hsiao, L. Yang, and Y. Yang, "MuseGAN: Multi-track sequential generative adversarial networks for symbolic music generation and accompaniment," in *32nd AAAI Conf. on Artificial Intelligence (AAAI)*, 2018.

[3] C. A. Huang, A. Vaswani, J. Uszkoreit, N. Shazeer, C. Hawthorne, A. M. Dai, M. D. Hoffman, and D. Eck, "An improved relative self-attention mechanism for transformer with application to music generation," *arXiv preprint arXiv:1809.04281*, 2018. [Online]. Available: http://arxiv.org/abs/1809.04281

[4] A. Roberts, J. Engel, C. Raffel, C. Hawthorne, and D. Eck, "A hierarchical latent vector model for learning long-term structure in music," in *Proc. 35th International Conf. on Machine Learning*, ser. Proceedings of Machine Learning Research, vol. 80. PMLR, 10–15 Jul 2018, pp. 4364–4373. [Online]. Available: https://proceedings.mlr.press/v80/roberts18a.html

[5] E. Frid, C. Gomes, and Z. Jin, "Music creation by example," in *Proc. 2020 Conf. on Human Factors in Computing Systems (CHI)*, ser. CHI '20. ACM, 2020, p. 1–13. [Online]. Available: https://doi.org/10.1145/3313831.3376514

[6] R. Manzelli, V. Thakkar, A. Siahkamari, and B. Kulis, "Conditioning deep generative raw audio models for structured automatic music," in *Proc. 19th Int. Society for Music Information Retrieval Conf. (ISMIR)*, 2018. [Online]. Available: https://ismir2018.ismir.net/doc/pdfs/192_Paper.pdf

[7] C. A. Huang, H. V. Koops, E. Newton-Rex, M. Dinculescu, and C. J. Cai, "AI song contest: Human-AI co-creation in songwriting," in *Proc. 21st Int. Society for Music Information Retrieval Conf. (ISMIR)*, 2020. [Online]. Available: https://program.ismir2020.net/static/final_papers/167.pdf

[8] C. A. Huang, C. Hawthorne, A. Roberts, M. Dinculescu, J. Wexler, L. Hong, and J. Howcroft, "The Bach Doodle: Approachable music composition with machine learning at scale," in *Proc. 20th Int. Society for Music Information Retrieval Conf. (ISMIR)*, 2019. [Online]. Available: https://archives.ismir.net/ismir2019/paper/000097.pdf

[9] T. Chakraborti, B. McCane, S. Mills, and U. Pal, "CoCoNet: A collaborative convolutional network," *arXiv preprint arXiv:1901.09886*, 2019. [Online]. Available: https://arxiv.org/abs/1901.09886

[10] R. Louie, A. Coenen, C. Z. Huang, M. Terry, and C. J. Cai, "Novice-AI music co-creation via AI-steering tools for deep generative models," in *Proc. 2020 Conf. on Human Factors in Computing Systems (CHI)*. ACM, 2020, p. 1–13. [Online]. Available: https://doi.org/10.1145/3313831.3376739

[11] K. Chen, C. Wang, T. Berg-Kirkpatrick, and S. Dubnov, "Music SketchNet: Controllable music generation via factorized representations of pitch and rhythm," in *Proc. 21st Int. Society for Music Information Retrieval Conf. (ISMIR)*, 2020. [Online]. Available: https://program.ismir2020.net/static/final_papers/146.pdf

[12] S. Park, T. Kwon, J. Lee, J. Kim, and J. Nam, "A cross-scape plot representation for visualizing symbolic melodic similarity," in *Proc. 20st Int. Society for Music Information Retrieval Conf. (ISMIR)*, 2019. [Online]. Available: https://archives.ismir.net/ismir2019/paper/000050.pdf

[13] M. Gotham and M. Ireland, "Taking form: A representation standard, conversion code, and example corpus for recording, visualizing, and studying analyses of musical form," in *Proc. 20th Int. Society for Music Information Retrieval Conf. (ISMIR)*, 2019. [Online]. Available: https://archives.ismir.net/ismir2019/paper/000084.pdf

[14] M. Sedlmair, C. Heinzl, S. Bruckner, H. Piringer, and T. Möller, "Visual parameter space analysis: A conceptual framework," *IEEE Trans. Visualization and Computer Graphics (TVCG)*, vol. 20, no. 12, pp. 2161–2170, 2014.