EFFECTS OF SINGLE-ATTRIBUTE CONTROL ON THE MUSIC GENERATED BY FIGARO

Anonymous authors
Paper under double-blind review

ABSTRACT

We have experimented with controlling the musical attributes of the music generated by the FIGARO model, while evaluating the success of control and the overall musical quality. The results suggest non-trivial correlations between the musical attributes and the musical quality metrics.

1 INTRODUCTION

The task of music generation has been increasingly interesting for the machine learning literature, especially since the evolution of language models such as the transformer Vaswani et al. (2017). Most practical applications require control over the generated music e.g. background music for videos. In this study, we attempt controlling the attributes of the generated music using the state-of-the-art model FIGARO, proposed by von Rütte et al. (2022). We aim to answer the following questions:

- How do different musical attributes respond to control?
- Which musical features affect the musical quality more when controlled?

2 RELATED WORK

The current state-of-the-art models for music generation employ the Music Transformer model architecture Huang et al. (2018), and generate music in the symbolic MIDI format Loy (1985) which can be converted to audio. When using transformers, the MIDI data is tokenized using format such as MIDI-like, REMI Ren et al. (2020), and CPWord Hsiao et al. (2021). More recent models also allow controlling the generated music e.g. Wu & Yang (2022), Shih et al. (2022).

The subject of this study is the FIGARO model, proposed by von Rütte et al. (2022). It allows the fine-grained control of music by using a description sequence representation that gets converted to the output sequence using a Transformer. The *description* is the combination of a learned latent representation of the input REMI-tokenized sequence, and a "domain expert" description with hand-crafted features. The FIGARO study did not experiment with altering individual attributes in the description sequence, hence the motivation behind our work.

3 METHODOLOGY

To answer the RQs for the FIGARO model, we generate music sequences with the `figaro_expert` checkpoint that uses only the expert description function (no latent representation). The generated sequences come from descriptions sampled from the Clean MIDI Subset of the LMD dataset [1], with an altered attribute in the controlled cases. We used MIDI files of real songs in order to be consistent with the distribution of a real music dataset. We then evaluate the extent to which the attribute control was achieved, and how did the attribute control affect the musical quality.

For each input description, we create 6 extra altered descriptions. For each description, only one attribute is altered over the whole sequence. For the Instruments attribute, we randomly remove one

[1] `https://colinraffel.com/projects/lmd/`

Table 1: Correlation between the attribute control amount and the evaluation metrics (evaluation metrics explained in Table 2)

Control	Ch	P	V	D	ND	H1	H4	GS	CPI
Chords	0.01	-0.31	0.03	-0.09	0.24	-0.12	-0.12	0.01	-0.16
Mean pitch	-0.30	0.23	0.02	-0.01	0.08	0.08	0.02	0.46	-0.17
Mean velocity	0.01	0.04	-0.12	0.10	-0.25	0.74	0.75	0.01	-0.13
Mean duration	0.20	0.02	-0.24	-0.07	-0.03	0.65	0.65	0.21	0.30
Note density	-0.14	-0.01	0.12	0.00	-0.02	-0.83	-0.87	-0.87	0.91

instrument from the whole sequence. For the chords attributes, we transpose all the chords of the sequence by a certain number of steps. For the Mean Pitch, Mean Duration, Mean Duration and Note Density, we shift their value by a uniformly-random delta over the whole sequence.

We evaluate the generated music with the metrics of the FIGARO study for attribute fidelity, and with 3 metrics of MusDr (Wu & Yang, 2020) for quality, the metrics are in 2 in Appendix A

4 Results and discussion

The mean values of the evaluation metrics are in table 2. We notice the expected drop in attribute fidelity for the controlled attribute, however the quality metrics don't seem to drastically vary, which suggests that the quality of FIGARO's output is consistent when controlled. Table 1 shows that some attributes are correlated with the metrics of other attributes e.g. controlling the pitch is negatively correlated with the chords f1-score. The control of certain attributes is also correlated with quality metrics, which might not seem related to the attribute e.g. controlling Velocity or Duration correlates with a change in the pitch histogram entropy. The correlations suggest a learned association that the FIGARO model has learned. Finally, figure 1 shows the response of each attribute's metric to control. Mean pitch and Velocity are harder to control with bigger shifts, while Duration is more controllable if decreased. Chord transposition does not seem to affect the f1-score of chords which is low to begin with.

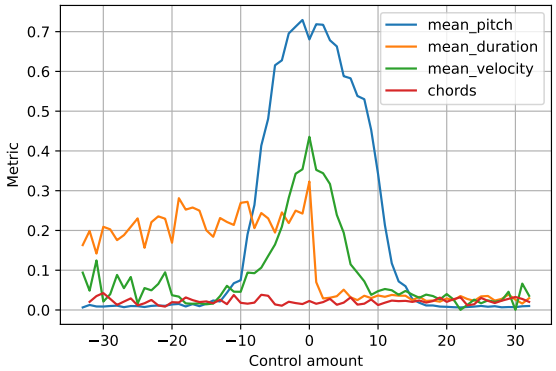

Figure 1: Response of the attribute-specific evaluation metrics to attribute control (More is better)

5 Conclusion

With the current results, we can already conclude that FIGARO maintains a consistent music quality when individual musical attributes are controlled, and that it has learned non-trivial correlations between these attributes. Certain attributes are less responsive to control (such as chords), which can be further understood with probe or ablation studies.

URM STATEMENT

The authors acknowledge that at least one key author of this work meets the URM criteria of ICLR 2023 Tiny Papers Track.

REFERENCES

Wen-Yi Hsiao, Jen-Yu Liu, Yin-Cheng Yeh, and Yi-Hsuan Yang. Compound Word Transformer: Learning to Compose Full-Song Music over Dynamic Directed Hypergraphs. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(1):178–186, May 2021. ISSN 2374-3468. doi: 10.1609/aaai.v35i1.16091. URL https://ojs.aaai.org/index.php/AAAI/article/view/16091. Number: 1.

Cheng-Zhi Anna Huang, Ashish Vaswani, Jakob Uszkoreit, Noam Shazeer, Ian Simon, Curtis Hawthorne, Andrew M. Dai, Matthew D. Hoffman, Monica Dinculescu, and Douglas Eck. Music Transformer, December 2018. URL http://arxiv.org/abs/1809.04281. arXiv:1809.04281 [cs, eess, stat].

Gareth Loy. Musicians Make a Standard: The MIDI Phenomenon. *Computer Music Journal*, 9(4): 8–26, 1985. ISSN 0148-9267. doi: 10.2307/3679619. URL https://www.jstor.org/stable/3679619. Publisher: The MIT Press.

Yi Ren, Jinzheng He, Xu Tan, Tao Qin, Zhou Zhao, and Tie-Yan Liu. PopMAG: Pop Music Accompaniment Generation. In *Proceedings of the 28th ACM International Conference on Multimedia*, pp. 1198–1206, Seattle WA USA, October 2020. ACM. ISBN 978-1-4503-7988-5. doi: 10.1145/3394171.3413721. URL https://dl.acm.org/doi/10.1145/3394171.3413721.

Yi-Jen Shih, Shih-Lun Wu, Frank Zalkow, Meinard Müller, and Yi-Hsuan Yang. Theme Transformer: Symbolic Music Generation with Theme-Conditioned Transformer, March 2022. URL http://arxiv.org/abs/2111.04093. arXiv:2111.04093 [cs, eess].

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is All you Need. pp. 11, 2017.

Dimitri von Rütte, Luca Biggio, Yannic Kilcher, and Thomas Hofmann. FIGARO: Generating Symbolic Music with Fine-Grained Artistic Control, March 2022. URL http://arxiv.org/abs/2201.10936. arXiv:2201.10936 [cs, eess, stat].

Shih-Lun Wu and Yi-Hsuan Yang. The Jazz Transformer on the Front Line: Exploring the Shortcomings of AI-composed Music through Quantitative Measures, August 2020. URL http://arxiv.org/abs/2008.01307. arXiv:2008.01307 [cs, eess].

Shih-Lun Wu and Yi-Hsuan Yang. MuseMorphose: Full-Song and Fine-Grained Piano Music Style Transfer with One Transformer VAE, December 2022. URL http://arxiv.org/abs/2105.04090. arXiv:2105.04090 [cs, eess].

A APPENDIX: FURTHER NUMERICAL RESULTS

Table 2: The mean value for each evaluation metric for each controlled attribute. The metrics are the Note Density NRMSE (ND), the Pitch OA as described in the FIGARO study (P), Duration OA (D), Velocity OA (V), Instrument f1-score (Instr), Chords f1-score (Ch), the pitch histogram entropy for 1 bar as implemented by the MusDr tool (H1), the pitch histogram entropy for 4 bars (H4), the grooving similarity across the track (GS), and the Chord Progression Irregularity (CPI)

Attribute control	**ND**	**P**	**D**	**V**	**Instr**	**Ch**	**H1**	**H4**	**GS**	**CPI**
None	0.50	0.73	0.40	0.45	0.95	0.56	1.87	2.32	0.98	0.71
Instrument	0.63	0.69	0.37	0.41	0.87	0.44	1.72	2.16	0.98	0.71
Chords	0.27	0.72	0.40	0.44	0.95	0.03	1.86	2.33	0.98	0.71
Pitch	0.65	0.18	0.33	0.39	0.93	0.50	1.74	2.17	0.98	0.69
Velocity	1.11	0.69	0.10	0.08	0.87	0.50	1.79	2.24	0.98	0.71
Duration	0.34	0.72	0.12	0.44	0.94	0.52	1.81	2.25	0.98	0.70
Note density	17.39	0.64	0.21	0.37	0.87	0.36	1.35	1.68	0.98	0.79