RhythGen: Controlling rhythmic complexity in generating symbolic music for interactive serious game interventions.

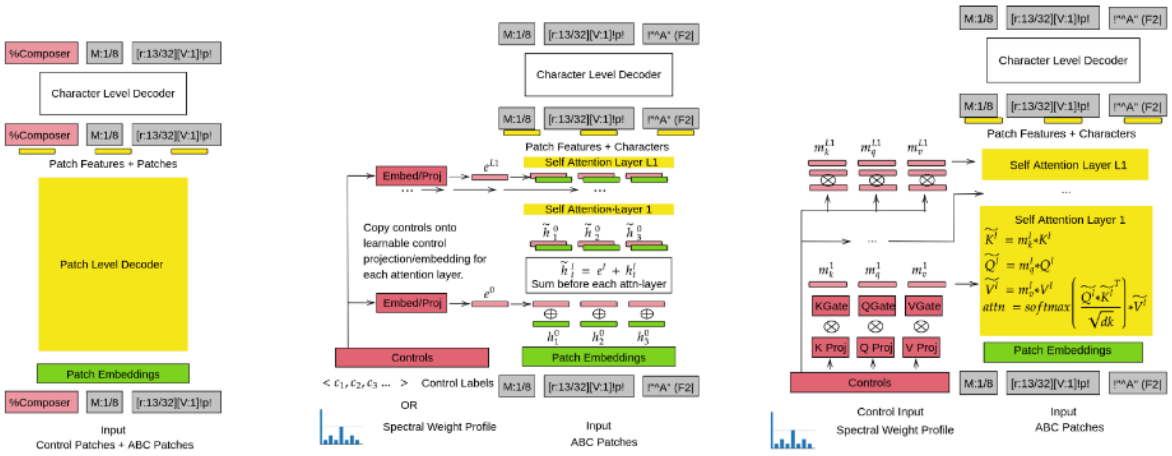
Authors
Efraim Dahl - Utrecht University - e.y.c.dahl@students.uu.nl
Peter van Kranenburg - Utrecht University - P.vanKranenburg@uu.nl
Anja Volk - Utrecht University - A.Volk@uu.nl

Music is essential in video games, enhancing immersion and engagement of players. However, players tend to disengage from the game music due to excessive repetition or due to differing musical preferences [1]. This is particularly problematic in serious games for therapeutic use such as Musical Attention Control Training (MACT) [2] where sustained engagement and interaction with the game music is essential for the intervention. Music generation offers a scalable solution, enabling a greater variety of music and adapting the music to a player's abilities and preferences, ultimately improving engagement and the effect of the intervention.

To this end, we introduce RhythGen, a transformer-based music generator [10,11] that extends the pretrained symbolic music generator NotaGen[3] with time-varying control over rhythmic structure and complexity. In rhythmic training applications, variations in rhythmic complexity, especially syncopation, are used to vary the cognitive load of music to stimulate attention and working memory [4,5]. Through incorporating control over syncopation levels in generated music, RhythGen expands the range of musical variety, and enables the use of progressively more challenging music, adjusting to a participants' abilities and training needs.

NotaGen is extended with control pathways and conditioned using either Inner Metric Analysis (IMA) [6], or syncopation labels derived thereof [7]. We introduce attention modulation as a novel control mechanism and compare it to adaptive in-attention conditioning [7,8]. Our evaluation explores the tradeoff between generation quality and control fidelity when using IMA versus discrete labels. Results show that label-conditioned models effectively produce music with targeted syncopation levels, while IMA-conditioned models primarily control note density, without capturing syncopation. The ability of our model to produce musical pieces of varying rhythmic complexity, to introduce a recognizable change in rhythmic structure within a piece and to create enjoyable music, is further demonstrated in a small-scale user study.

Figure



(a) Original Notagen model with standard conditioning

(b) RhythGen with In-Attention Conditioning

(c) RhythGen with Attention Modulation

Figure 1: The architecture of RhythGen using different conditioning mechanisms

References

- [1] Rogers, K., & Weber, M. (2019, September). Audio Habits and Motivations in Video Game Players. *Proceedings of the 14th International Audio Mostly Conference: A Journey in Sound*, 45–52. doi:10.1145/3356590.3356599

-[2] Volk, A., Hakvoort, L., Schaefer, R. S., Chalkiadakis, E., & Bakkes, S. (2023). Developing and evaluating a Musical Attention Control Training computer game application. SMC-*Proceedings* , *2023-June*, 261–268.

[3] Wang, Y., Wu, S., Hu, J., Du, X., Peng, Y., Huang, Y., … Sun, M. (2025). NotaGen: Advancing Musicality in Symbolic Music Generation with Large Language Model Training Paradigms. *arXiv [Cs.SD]*.

[4] Zanto TP, Giannakopoulou A, Gallen CL, Ostrand AE, Younger JW, Anguera-Singla R, Anguera JA, Gazzaley A. Digital rhythm training improves reading fluency in children. Dev Sci. 2024 May;27(3)

[5] Kirschner S, Tomasello M. Joint drumming: social context facilitates synchronization in preschool children. J Exp Child Psychol. 2009 Mar;102(3):299-314. doi: 10.1016/j.jecp.2008.07.005. Epub 2008 Sep 12. PMID: 18789454.

[6] Volk, A. (2008). The Study of Syncopation Using Inner Metric Analysis: Linking Theoretical and Experimental Analysis of Metre in Music. *Journal of New Music Research*, *37* (4), 259–273.

[7] Bemman, B., & Christensen, J. (11 2024). Inner Metric Analysis as a Measure of Rhythmic Syncopation. ISMIR-*Proceedings* 389–396.

[8] Lan, Y.-H., Hsiao, W.-Y., Cheng, H.-C., & Yang, Y.-H. (2024, July). MusiConGen: Rhythm and Chord Control for Transformer-Based Text-to-Music Generation. ISMIR - Proceedings

[9] Wu, S.-L., & Yang, Y.-H. (2023). MuseMorphose: Full-Song and Fine-Grained Piano Music Style Transfer with One Transformer VAE. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*.

[10] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention Is All You Need. *NeurIPS*:

[11] -Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). *Language models are unsupervised multitask learners*. OpenAI

-Nestke, A., & Noll, T. (2001). Inner Metric Analysis. Tatra Mountains Publications, 91-111

-Koops, H. V., Volk, A., & de Haas, W. B. (2015). Corpus-Based Rhythmic Pattern Analysis of Ragtime Syncopation. ISMIR - *Proceedings* 483–489.

-Chalkiadakis, E. (2022). *Developing and evaluating a Musical Attention Control Training computer game application* (Utrecht University, Utrecht, NL).

-Copet, J., Kreuk, F., Gat, I., Remez, T., Kant, D., Synnaeve, G., … Défossez, A. (2023). Simple and Controllable Music Generation. *NeurIPS*.

- Wu, S., Wang, Y., Yuan, R., Guo, Z., Tan, X., Zhang, G., Zhou, M., Chen, J., Mu, X., Gao, Y., Dong, Y., Liu, J., Li, X., Yu, F., & Sun, M. (2024). CLaMP 2: Multimodal Music Information Retrieval Across 101 Languages Using Large Language Models.

Thul, E., & Toussaint, G. (2008). Rhythm Complexity Measures: A Comparison of Mathematical Models of Human Perception and Performance. Proceedings of ISMIR.