Interactive Experiences with a Web-based Drummer Bot for Finger-Tapping

Çağrı Erdem and Carsten Griwodz
Department of Informatics
University of Oslo, Norway
{cagrie, griff}@ifi.uio.no

Abstract

This extended abstract summarizes the findings of a user study conducted with *dB*, a web-based "drummer bot" designed for interactive groove-making. dB utilizes a Variational Autoencoder (VAE) to transform simple rhythmic inputs into complex drum patterns with microtiming and dynamics. We explore the interactive experiences of a diverse user base in connection with generative parameters.

Introduction

Artificial intelligence (AI) has been used to translate body movements into musical outputs since the early 1990s (Lee et al., 1991), accompanied by a growing interest in interactive multi-agent systems (Collins, 2006; Tatar and Pasquier, 2019). Our previous research focused on developing AIs equipped with what Erdem (2022) calls "embodied perspectives" to sense and interpret human bodily signals in live music (Erdem and Jensenius, 2020; Erdem et al., 2020, 2022). A recurring common critique in the user studies of these systems was the need for co-creative AI to progress beyond sounding accuracy in its responses to "understand" performers' changing moods for better communication, especially in improvisational contexts (Erdem et al., 2023).

Therein lies a gap in our knowledge of what genuinely engages or disengages users in AI music interactions, prompting a need for more inclusive and diverse user studies. In response, we developed *dB*, a system designed to generate musical grooves through finger-tapping, leveraging the universal appeal of rhythm (Winkler et al., 2009) and enabling user interaction without prior musical knowledge. We seek to answer how users interact with and perceive an artificial drummer entity living in the browser through finger-tapping, providing insights into user engagement under varied randomness and rhythmic density conditions.

Rhythm Generation

AI techniques are widely utilized in audio and music generation (Ji et al., 2020; Yin et al., 2023), with notable implementations such as GrooVAE's sequence-to-sequence network for drum patterns (Gillick et al., 2019), temporal convolutional networks for human improvisation (McCormack et al., 2019), and studies on microtiming in Brazilian percussion (Wright and Berdahl, 2006). Innovations extend to intelligent agents learning drumming from human movements (Tidemann et al., 2009), composing rhythms (Makris et al., 2019), generating electronic dance music (EDM) beats (Vogl et al., 2019), and rhythm accompaniment (Nuttall et al., 2021; Haki et al., 2022). Our project draws inspiration from these diverse applications but focuses on an accessible and engaging web application, emphasizing playfulness and accessibility through user-friendly interfaces and dynamic generative music elements.

From Tapping to Complex Rhythms

The technical implementation of data curation and representation, model architecture, and interface design are subject to an upcoming publication. But briefly, at the core of our Variational Autoencoder (VAE) model's architecture is the processing of quantized representations of users' finger-tapped rhythms (an array named H_t) on the computer keyboard. The true H_t inputs are created utilizing the *Pulse* layer of the "Pattern Category" concept (Senn et al., 2018), which compresses drum tracks predominantly based on cymbals. Our generative system processes tapped rhythm arrays without dynamics and microtiming information into full drum grooves encompassing Hits (H'), Time Offsets (O'), and Velocities (V') for nine predefined parts (e.g., kick, snare, etc.) of a standard drum kit.

Experiments

A preliminary 30-minute user study was conducted with 63 participants using a simplification of the original design.[1] The study comprised an initial briefing, a questionnaire, a practice session, nine 2-minute interactive dB sessions (each followed by a survey), and concluded with a final questionnaire. Participants performed three tapped rhythms per session under specific conditions detailed in the appendices and responded on a 5-point Likert scale. After discarding data from 11 participants due to incomplete entries, data from

[1]The uncluttered user interface is available at https://2groove.live/

468 interactions involving 52 participants ($\mu_{\text{age}} = 29.79$ $[15, 74]$, $\sigma = 12.50$, including 19 females, 1 nonbinary, and 32 males) were analyzed through an initial grouping as *Musicking* (23 participants) and *Non-musicking* (29 participants).

Test Conditions

We categorized each test parameter into three specific levels: low, medium, and high, creating predefined pairings that contrasted temperature and threshold settings, except in one instance where both were uniformly set to medium. This contrasting approach is crucial as it balances the novelty of generated rhythms with the model's predictive confidence. Table 1 illustrates the relationships among the three dimensions: tempo, temperature, and threshold.

Musicking vs. *Non-Musicking*

Comparisons of these base groups with respect to the "I felt bored," "I felt good," "It was tiresome," and "I felt inspired" ratings using Kruskal-Wallis H tests showed no significance. However, the post hoc Nemenyi test comparing all sessions revealed significant variations in boredom levels among further subgrouped (as detailed in Table 2) comparisons:

$$\text{"I felt bored" (group 3)} : \begin{cases} S_{1,mid} \text{ vs. } S_{3,mid},\ z = 4.690, p = 0.045 \\ S_{2,slow} \text{ vs. } S_{3,mid},\ z = 4.432, p = 0.003 \end{cases}$$

Engagement and Boredom

Our study participants' responses to the "I felt bored" statement suggest that randomness, rhythmic variation, and faster tempos contribute to the reduction of boredom. This is the conclusion of a repeated measures ANOVA that revealed significant differences ($F = 3.133, p = 0.004$) for the statement across sessions. Due to sphericity violations confirmed by Mauchly's test, the Greenhouse-Geisser correction was employed, adjusting the degrees of freedom with a correction factor ($\epsilon = 0.8139$). Subsequent post hoc analysis showed notable differences in boredom levels: the analysis of session means (Table 3) indicates that sessions $S_{3,mid}$ and $S_{3,fast}$ were consistently less boring. This aligns with findings highlighting surprise as pivotal for playful experiences (Krzyzaniak et al., 2022).

In addition, a temporal analysis of responses suggested a decrease in boredom over time, implying a potential development of mastery. Conversely, the *Musicking* group experienced increased fatigue, possibly due to the monotony or perceived uncontrollability of the system, yet they also sensed more opportunities for exploration compared to their non-musical counterparts.

Age Factor

Responses of participants identified as part of the *Generation Z* cohort (born 1997 or later) Dimock (2019) varied compared to the older group. Linear mixed models (LMMs) regression partly confirmed the significant non-normality of the Shapiro-Wilk test applied to the "I felt good," "It was tiresome," and "I felt bored" responses: positive feelings tend to increase with age ($\beta = 0.386, p = 0.003$), tiredness tends to decrease ($\beta = -0.160, p = 0.187$), and boredom levels also decrease with age ($\beta = -0.615, p < 0.001$).

Tablo 1: Cross-tabulation of session labels organizes temperature and threshold settings into low, mid, and high. Session labels ($S_{x,y}$) denote combinations of tempo (slow, mid, fast) and sequence number (1 to 3), arranged vertically by tempo and horizontally by temperature levels, with tempos of 60 BPM (slow), 90 BPM (mid), and 120 BPM (fast).

temperature		low	mid	high
threshold		high	mid	low
tempo	slow	$S_{1,slow}$	$S_{2,slow}$	$S_{3,slow}$
	mid	$S_{1,mid}$	$S_{2,mid}$	$S_{3,mid}$
	fast	$S_{1,fast}$	$S_{2,fast}$	$S_{3,fast}$

Tablo 2: Participant categories based on their perceived control during interaction and prior experience with interactive generative music, assuming experienced participants tend to waive control to artificial agents more comfortably.

	Not experienced	**Experienced**
Less control	Group 1 (15 ppl)	Group 3 (12 ppl)
More control	Group 2 (13 ppl)	Group 4 (12 ppl)

Tablo 3: The post hoc Tukey HSD test results show that Sessions 8 ($S_{3,mid}$) and 9 ($S_{3,fast}$) received significantly lower boredom levels.

A	**B**	**Mean diff.**	**T-stat**	**p-value**
$S_{1,slow}$	$S_{3,mid}$	0.5	3.01	0.004
$S_{1,mid}$	$S_{3,mid}$	0.46	3.05	0.003
$S_{2,slow}$	$S_{3,mid}$	0.57	3.48	0.001
$S_{2,slow}$	$S_{3,fast}$	0.46	3.15	0.002
$S_{2,mid}$	$S_{3,mid}$	0.4	3.27	0.001
$S_{2,fast}$	$S_{3,mid}$	0.48	3.4	0.001

Toward Perceptual Congruity

The computer keyboard is inadequate for playing rhythms comfortably. Quantization can be perceived as "low resolution," echoing the interaction vs. influence distinction of Boden and Edmonds (2009). The *Pulse Layer approach* to compress full grooves into 1D arrays was adequate, but we need to seek perceptually-based machine learning metrics for better modeling of action-response congruity. Higher randomness and rhythmic variation lead to less boredom. A balance between surprise and perceived control is necessary for feeling "skillful."

References

Boden, M. A. and Edmonds, E. A. (2009). What is generative art? *Digital Creativity*, 20(1-2):21–46. Routledge.

Collins, N. M. (2006). *Towards Autonomous Agents for Live Computer Music: Realtime Machine Listening and Interactive Music Systems*. PhD thesis, University of Cambridge.

Dimock, M. (2019). Defining generations: Where Millennials end and Generation Z begins.

Erdem, C. (2022). *Controlling or Being Controlled? Exploring Embodiment, Agency and Artificial Intelligence in Interactive Music Performance*. Doctoral thesis, University of Oslo.

Erdem, C. and Jensenius, A. R. (2020). RAW: Exploring Control Structures for Muscle-based Interaction in Collective Improvisation. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 477–482, Birmingham, UK. Zenodo.

Erdem, C., Lan, Q., Fuhrer, J., Martin, C. P., Tørresen, J., and Jensenius, A. R. (2020). Towards Playing in the 'Air': Modeling Motion-Sound Energy Relationships in Electric Guitar Performance Using Deep Neural Networks. In *978-88-945415-0-2*, pages 177–184. Axea sas/SMC Network. Accepted: 2020-09-15T18:05:18Z.

Erdem, C., Wallace, B., and Jensenius, A. R. (2022). CAVI: A Coadaptive Audiovisual Instrument–Composition. PubPub.

Erdem, , Wallace, B., Glette, K., and Jensenius, A. R. (2023). Tool or Actor? Expert Improvisers' Evaluation of a Musical AI "Toddler". *Computer Music Journal*, pages 1–17.

Gillick, J., Roberts, A., Engel, J., Eck, D., and Bamman, D. (2019). Learning to Groove with Inverse Sequence Transformations. arXiv:1905.06118 [cs, eess, stat].

Haki, B., Nieto, M., Pelinski, T., and Jordà, S. (2022). Real-Time Drum Accompaniment Using Transformer Architecture. Publication Title: Proceedings of the 3rd Conference on AI Music Creativity Publisher: AIMC.

Ji, S., Luo, J., and Yang, X. (2020). A Comprehensive Survey on Deep Music Generation: Multi-level Representations, Algorithms, Evaluations, and Future Directions. arXiv:2011.06801 [cs, eess].

Krzyzaniak, M., Erdem, , and Glette, K. (2022). What Makes Interactive Art Engaging? *Frontiers in Computer Science*, 4.

Lee, M., Freed, A., and Wessel, D. (1991). Real-Time Neural Network Processing of Gestural and Acoustic Signals. pages 277–280, Montreal, Quebec, Canada. International Computer Music Association.

Makris, D., Kaliakatsos-Papakostas, M., Karydis, I., and Kermanidis, K. L. (2019). Conditional neural sequence learners for generating drums' rhythms. *Neural Computing and Applications*, 31(6):1793–1804.

McCormack, J., Gifford, T., Hutchings, P., Rodriguez, M. T. L., Yee-King, M., and d'Inverno, M. (2019). In a Silent Way: Communication Between AI and Improvising Musicians Beyond Sound. arXiv:1902.06442 [cs].

Nuttall, T., Haki, B., and Jorda, S. (2021). Transformer Neural Networks for Automated Rhythm Generation. In *International Conference on New Interfaces for Musical Expression*.

Senn, O., Kilchenmann, L., Bechtold, T., and Hoesl, F. (2018). Groove in drum patterns as a function of both rhythmic properties and listeners' attitudes. *PLOS ONE*, 13(6):e0199604. Publisher: Public Library of Science.

Tatar, K. and Pasquier, P. (2019). Musical agents: A typology and state of the art towards Musical Metacreation. *Journal of New Music Research*, 48(1):56–105.

Tidemann, A., Öztürk, P., and Demiris, Y. (2009). A Groovy Virtual Drumming Agent. In Ruttkay, Z., Kipp, M., Nijholt, A., and Vilhjálmsson, H. H., editors, *Intelligent Virtual Agents*, Lecture Notes in Computer Science, pages 104–117, Berlin, Heidelberg. Springer.

Vogl, R., Eghbal-Zadeh, H., and Knees, P. (2019). An automatic drum machine with touch UI based on a generative neural network. In *Proceedings of the 24th International Conference on Intelligent User Interfaces: Companion*, IUI '19, pages 91–92, New York, NY, USA. Association for Computing Machinery.

Winkler, I., Háden, G. P., Ladinig, O., Sziller, I., and Honing, H. (2009). Newborn infants detect the beat in music. *Proceedings of the National Academy of Sciences*, 106(7):2468–2471. Publisher: Proceedings of the National Academy of Sciences.

Wright, M. and Berdahl, E. (2006). Towards Machine Learning of Expressive Microtiming in Brazilian Drumming.

Yin, Z., Reuben, F., Stepney, S., and Collins, T. (2023). Deep learning's shallow gains: a comparative evaluation of algorithms for automatic music generation. *Machine Learning*.