EasyChair Preprint

№ 13845

Research on the Emotional Vocabulary Used in the User Interface of Generative AI-Based Music Production System

Yeongju Seo and Kiechan Namkung

July 7, 2024

RESEARCH ON THE EMOTIONAL VOCABULARY USED IN THE USER INTERFACE OF GENERATIVE AI-BASED MUSIC PRODUCTION SYSTEM

Yeongju Seo[1][0009-0009-4892-3836] and Kiechan Namkung[2*][0000-0003-4009-4892]

[1] Kookmin University, Department of Smart Experience Design, TED. Seoul 02707, Korea
[2*] Kookmin University, Department of AI Design, College of Design. Seoul 02707, Korea
*soundux@kookmin.ac.kr

Abstract. Generative AI-based music production systems are useful tools for generating text, images, and sounds, with particularly active use in the fields of text and image generation. However, most of these AI systems' user interfaces (UI) does not yet sufficiently reflect user experience (UX) to satisfy users. This study aims to analyze the emotional vocabulary that is currently used in generative AI systems and to derive and compare the emotional vocabulary used by users to express the atmosphere of the music through experiments. This serves as a foundational study that demonstrates the necessity for more user centered development of emotional vocabulary.

Keywords: User Interface, Emotional Vocabulary, AI-Based music production System

1 Introduction

AI systems are useful tools for generating text, images, and sounds, and they are actively used, particularly in text and image generation [1].

However, most of these AI systems' user interfaces (UI) does not yet sufficiently reflect user experience (UX) to satisfy the users [2]. User experience encompasses the overall experience of user interactions within a product or a service [3], and the user interface plays a crucial role in utilizing the system as it is a key element in the creative process provided by generative AI systems [4].

The user interface of AI systems consists of various elements, including audio, voice, and text input [5]. Among these, selecting or entering text is essential for generative AI systems to generate appropriate content [6]. Therefore, the users' choice of vocabulary used in the user interface is important.

Generative AI-based music production systems use artificial intelligence algorithms to generate music [7]. The user interface of these systems is designed to allow users to generate music based on their preferences by manipulating parameters such as emotion, tempo, and length through text. Similar to other generative AI systems that rely on text, this system enables users to create music by adjusting these variables to suit the users' tastes [7].

Various research on the interaction between generative AI-based music production systems and users indicate that most current user interfaces are very difficult to use in actual music generation scenarios. Moreover, there are few preceding studies on user interfaces and emotions [2]. Since emotion plays a decisive role in music composition [8], the emotional vocabulary provided in the user interface becomes a critical element in generative AI-based music production systems.

This study aims to analyze the emotional vocabulary currently used in generative AI-based music production systems and to derive and compare the emotional vocabulary used by users to express the atmosphere of the music through experiments. This is a foundational study for designing user-centered emotional vocabulary in generative AI-based music production systems.

2 Literature Review

2.1 Current Status of Generative AI-Based Music Production Systems

Current AI systems that generate music by selecting given emotional vocabulary are Soundraw [9], Mubert [10], Evoke Music [11], Aiva [12], Ecrett Music [13], Beatoven [14], and MusicStar [15]. Most of these systems are rule-based systems, which use predefined emotional vocabulary as parameter tools to encode and express specific emotions, thus allowing the creation of musical works [16]. The emotional vocabulary used in each system is shown in Table 1.

Table 1. Emotion Vocabulary Used in Music Production AI System Interface

Music Production AI systems interface	Emotional Vocabulary
Soundraw	Epic, Happy ,Hopeful, Angry, Sentimental, and 20 others
Mubert	Fun, Beautiful, Inspirational, Peaceful, and 50 others
Evoke music	Aggressive, Busy, Danger, Frantic, Cold, and 55 others
Aiva	Calm, Excited, Fearful, Tension, Sad
Ecrett music	Chill, Dark, Uplifting, Happy, Serious, and 6 others
Beatoven	Sad, Calm, Motivational, Happy, Scary, and 11 others
MusicStar	Happy, Sad, Blessed, Loved, Adventurous, and 4 others

2.2 Research on the Relationship Between Music and Emotional Vocabulary

Previous studies dealing with emotional vocabulary in music have mostly focused on improving the technical performance of AI music tools. In particular, studies have been conducted on categorical models and dimensional models, which form the theoretical basis for designing and implementing AI-based emotional music production systems [17][18]. Research on label-based generative models, which use emotion labels to train music, has also been conducted [19][20].

However, the emotion models and emotion labels covered in these preceding studies have not beneficially impacted user-centric interfaces [19]. Users often experience

problems where music with different emotions are generated from their selected emotional vocabulary, significantly lowering the user experience in the creative process.

3 Methods

The experiment was conducted with 15 content designers who has experience with using generative AI-based music production systems. These core users of generative AI-based music production systems are mainly identified as music composers and content designers. The experiment was conducted in three main stages.
First, a focus group interview was conducted with the participants. Participants listened to 34 randomly pre-selected pieces of music, discussed the emotions they felt when listening to the music, and recorded the emotional vocabulary. Then, the derived emotional vocabulary was grouped through card sorting. Card sorting was conducted in two groups using an open card sorting method. Finally, the emotional labels obtained from the card sorting were compared with the emotional vocabulary used in existing AI generative systems (see Fig. 1).

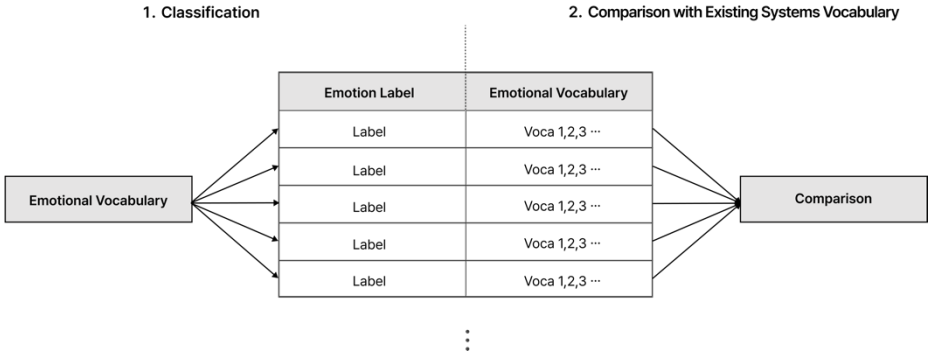

Fig. 1. Emotional Vocabulary classification and detailed comparison procedure

4 Results

The focus group interview resulted in 192 emotional vocabularies for the 34 pieces of music. This indicates that listeners can describe their emotions with various emotional vocabularies for the same music. Through the card sorting of the two groups, 14 labels were derived from Group A and 11 labels from Group B, totaling 25 emotional labels. Among these, 4 common labels—Dull, Powerful, Dreamlike, Exciting—were derived, including most of the common emotional vocabulary. However, the remaining labels showed differences, indicating that emotional expressions are subjective and can be perceived differently by individuals.

Table 2. Emotion Label

	Group A	Group B
Emotion Label	Powerful, Exciting, Dull, Dreamlike, Tranquil, Refreshing, Youthlike, Sweet, Rough, Attractive, Peppy, Overwhelming, Delicate, Sacred	Powerful, Exciting, Dull, Dreamlike, Warm, Dark, Fluttering, Emotional, Exhilarating, Distinctive, Tortuous

The comparison results with the emotional vocabulary used in existing generative AI systems are shown in Table 3.

Table 3. Comparison of Emotional Vocabulary Used in Existing Generative AI Systems with User-Derived Emotional Labels

	Emotion Vocabulary Count	Representative Vocabulary Match Count(Group A)	Representative Vocabulary Match Count(Group B)	Total Matches with AI System Vocabulary (192 terms)
Soundraw	25	0	1	1
Mubert	44	0	0	10
Evoke Music	57	1	2	20
Aiva	5	0	0	1
Ecrett Music	11	0	1	6
Beatoven	16	0	0	6
MusicStar	9	0	0	1

The analysis of the matching items between these sets showed a low level of agreement and poor matching between the entire derived emotional vocabulary and the representative emotional vocabulary.

5 Conclusions and Limitations

This study confirmed through experiments that the emotional vocabulary used in generative AI-based music production systems are not user-centered. The results are as follows: first, it was found that the emotional vocabulary used in most generative AI-based music production systems differs significantly from the emotional vocabulary used by actual users to express the emotions of music. This discrepancy is a major cause of the system's inability to accurately capture and reflect users' intentions and emotions, resulting in poor user experience. Second, the potential to design user-centered emotional vocabulary was discovered. The experiment conducted with the same emotional vocabulary yielded many common emotional labels and commonly used emotional vocabulary with similar meanings.

This study has several limitations. The sample size of the user group participating in the study is small and not extensive, limiting the generalizability of the research

outcomes. Specifically, the individual musical diversity and personal experiences of the participants were not adequately reflected. Additionally, this study focused on a few specific generative AI-based music production systems and did not address all available options, which is somewhat limiting. This suggests that the study results may primarily apply to the tools examined, and different outcomes may be observed with other generative AI-based systems. Furthermore, the detailed relationships between emotional domains arising from musical elements such as harmony, melody, timbre, and chord progression and the emotional vocabulary were not sufficiently considered. This indicates that there was a limitation in precisely analyzing the emotional impact of musical elements on users.

This foundational study demonstrates the need for more user-centered emotional vocabulary development of in the interfaces of generative AI-based music production systems. When interpreting the results of this study, it is necessary to consider several limitations. Future research should focus on specific measures to design user-centered emotional vocabulary and practical methods for their application.

References

1. Author, F: Article title. Journal 2(5), 99–110 (2016). Roberto Gozalo-Brizuela, Eduardo C. Garrido-Merchán.: A survey of Generative AI Applications. *arXiv preprint arXiv:2306.02781v2* (2023).
2. Miguel Civit, Javier Civit-Masot, Francisco Cuadrado, Maria J. Escalona: A systematic review of artificial intelligence-based music generation: Scope, applications, and future trends. *Expert Systems With Applications*, 209, 118190 (2022). DOI: 10.1016/j.eswa.2022.118190
3. Yang, B., Wei, L., Pu, Z.: Measuring and Improving User Experience Through Artificial Intelligence-Aided Design. *Frontiers in Psychology*, 11:595374 (2020). DOI: 10.3389/fpsyg.2020.595374
4. Maddalena Torricelli, Mauro Martino, Andrea Baronchelli, Luca Maria Aiello: The role of interface design on prompt-mediated creativity in Generative AI. In: *ACM Web Science Conference (WebSci '24)*, May 21–24, 2024, Stuttgart, Germany. ACM, New York, NY, USA, pp. 1-6. DOI: 10.1145/3614419.3644000
5. Interaction Design Foundation - IxDF: What is User Interface (UI) Design? *Interaction Design Foundation - IxDF*. https://www.interaction-design.org/literature/topics/ui-design, last accessed 2024/05/29.
6. Sangwon Lee, Jin Yan: The Potential of a Text-Based Interface as a Design Medium: An Experiment in a Computer Animation Environment. *Interacting with Computers* 28(1), 85–101 (2016). DOI: 10.1093/iwc/iwv045
7. Yueyue Zhu, Jared Baca, Banafsheh Rekabdar, Reza Rawassizadeh: A Survey of AI Music Generation Tools and Models. *arXiv preprint arXiv:2308.12982v1* (2023).
8. Juslin, P. N., Sloboda, J. A.: Music and emotion: Theory and research. 1st edn. *Oxford University Press*, Oxford (2001).
9. Soundraw: https://soundraw.io, last accessed 2024/05/30.
10. Mubert: https://mubert.com, last accessed 2024/05/30.
11. Evoke Music: https://evokemusic.ai, last accessed 2024/05/30.
12. Aiva: https://www.aiva.ai, last accessed 2024/05/30.
13. Ecrett Music: https://ecrettmusic.com, last accessed 2024/05/30.
14. Beatoven: https://www.beatoven.ai, last accessed 2024/05/30.

15. MusicStar: https://www.musicstar.ai, last accessed 2024/05/30.
16. Adyasha Dash, Kat R. Agres: AI-Based Affective Music Generation Systems: A Review of Methods, and Challenges. In: *Proceedings of the XX International Conference*, ACM, New York, NY, USA, pp. 1-26 (2022)
17. Duncan Williams, Alexis Kirke, Eduardo R. Miranda, Etienne Roesch, Ian Daly, Slawomir Nasuto: Investigating affect in algorithmic composition systems. *Psychology of Music*, 43(6), 831–854 (2015). DOI: 10.1177/0305735614543282
18. Lucas N. Ferreira, Jim Whitehead: Learning to Generate Music with Sentiment. In: *20th International Society for Music Information Retrieval Conference (ISMIR 2019)*, Delft, The Netherlands, pp. 384-390 (2019).
19. Kaitong Zheng, Ruijie Meng, Chengshi Zheng, Xiaodong Li, Jinqiu Sang, Juanjuan Cai, Jie Wang, Xiao Wang: EmotionBox: A music-element-driven emotional music generation system based on music psychology. *Frontiers in Psychology*, 13:841926 (2022). DOI: 10.3389/fpsyg.2022.841926
20. Kun Zhao, Siqi Li, Juanjuan Cai, Hui Wang, Jingling Wang: An Emotional Symbolic Music Generation System based on LSTM Networks. In: *2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC 2019)*, IEEE, New York, NY, USA, pp. 2039-2043 (2019). DOI: 10.1109/ITNEC.2019.8729352