Audio Engineering Society

Convention Express
Paper 294

Presented at the AES 157th Convention
2024 October 8-10, New York, NY, USA

This Express Paper was selected on the basis of a submitted synopsis that has been peer-reviewed by at least two qualified anonymous reviewers. The complete manuscript was not peer reviewed. This Express Paper has been reproduced from the author's advance manuscript without editing, corrections, or consideration by the Review Board. The AES takes no responsibility for the contents. This paper is available in the AES E-Library (http://www.aes.org/e-lib) all rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.

Subjective Evaluation of Emotions in Music Generated by Artificial Intelligence

Holt Stairs[1], Wesley Bulla[1], and Song Hui Chon[1]

[1] *M.S. Audio Engineering, Mike Curb College of Entertainment and Music Business, Belmont University, Nashville, TN 37212*

Correspondence should be addressed to Holt Stairs (holt.stairs@gmail.com)

ABSTRACT

Artificial Intelligence (AI) models offer consumers a resource that will generate music in a variety of genres and with a range of emotions with only a text prompt. However, emotion is a complex human phenomenon which becomes even more complex when attempted to convey through music. There is limited research assessing AI's capability to generate music with emotion. Utilizing specified target emotions this study examined the validity of these emotions as expressed in AI-generated musical samples. Seven audio engineering graduate students listened to 144 AI-generated musical examples with 16 emotions in three genres and reported their impression of the most appropriate emotion for each stimulus. Using Cohen's kappa minimal agreement was found between subjects and AI. Results suggest that generating music with a specific emotion is still challenging for AI. Additionally, the AI model here appeared to operate with a predetermined group of musical samples linked to similar emotions. Discussion includes how this rapidly changing technology might be better studied in the future.

1 Introduction

The rate of improvement in Artificial Intelligence (AI) continues to advance in many fields, including music [1]. The use of machine learning has advanced methods to automate complex production processes [2] and offers ways to assist with the expression of creativity in musical composition [3] [4]. Increases in computing power and the size and quality of training data have led to a new rush to produce better and more sophisticated AI models [5]. Several AI models are already available that claim to deliver complete songs that will convey any emotion specified by the user [6] [7]. Assertions include the range of human emotions described by psychologists such as happiness, sadness, fear, anger, surprise, and disgust [8, 9] along with a host of other degrees or gradients of one or more of those emotions [10]. However, because the deployment of AI has advanced so rapidly, there is limited research [11] testing vendors' claims of AI's ability to compose music with an intended emotional response.

This study focused on evaluating a currently available AI model's capability to generate a musical segment with a specific emotion and whether or not participants would agree with the target emotion.

2 Methods

The AI vendor being evaluated offered a choice among 16 so-called emotions including, "sad, calm, motivational, happy, scary, cheerful, angry, triumphant, relaxing, depressing, dreamy, inspirational, tense, joyful, fearful, and energetic," as shown in Figure 1. Stimuli were three groups of 48 unique samples each generated using the different emotional categories in three musical genres (Rock, Electronic, and Pop). The first two sets included 96 samples generated from the AI model. The third set was curated to only include samples that passed a

preliminary assessment (by the first author) of relevance to the given keyword prompt of emotion. This preliminary assessment was not performed on the first two sets.

To evaluate if the AI model was successful in generating music with a specific target emotion, a listening experiment was carried out with seven audio engineering graduate students (one female and six males). They had finished one semester of grad-level critical listening; therefore, they would be considered between "novice" and "expert" in their listening skills. Participants listened to each of 30-second samples and were asked to choose one out of the 16 listed emotions that they thought or felt should be associated with the sample. The experiment was conducted in a group setting in an acoustically treated classroom with professional stereo loudspeakers. Subjects were asked if the initial listening level was appropriate but were not permitted to adjust the loudspeaker volume individually. Subjects could request to repeat samples if desired. Each set of 48 stimuli was randomized in order, which was presented in three evaluations (eight minute long). There was a short a 1-minute break after each eight-minute evaluation to minimize possible fatigue. Each set of 48 unique samples were evaluated twice by the subjects.

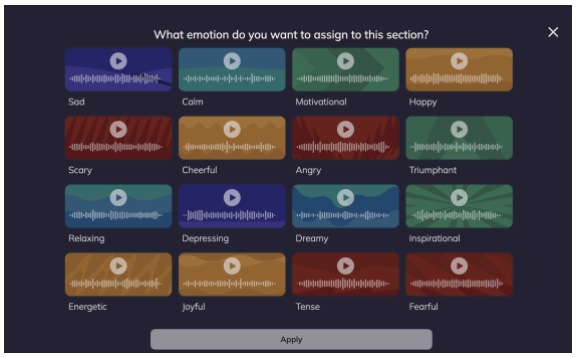


Figure 1. Screenshot of Beatoven.AI consumer interface, which was used to generate stimuli

3 Results

The agreement between the AI and human participants was measured using *Cohen's kappa*. Hits were counted where AI and human participants agreed on the expressed emotion in a particular sample. If human subjects thought a sample was happy but it was generated with a target emotion of "joyful" that would count as non-agreement. However, when samples are grouped by similar emotions, if human participants selected "happy" and the AI was targeting a grouped emotion like "joyful," this would be an example of agreement.

Results found the agreement of the participants with the emotion the AI was attempting to generate was extremely low (6.5%) with a kappa indication of "none to minimal agreement" (κ = 0.17) for the two non-curated sets. The curated set showed increased the hit rate to 15%, a kappa of κ = 0.22 (minimal agreement).

The 16 emotions were further categorized into seven groups based on lexical similarity: (angry, tense), (calm, dreamy, relaxing), (cheerful, happy, joyful), (depressing, sad), (energetic, inspirational), (fearful, scary), (motivational, triumphant). The AI's target emotion for a stimulus was put into its appropriate group and the subjects' responses were also grouped in the same way. Agreement was then calculated between the grouped target emotion and the grouped subjects' responses. This grouping into seven categories still showed minimal agreement with a κ = 0.35.

Cohen's kappa (κ)	Set 1	Set 2	Set 3
Min	0.00	0.00	0.00
Max	0.66	1.00	1.00
Mean	0.17	0.18	0.22
Min (grouped)	0.07	0.07	0.00
Max (grouped)	1.00	1.00	1.00
Mean (grouped)	0.35	0.31	0.35

Table 1. Agreement between AI and human subjects for each set of stimuli (in Cohen's kappa)

Table 1 presents the average, minimum, and maximum agreements between AI and human subjects for each set of stimuli, first without then with grouping of emotions. The agreements ranged from 0% (no agreement) to 100% (perfect agreement) with seemingly no correlation to a particular emotion or genre. Without grouping of emotions, Set 3 showed slightly higher average agreement than the other two sets. However, this difference disappeared when emotions were grouped.

Participants occasionally had a strong agreement (100%) with one another but not with the AI. For instance, all participants responded to one particular sample as conveying the "angry" emotion when it was generated with the keyword prompt of "fearful."

When analysing data by a single emotion or genre, no significant agreement was found between participants' responses and the input emotion to the AI model.

4 Discussion

The goal of this research was to evaluate the current state of AI models' ability to generate music with specific emotions. Results show that the particular AI model tested was not able to generate music with emotional cues with which listeners could reliably agree. Additionally, there were some stimuli generated in an unexpected way. For example, a prompt for a "scary rock track" resulted in music including a ukulele in its instrumentation with a lot of dissonance and a very awkward syncopated drum beat. Removing outlier samples such as this resulted in slightly higher agreement in the third set, although it still failed to achieve any statistical significance.

Listening to the samples in one category also revealed that the model would repeat certain features and even generate near identical samples for two different emotions. These patterns suggest that this particular AI model may use an internal grouping of emotions. It appears that this AI engine does not generate an entirely unique song, but instead attaches pre-generated elements together and mixes them in different ways. Therefore, human intervention may still be required to best utilize AI compositions.

The results from this study should be interpreted with caution due to the limitation of using only one AI model and a small subject size. Future studies should compare multiple AI models to get a broader understanding of the overall current state of AI and music generation. This field is growing so fast that now, a few months after this current study, there are services like [12] that offer more advanced text to music AI programming options than what was possible in the AI engine that was used in this research. A larger pool of qualified listeners is also required in future studies. Additionally, it may be beneficial to conduct research with the nature of the AI composer hidden to examine if the identity of the composer reflects bias against non-human artistic composition [13].

5 Conclusions

This study examined AI-generated music with 16 specific emotions in three genres but failed to find any significant agreement between human listeners and the target emotions. Even grouping lexically similar emotions together did not find any significant agreements between AI and human subjects' responses. Emotion recognition in music created by humans [14] [15] is not necessarily a straightforward task due to its complexity, which may be more difficult for AI at least currently. Given the rate of change in AI performance and the amount of new AI music tools, there are newer and better AI engines than what was used in this study. Therefore, this area of study needs to be constantly evaluated to better understand the capabilities and limitations of this changing technology.

References

[1] V. C. Muller and N. Bostrom, "Future Progress in Artificial Intelligence: A Survey of Expert Opinion," *Fundamental Issues of Artificial Intelligence*, vol. 376, pp. 555-572, 2016, doi: http://doi.org/10.1007/978-3-319-26485-1_33.

[2] Jean-Pierre Briot, Gaëtan Hadjeres, and François-David Pachet, *Deep Learning Techniques for Music Generation*. Springer, 2019.

[3] S. Knotts and N. Collins, "A Survey on the Uptake of Music AI Software," 2020. Available: http://nime.org/proceedings/2020/nime2020_paper95.pdf

[4] C. Roads, "Artificial Intelligence and Music," *Computer Music Journal*, vol. 4, no. 2, p. 13, 1980, doi: https://doi.org/10.2307/3680079.

[5] Epoch (2024), "Datapoints used to train notable artificial intelligence systems," *Our World in Data*. http://ourworldindata.org/grapher/artificial-intelligence-number-training-datapoints (accessed Sept. 5, 2024).

[6] "Royalty-Free, AI Music Generator | Beatoven.ai," *www.beatoven.ai*. https://www.beatoven.ai/ (accessed Aug. 7, 2024).

[7] "Soundraw," *soundraw.io*. https://soundraw.io/ (accessed Aug. 15, 2024).

[8] James Madison University, "About Emotions." Available: https://www.jmu.edu/counselingctr/files/About%20Emotions.pdf (accessed Aug. 27, 2024).

[9] UWA, "Our Basic Emotions Infographic | List of Human Emotions," UWA Online, May 17, 2019. https://online.uwa.edu/infographics/basic-emotions/ (accessed Aug. 27, 2024).

[10] C. Bao and Q. Sun, "Generating Music with Emotions," *IEEE Transactions on Multimedia*, pp. 1–1, 2022, doi: https://doi.org/10.1109/tmm.2022.3163543. (accessed Aug. 28, 2024).

[11] A. Dash and K. Agres, "AI-Based Affective Music Generation Systems: A Review of Methods and Challenges," *ACM computing surveys*, vol. 56, no. 11, , doi: https://doi.org/10.1145/3672554.

[12] "Udio | AI Music Generator - Official Website," Udio. https://www.udio.com/ (accessed Apr. 24, 2024). (accessed Aug. 15, 2024).

[13] H. Chu, H. Lim, J. Lee, J. Kim, H. Lee, T. Kim, S. Kim, S. Jin, and S. Ko, "An Empirical Study on How People Perceive AI-generated Music," *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, Oct. 2022, doi: https://doi.org/10.1145/3511808.3557235.

[14] T. Li and M. Ogihara , "Detecting Emotion in Music," *Proceedings of the 4th International Symposium on Music Information Retrieval* (ISMIR), Oct. 2003, Accessed: Sept. 05, 2024. [Online]. Available: http://jhir.library.jhu.edu/handle/1774.2/41

[15] P. Juslin, "Communicating emotion in music performance: A review and a theoretical framework," In *Music and emotion: Theory and research*, P.N. Juslin & J.A. Sloboda, Eds., Oxford University Press, 2001, pp. 309–337.