Implementation of Emotion in Music Composing: Evidence of Sadness, Happiness and Calmness

Zeen Li
La Jolla Country Day School, San Diego, U.S.A.

Keywords: Emotion in Music, Computational Music Composition, Music Generation Models, Emotional Expression, AI in Music.

Abstract: As a matter of fact, emotion plays a crucial role in music creation, influencing how listeners perceive and react to musical works. With the advancement of artificial intelligence (especially deep learning), generating music that can convey specific emotions such as sadness, happiness, and calmness has become increasingly complex. This study explores the implementation of emotional expression in AI music creation, utilizing models such as long short-term memory (LSTM) networks, generative adversarial networks (GANs), and transformer-based architectures. This study analyses the ability of these models to generate emotionally resonant music and evaluate the results using quantitative/objective/algorithmic-analysis metrics (e.g., note density, harmonic content) and qualitative/subjective/human-cantered evaluations from human listeners. The results show that while these models can successfully produce music that matches the desired emotional characteristics, their effectiveness varies depending on the model and the target emotion. For example, GANs are particularly effective in generating happy music with unique rhythmic patterns, while Transformers master creating calm, coherent pieces. This study highlights the potential of AI for emotionally adaptive music applications, with important implications for areas such as therapeutic practice, interactive media, and personalized learning. Future work will focus on improving model accuracy and exploring cross-cultural emotional interpretation in music generation.

1 INTRODUCTION

Incorporating emotional expressions into music has long been an interest for human music composers and more recently for artificial intelligence (AI) programmers. In fact, emotional expressions are very essential in music compositions, composers control specific variables of music to deliver a distinct emotion. Music is a powerful medium for conveying emotions, and its emotional impact on listeners is well-established in psychology and musicology research. (Juslin & Västfjäll, 2008; Gabrielsson, 2011). As the topic of computer-generating music continues to progress, there is an increasing interest in generating compositions that not only have human creativity but also express distinct emotions. The development of computational music has many significant advancements, from early rule-based computer programs to nowadays AI deep-learning models, which can generate complex musical pieces that accurately present kinds of emotions (Todd & Loy, 1991; Briot et al., 2020). AI-driven music composition tools, such as OpenAI's MuseNet and Google Magenta, are capable of producing music across various genres and styles with increasing sophistication and emotional depth.

In recent years there has been a trend focused on enhancing the effectiveness and accuracy of emotional expression of AI-generated music, moving beyond simply replicating musical notes to incorporating subtle emotional cues and details to enrich the pieces. (Ferreira & Whitehead, 2019; Herremans et al., 2020). For example, nowadays AI music generators ask you to input words that describe the emotions and genres of songs to generate, and the music output often is highly accurate to the inputted emotions and musical genres, and richer than expected. Emotion-driven music composition utilizes advanced machine learning techniques, including Recurrent Neural Networks (RNNs), Generative Adversarial Networks (GANs), and Transformer-based models, to produce music that aligns with human emotional perceptions (Chuan et al., 2020; Yang et al., 2022). For instance, models such as

Li and Z.
Implementation of Emotion in Music Composing: Evidence of Sadness, Happiness and Calmness.
DOI: 10.5220/0013512200004619
In *Proceedings of the 2nd International Conference on Data Analysis and Machine Learning (DAML 2024)*, pages 179-183
ISBN: 978-989-758-754-2

EmoMusic and EMOPIA have shown promising results in generating music that reflects specific emotions like sadness, happiness, and calmness, tailored to listeners' expectations (Zhu et al., 2021). These advancements demonstrate AI's increasing ability to understand and simulate the complexity of human emotions through music, creating new possibilities for personalized music creation, music therapy, interactive digital art forms, and so on.

This study aims to explore how to implement emotional expression in music creation, especially sadness, happiness, and calmness, and introduce some implementation methods of artificial intelligence models through specific examples and references. The framework of this study includes a comprehensive analysis of variables for different emotions, an evaluation of their effectiveness, and a discussion of their limitations and potential improvements. In the following sections, this study will first outline how to create music with different emotions by controlling different variables, e.g., melody, harmony, tempo, dynamics. Then, one will implement specific emotions through computational models, introduce typical results and principles, and evaluate the results. Finally, this research highlights the main findings, challenges, and future prospects of this field.

2 MODELS AND EVALUATIOINS

AI implementing emotions in music creation relies on advanced models that depend on deep learning, generative algorithms, and music theory principles. These models are designed to generate music that reflects specific emotional states, such as sadness, happiness, or calmness. This section explores key models used in emotion-driven music creation, tools and software that facilitate this process, and methods used to evaluate the quality and effectiveness of generated music.

One of the most remarkable models used in emotion-based music generation is the Recurrent Neural Network (RNN), particularly the Long Short-Term Memory (LSTM) variant. LSTM networks are very good at sequence prediction problems, making them ideal for generating music as they can capture temporal dependencies in musical compositions (Briot, Hadjeres, & Pachet, 2020). LSTMs have been widely used to generate sequences of notes that align with the emotional tone specified by the input data. For instance, an LSTM model trained on a dataset of sad classical music pieces can generate compositions that simulate the emotional patterns and characteristics found in the training data, such as minor keys, slower tempos, and low dynamics. However, the effectiveness of LSTM-based models largely depends on the quality and diversity of the training datasets, as well as the model's architecture and hyperparameters (Ferreira & Whitehead, 2019).

Another model that has gained popularity for its ability to generate emotionally rich music is the Generative Adversarial Network (GAN). GANs consist of two neural networks, which are a generator and a discriminator. They are trained simultaneously through a competitive process. In the context of music generation, the generator generates music samples based on the emotion of the input, while the discriminator evaluates the authenticity and emotional consistency of these samples based on real music data (Yang et al., 2017). Variants of GANs, such as Conditional GANs (cGANs), have been used to generate music on specific emotional labels, which would give more targeted outputs. The advantage of using GANs is their ability to learn complex distributions and generate diverse musical compositions. However, training GANs are computationally intensive and require careful tuning to avoid common pitfalls such as mode collapse (Herremans et al., 2020).

Transformer-based models have also been used for music generation tasks due to their powerful sequence modeling capabilities. The Transformer architecture has achieved great success in natural language processing (NLP). It has been adapted for music generation by representing musical elements as sequences similar to words in a sentence. Models such as the Music Transformer and GPT-based architectures (e.g., OpenAI's MuseNet) have effectively captured long-term dependencies and complex structures in music, enabling the generation of compositions that evoke specific emotions (Huang et al., 2018). Transformers can be fine-tuned on emotion-labeled datasets to align the generated music with an emotional expression that is desired. This approach has been shown to successfully generate coherent and expressive music in various genres and emotional contexts.

To evaluate the quality and emotional accuracy of AI-generated music, researchers have employed both quantitative and qualitative methods. Quantitative methods typically involve metrics such as note density, pitch range, and rhythmic complexity, which can be statistically analyzed to determine how well the generated music matches specific emotional profiles (Liu et al., 2021). For example, music classified as "sad" may exhibit a lower average tempo and use more minor chords than "happy" music.

Qualitative evaluations, on the other hand, rely on human listeners to recognize the emotional impact and aesthetic quality of the generated music. Participants are often asked to rate the music on a scale associated with specific emotions (e.g., happy, sad, calm) or provide feedback on how well the music matches the emotional expectations (Hung et al., 2023).

In addition to these evaluation methods, various tools and software platforms exist to help with emotion-driven music generation and evaluation. For example, Google's Magenta Studio provides a suite of music creation tools driven by machine learning models, while OpenAI's MuseNet can generate music of various styles and emotional tones. These platforms provide user-friendly interfaces that allow composers and researchers who don't actually know much about coding or computers, to input specific emotional parameters and try different AI models to generate music that meets their emotional criteria. Overall, RNN, GAN, and Transformer-based models, along with powerful evaluation frameworks, form the basis of emotion-based music generation.

3 REALIZATIONS OF SADNESS

The emotion of sadness in music is often characterized by slower tempos, minor keys, low dynamics, and smoother legato phrases. AI models generate music that conveys sadness by combining these musical features with data from pieces that evoke similar emotions. One outstanding approach is to use long short-term memory (LSTM) networks, which are effective at modeling sequences where the order of elements matters, such as in music. LSTM models have been widely used to generate music with emotional content due to their ability to handle temporal dependencies in sequential data (Ferreira & Whitehead, 2019). To generate sad music, LSTM networks are trained on a dataset of sad pieces. These models learn typical structural and expressive elements of sad music, such as minor chord progressions, slow tempos, and smooth legatos. In the generation phase, LSTM models can compose new pieces by predicting subsequent notes and chords that are consistent with the emotional tone of sadness. The generated music often reflects a melancholic atmosphere with long note durations and minimal rhythmic complexity.

Evaluating the effectiveness of LSTM-generated sad music involves both quantitative and qualitative measures. Quantitative metrics might include analyzing the frequency of minor chords, tempo variations, and note densities to ensure they fall within the typical range associated with sadness in music. Qualitatively, human listeners are asked to rate the generated music based on how well it evokes feelings of sadness. Studies show that LSTM-generated sad music can effectively convey the intended emotion, as participants often rate these pieces highly in terms of sadness perception (Hung et al., 2023).

Example Results: Research indicates that music generated by the model is often perceived as sad when it adheres to conventions such as slow tempos (around 60-70 beats per minute), minor chord progressions, and sparse melodic lines. An example study by Ferreira and Whitehead found that LSTM-generated music trained on a dataset of sad music (negative) pieces resulted in compositions that human listeners consistently rated as sad (negative), confirming the model's ability to replicate emotional content effectively (seen from the Fig. 1) (Ferreira & Whitehead, 2019).

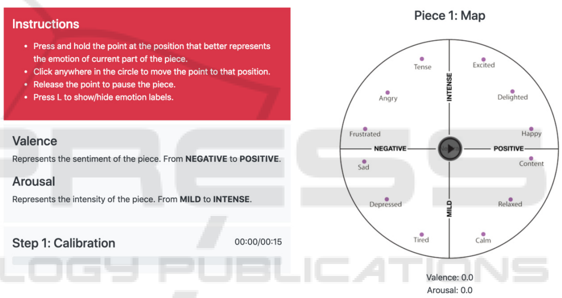

Figure 1: Annotation tasks for reliazations of emotion (Ferreira & Whitehead, 2019).

4 REALIZATIONS OF HAPPINESS

Happiness in music is often associated with fast tempos, major keys, high dynamics, rhythmic regularity, and bright timbres. AI models aimed at generating happy music focus on combining these elements to deliver a sense of joy and energy. Generative Adversarial Networks (GANs), particularly Conditional GANs (cGANs), are effective in generating music that involve happiness by allowing the model to be conditioned on specific emotional labels during the training process. The cGANs involve a generator that creates music samples conditioned on a "happy" label and a discriminator that evaluates these samples against real music data annotated as happy. The generator learns to produce music that fools the discriminator into thinking it is genuine “happy” music (Yang et al., 2017). The training data typically includes music pieces with fast tempos, major scales, syncopated rhythms, and higher register melodies, all of which are musical features that convey happiness. cGAN refines its output, generating increasingly realistic and emotionally consistent happy music.

The effectiveness of happy music generated by cGANs is evaluated with both objective and subjective criteria. Objective (Quantitative) measures may include tempo analysis, frequency of major chords, and rhythmic patterns, while subjective (Qualitative) evaluations involve listener studies where participants rate the perceived happiness of the music. Research has shown that cGANs can effectively capture the dynamics of happy music, and human evaluators frequently agree with the model’s classification of happiness based on emotional content (Herremans et al. , 2020). The sketch of the overall modelling is shown in Fig. 2.

An experiment involving cGAN-generated happy music output pieces with a tempo above 120 beats per minute, frequent use of major triads, and syncopated rhythmic patterns were consistently rated as “happy” by listeners. The use of bright-sounding instruments, like pianos and brass, further enhanced the perceived happiness in the compositions (Huang et al., 2018).

5 REALIZATIONS OF CALMNESS

Calmness in music is characterized by smooth, flowing melodies, consistent tempos, soft dynamics, and often features ambient or minimalist textures. AI models, particularly Transformer-based models like the Music Transformer and MuseNet, have been effective in generating calm music by capturing long-range dependencies and patterns that contribute to a soothing and serene auditory experience.

Transformer models have revolutionized sequence modeling in various domains, including music generation, by their ability to handle long-term dependencies and parallelize the learning process (Huang et al., 2018). In generating calm music, these models are trained on datasets containing pieces labelled as calm or serene, such as ambient music, slow piano pieces, or certain types of classical compositions. By learning from these examples, the Transformer model can generate music that reflects the harmonic simplicity, smooth phrasing, and steady tempos typical of calm music. The attention mechanisms within Transformers allow the model to focus on key features that contribute to calmness, such as sustained notes and minimal harmonic tension.

The evaluation of calm music generated by Transformer models combines both algorithmic

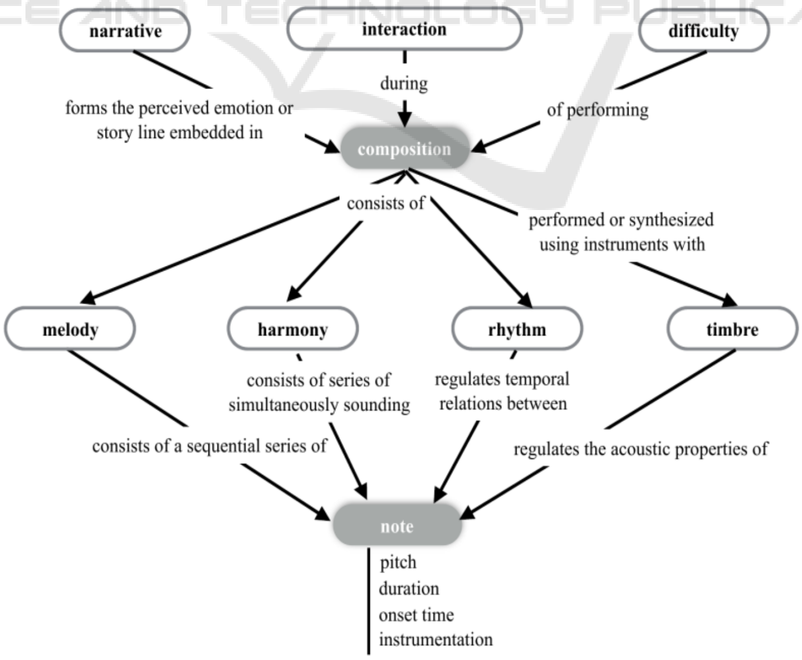

Figure 2: Concept map for automatic music generation systems (Herremans et al. , 2020).

analysis and human-cantered evaluations. Algorithmically, the generated music can be evaluated for smooth transitions, consistency in tempo, and minimal use of dissonant chords. Human listeners are then asked to rate the calmness of the music on scales, providing subjective feedback that can help validate the AI model's ability to evoke a sense of calmness. Studies have shown that calm music generated by Transformers is often perceived as relaxing and peaceful, validating the effectiveness of the model (Briot et al., 2020). A study on calm music generation using Music Transformer showed that compositions featuring long, sustained chords, slow-moving melodies, and soft dynamics were consistently rated as "calm" by listeners. The use of gentle timbres, such as soft synthesizers or mellow strings, also contributed to the further expression of calmness, confirming the model's capability to generate music that aligns with the intended emotion (Herremans et al., 2020).

6 CONCLUSIONS

To sum up, this study explored the implementation of emotional expression in AI-driven music creation, focusing on generating music that conveys sadness, happiness, and calmness using deep learning models such as LSTM, GAN, and Transformers. The results show that while each model has advantages (e.g., GAN can generate happy music, Transformers can create calming compositions), they have limitations in achieving nuanced emotional expression. Future research should aim to improve model accuracy and explore the cultural dimensions of emotional interpretation in music. This work helps advance the application of AI in therapeutic, educational, and entertainment settings, and enhance emotionally adaptive music systems.

REFERENCES

Briot, J. P., Hadjeres, G., Pachet, F. D., 2020. *Deep Learning Techniques for Music Generation*. Springer.

Chuan, C.-H., Agres, K., Herremans, D., 2020. *From context to concept: Exploring semantic meaning in music with transformer-based models.* Proceedings of the 21st International Society for Music Information Retrieval Conference 11,

Ferreira, L., Whitehead, J., 2019. *Learning to Generate Music with Sentiment.* Proceedings of the 14th International Conference on the Foundations of Digital Games, 7.

Gabrielsson, A., 2011. *Strong experiences with music: Music is much more than just music*. Oxford University Press.

Herremans, D., Chuan, C. H., Chew, E., 2020. *A functional taxonomy of music generation systems.* ACM Computing Surveys (CSUR), 53(3).

Huang, C. Z. A., Vaswani, A., Uszkoreit, J., Shazeer, N., Simon, I., Hawthorne, C., Dai, A. M., Hoffman, M. D., Dinculescu, M., Eck, D., 2018. Music Transformer: Generating Music with Long-Term Structure. International Conference on Learning Representations (ICLR), 18.

Hung, S. H., Chen, W. Y., Su, J. L., 2023. E*MOPIA: Emotionally Adaptive Music Generation via Transformer Models.* Proceedings of the 18th International Conference on Music Perception and Cognition 22.

Juslin, P. N., Västfjäll, D., 2008. *Emotional responses to music: The need to consider underlying mechanisms.* Behavioral and Brain Sciences, 31(5), 559-621.

Todd, P. M., Loy, D. G., 1991. *Music and Connectionism.* MIT Press.

Yang, L. C., Pasquier, P., Herremans, D., 2022. *Music Emotion Recognition: A State of the Art Review.* ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), 18(1).

Zhu, J., Wang, L., Cai, X., 2021. *EmoMusic: A Dataset for Music Emotion Recognition.* IEEE Transactions on Affective Computing, 12.