Journal of Research High School

Vol. 2025 (5) 393 – 401
ISSN 2688-3651 [online]

A Systematic Review of Artificial Intelligence Models Used in Affective Music Generation

Aiden Hong[1] *

[1]Bergen County Academies, Hackensack, NJ, USA
*Corresponding Author: aidenhong08@gmail.com

Advisor: Ivy Wang, ivywan@bergen.org

Received November 12, 2024; Revised March 6, 2025; Accepted March 18, 2025

Abstract

Affective Music Generation (AMG) is a type of music generation that incorporates specific emotions into musical compositions. The challenge of accurately integrating emotional expression into AMG remains underexplored, particularly in the evaluation of model performance across various architectures. Given the growing role of affective music in entertainment and therapy, this review's objective was to evaluate current AI models to identify the most effective architectures for emotional accuracy. This study covered foundational AMG concepts and models such as Russell's Circumplex Model of Emotion, long short-term memory (LSTM) models, and transformer models. The current state of affective music generation is reviewed by comparing machine learning model architectures and their emotional accuracy. After applying the inclusion and exclusion criteria, fifteen studies were selected through a systematic database search of Google Scholar and PubMed. The methodology followed PRISMA guidelines for a systematic review. Key findings through a narrative analysis of the studies showed that deep learning models, particularly transformers, are the most common approach for AMG due to their ability to handle sequential data efficiently. In addition, Russell's circumplex model of emotion was often used for emotional representation and quantification.

Keywords: Systematic Review, Affective Music Generation, Emotion, Transformers, Russell's Circumplex Model of Affect, Long Short-Term Memory, Music Generation

1. Introduction

Music generation is a relatively new field that involves rule-based or model-based music composition. It is the process of automating music creation by incorporating aspects of music theory and musical features such as tempo, rhythm, and key into a generation algorithm or model. With enough development in the field, music generation has the potential to become a handy tool in entertainment, media, and therapy. Affective music generation (AMG) is music generation that incorporates a specified emotion into the melody. Affective music generation can help induce positive moods in an individual and be used for therapeutic purposes. Generated music can also improve the conditions of patients with neurodegenerative diseases. Practical applications of AMG include AI-composed soundtracks for gaming or movies, promotion of physical activity during rehabilitation, and mood improvement for mental disorders (Dash and Agres, 2024). Given AMG's multitude of benefits, it is beneficial to review the current state of research in this field and set a standard for the most effective architectures. This study aimed to systematically evaluate AI models used in AMG, comparing each model's emotional accuracy and general performance. A key gap identified is the lack of standardized evaluation metrics, which this paper sought to address.

Past studies have reviewed the state of affective music generation by analyzing different approaches and methods of state-of-the-art systems. For example, Wiafe and Franti (2023) reviewed affective algorithmic composition by exploring common techniques and methods such as emotion models, machine learning techniques, and composition

Vol. 2025 (5) 393 – 401
ISSN 2688-3651 [online]

approach. More recently, Dash and Agres (2024) have investigated progress, conventional models, and challenges in affective music generation systems. However, there is a need for more thorough analysis on the quantifiable results of state-of-the-art AMG systems (i.e., comparing emotional accuracy between two models). More research is needed in the field, and the inherent difficulty of comparing complex models with differing architectures makes data-supported conclusions scarce. To overcome a lack of data-supported findings, this study aimed to review current optimal models based on the statistical results of previous AMG studies. This work provided numerical results in model performance by grouping studies by their evaluation method (subjective, objective, etc.) and comparing their performance values.

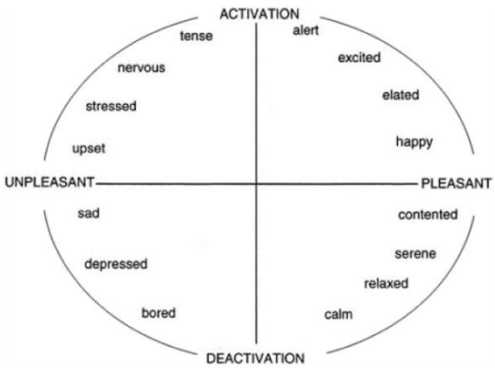

Figure 1. Graphical representation of Russell's circumplex model of affect (Posner et al., 2005)

2. Theory

AMG systems often employ machine learning and emotion models to interpret emotion and incorporate it into music. Specifically, Russell's circumplex model of affect and neural networks (long short-term memory, transformer) are key components. LSTM and transformers are more prevalent in AMG as compared to other models due to their ability to handle the sequential data structure of music. Transformers have shown to be especially effective in maintaining long term structure in its output (Dash and Agres, 2024). It is therefore important to understand the underlying structure of these models.

2.1 Russell's Circumplex Model of Affect

Russell's circumplex model of affect is a widely used format for emotion representation in affective music generation systems. It is an alternative to discrete labeled emotions and represents emotions in a 2-D continuous plane (i.e., maps all emotions to a pair of values rather than a categorical label). The circumplex model of emotion proposes that emotional states are a product of two independent neurological systems. This model of emotion contradicts basic emotion models, which assume that a discrete and independent neural system can define each emotion. The axes on this plane are comprised of a horizontal valence value, which describes the pleasant/unpleasantness of a given emotion, and a vertical arousal value, which represents the intensity of the emotion. These values range from 0 to 1, with 0 being the least intense or pleasant and 1 being the most. The transition from discrete to continuous allows for more precise emotion inputs for affective models (Russell, 1980). Though commonly used in AMG, this emotion model has limitations in its ability to fully capture the complexity of individual emotion, as it assumes a relatively simple, linear relationship between valence and arousal. Other alternatives, such as pleasure-arousal-dominance, use higher dimensionality (pleasure-displeasure, arousal-nonarousal, dominance-submissiveness) than Russell's 2D model, which can be more effective in representing complex emotions (Mehrabian, 1996).

2.2 Long Short-Term Memory

LSTM is a recurrent neural network model with frequent applications in affective music generation. Its design is meant to prevent the vanishing/exploding gradient problem, which is especially prevalent in RNNs during backpropagation over time due to their feedback loop. LSTMs handle sequential data, with inputs at each point in time. At time t, an external input is given, and two inputs are received from the outputs of the LSTM at time t-1, called the hidden state and the cell state (short and long-term memory). These inputs are generally processed through 3 main gates: the forget gate, the input gate, and the output gate. The forget gate changes the cell state and determines how much the current cell state should be kept, generally through a sigmoid activation function. Depending on the external input, the input gate adds memory to the newly changed cell state, and the output gate determines the new hidden state.

Vol. 2025 (5) 393 – 401
ISSN 2688-3651 [online]

LSTM's ability to handle long-term dependencies in sequential data makes it extremely useful for applications such as AMG and machine translation. However, it is limited in its high computational costs and difficulty with parallel processing.

2.3 Transformer

A transformer is a neural network that uses an attention mechanism to understand contextual relationships between elements. Its ability to use attention values allows it to process sequential data parallelly, unlike typical RNNs. Transformers keep track of the positions of sequential data through positional encoding by modifying the input token depending on its position in the sequence. The key, value, and query vectors consider the element's context through self-attention. The queries, keys, and values are created by changing the positional encoded values through a matrix of set values derived from training. Relationships between different elements are calculated by comparing their keys and queries, often done by getting the dot product of the two vectors. The outputs of these calculations are then put into a softmax operation and multiplied by their respective value vectors to get the self-attention values. The decoder can use these values for encoder-decoder attention and return the desired output. Transformers' ability to process sequences in parallel make them more efficient than traditional RNNs, particularly for larger datasets and sequences. The self-attention mechanism allows transformer models to retain long-range dependencies and understand contextual information, making them suitable for many of the same applications as LSTM (AMG, machine translation, natural language processing). Despite these advantages, transformers face challenges due to their high computational cost and immense data requirements for training.

3. Methodology

3.1 Database/Search Query

The structure of this review is per the guidelines of the PRISMA 2020 statement. A comprehensive search was conducted through Google Scholar and PubMed databases to analyze overarching trends in affective music generation models. These databases were chosen for the literature review due to their abundance of material relating to the topic and their prestige. To maximize efficiency in discovering relevant articles, the databases were searched with the following queries: ("music" AND "emotion" AND "generation), ("music" AND "affective" AND "generation"), ("music" AND "mood" AND "generation"), ("music" AND "emotion" AND "synthesis").

3.2 Inclusion and Exclusion Criteria

From the initial search through both databases, 103 results were found. Note that, for the Google Scholar database, only the top 5 pages of results were considered due to the sheer abundance of search results. Upon filtering out duplicates, 81 articles remained. After reviewing the title and abstract, inclusion criteria were applied to the initial search. These criteria included a) accessible through the web as free access or institutional login under MIT b) assessing the quality/emotional accuracy of an original affective music generation model, "original" meaning created for the article, not a review of an existing model c) written or translated in English. After removing articles that did not fit the inclusion criteria, 52 were left.

The articles remaining after the initial screening process were excluded if they fit the exclusion criteria after a full-text review a) algorithms whose output does not have a melodic component, for example, an affective lyric text generator b) non-peer-reviewed articles such as books and theses c) emotion-based song/playlist recommendation systems. Considering exclusion criteria left 12 articles to be used in the systematic review.

3.3 Data Synthesis

Data regarding purpose, affective computing methods, user experience enhancements, approaches, and quantitative and qualitative results were recorded and compared to gather knowledge on the current state of affective

Vol. 2025 (5) 393 – 401
ISSN 2688-3651 [online]

music generation. If a paper had more than one type of AMG model, only the most successful one was considered (i.e. the highest accuracy and quality). The type of model used was recorded for documents that used machine learning or deep learning; for rule-based algorithms, the specific features that were manipulated were recorded. Finally, only the music generation aspect was considered for papers that combined different systems with AMG (affective music generation), such as image classification and movement tracking.

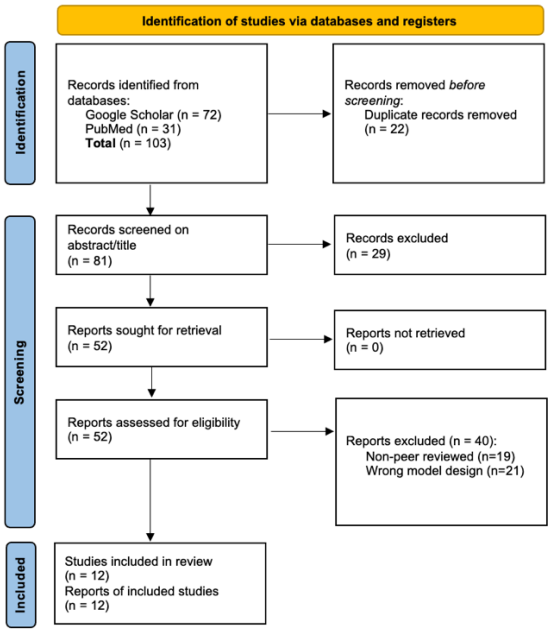

Figure 2. Flowchart of Study Selection Methodology

4. Results

The general findings of the outcomes of each study can be found in Table 1. Nine studies used a deep learning model (i.e. LSTM, transformer, GAN), making it the most commonly used generation form; the remaining studies (n=3) used rule-based algorithms. Three of the deep learning models used a variant of the transformer, making it the most commonly used model among the included studies. The emotion model most frequently used was by far Russell's emotion circumplex (n=10), though the most common emotion input type was nearly evenly split between label-based (n=5) and continuous (n=4).

Table 1. Summary of Included Studies

Author (Year)	Affective Computing Method	Features Used	Emotion Input Type	Emotion Model
Zheng et al. (2022)	GRU	Pitch histogram, note density	Histogram, Note Density	Russell's Emotion Model
Grekow and Dimitrova-Grekow (2021)	VAE	N/A	Label Based, Random Sample	Russell's Emotion Model
Sulun et al. (2022)	Transformer	Note density, tempo, instrument	Continuous Valence/Arousal	Russell's Emotion Model
Zhao et al. (2019)	LSTM	Pitch, rhythm	Label Based	Russell's Emotion Model
Bao and Sun (2022)	Transformer	Syllable, pitch, duration, rest	Label Based	Russell's Emotion Model
Wallis et al. (2008)	Rule-based	Pitch register, loudness, rhythmic roughness, tempo, articulation, upper extensions, harmonic mode	Continuous Valence/Arousal	Russell's Emotion Model
Miyamoto et al. (2020)	Rule-based	Tempo, rhythm, pitch, loudness, mode	Continuous Valence/Arousal	Russell's Emotion Model
Agres et al. (2023)	Rule-based	Tempo, velocity range, velocity variation, instrumentation, rhythm, voice leading, pitch register	Continuous Valence/Arousal	Russell's Emotion Model
Pangetsu and Suyanto (2021)	Transformer	N/A	Label Based	Russell's Emotion Model
Zheng and Li (2024)	GAN	N/A	Sample Piece	N/A
Huang et al. (2021)	GAN	Tempo, rhythm, pitch, loudness, mode	Label Based	Russell's Emotion Model
Madhok et al. (2018)	LSTM	N/A	Label Based	Categorical Model

Vol. 2025 (5) 393 – 401
ISSN 2688-3651 [online]

4.1 Affective Computing Methods

Transformers were a commonly used model for affective music generation due to their ability to compute sequences parallelly. A study by Sulun et al. (2022) proposed a decoder-only transformer system with continuous valence-arousal values and MIDI event tokens. Bao and Sun (2022) used an encoder-decoder GRU and transformer to generate melody and lyrics based on a seed lyric/melody and a modified beam search via an emotion classifier to generate music with a desired emotion. Other studies such as one by Pangetsu and Suyanto (2021) used REMI (REvamped MIDI-derived event) tokens and emotion tokens in a Transformer-XL, with randomly generated chords as input and a predicted sequence of musical events as output.

GANs (Generative Adversarial Networks) are another standard model in affective music generation. For example, Zheng and Li (2024) used GAN combined with self-attention and learning automata to keep track of long-distance dependencies and relationships during generation. The self-attention mechanism also benefited from retaining emotional features from the input and incorporating them into the output. Another example is a study by Huang et al. (2021), which used multiple discriminators and a generator with an emotion encoder. The model's discriminators include two classifiers that constrain the melody's emotional features (valence and arousal) and match it with the given emotion input.

LSTMs are able to maintain long term dependencies over time by preventing the vanishing/exploding gradient problem, making it another suitable model for music generation. Madhok et al. (2018) used a doubly stacked LSTM network that, upon receiving an emotion input translated from an image, selects music pieces corresponding to the emotion from a pre-existing dataset and feeds its one hot representation to the LSTM model.

Rule-based algorithms were also used as an alternative to machine learning models. Wallis et al. (2008) used a direct linear relationship between musical features and valence/arousal values. Valence values were assumed to have a direct relationship with mode and pitch, while arousal values were associated with tempo, articulation, volume, and rhythmic roughness. This simplistic algorithm made large training datasets and training time unnecessary, while lowering computation time. A study by Agres et al. (2023) followed a similar structure, determining tempo, rhythm, and volume from arousal and pitch and mode from valence. Miyamoto et al. (2020) used valence values to probabilistically determine chord progressions, voice leading, instrumentation, and pitch registers. Rhythmic and timbral features such as tempo, note density, and velocity were controlled by arousal.

4.2 Model Evaluation Methods

Model evaluation methods for affective music generation can be categorized into four groups: emotion label comparison, music quality rating, objective, and subjective. Emotion label comparison methods check how accurately generated melodies portray the intended emotion. Music quality rating determines the smoothness and enjoyability of the music without considering affective elements. Objective methods use algorithms to determine the above qualities, while subjective methods use the opinions of human participants.

The reviewed studies used both objective (algorithmic) and subjective (human participants) methods to test the quality and accuracy of their model. The algorithmic evaluation methods mainly consisted of comparing inputs of generated music with the output of a trained emotion classifier/regression model. For subjective evaluation methods, MOS (mean opinion score) and SAM (self-assessment manikin) were commonly used to determine quality and emotional accuracy of generated music. The methods and results for the included articles can be found in Table 2.

Subjective Emotion Accuracy

The subjective evaluation of AMG models is helpful because they do not require the setup of a classifier and allow for necessary human feedback. For example, Zheng et al. (2022) used 26 participants and played 3 generated samples of each of 4 emotions (happy, tensional, sad, peaceful). The participants were then asked to categorize the music out of the same 4 emotions, and the results were compared with the intended emotion. Madhok et al. (2018) used an MOS approach, where 30 participants played 30 melodies and were asked to rate them from 0-10 (sad-neutral-happy). The average of the results was then compared with the emotion rating of an input image and a correlation was

Vol. 2025 (5) 393 – 401
ISSN 2688-3651 [online]

noted. Pangetsu and Suyanto (2021) also used an MOS rating scale of 0-10 (sad-neutral-happy) and recorded the matches between the 3 emotions and the input.

Table 2. Evaluations Methods and Results of Included Studies

Author (Year)	Emotional Accuracy Evaluation Method	Emotion Accuracy Rating
Zheng et al. (2022)	Subjective	(happy, tensional, sad, peaceful) 71%, 74% 56%, 63%
Grekow and Grekow (2021)	Objective	N/A
Sulun et al. (2022)	Objective	(error) 0.1948
Zhao et al. (2019)	Subjective	(happy, tensional, sad, peaceful) 72%, 50% 56%, 62%
Bao and Sun (2022)	Objective - Classifier	(negative, positive) 61.7%, 78.3%
Wallis et al. (2008)	N/A	N/A
Miyamoto et al. (2020)	N/A	N/A
Agres et al. (2023)	Subjective - SAM	(valence & arousal R-squared values) 0.90, 0.96
Pangetsu and Suyanto (2021)	Subjective - MOS	(total accuracy) 70%
Zheng and Li (2024)	N/A	N/A
Huang et al. (2021)	Objective - Classifier	(4-quadrants, valence-half, arousal-half) 39.58%, 58.96%, 63.96%
Madhok et al. (2018)	Subjective - MOS	(correlation) 0.93

SAM was also used to obtain the music's perceived valence and arousal values directly. This scale measures valence and arousal from 1-9 using representative figures to guide the participant towards the most accurate emotion. In a study by Agres et al. (2023), 26 participants recorded the perceived emotion of generated samples using SAM. The results were converted into values between 0 and 1. Linear regression analysis was then used to determine the similarity between the results and the intended emotion.

Subjective Music Quality

Subjective music quality was the most common method used to determine music quality because it is difficult to algorithmically and objectively determine the quality of generated music. All studies that recorded subjective music quality used a form of MOS rating scale to obtain average satisfaction from generated musical sequences. For example, Huang et al. (2021) used two metrics, authenticity and fluency, as values between 1-5 and compared them with both the actual training data and another existing AMG model. Zheng and Li (2024) used 5 music experts and 5 regular participants to rate the quality of generated music on a scale of 1-10. Bao and Sun (2022) also used a scale of 1-5 (very bad-very good) to record melody quality.

Objective Emotion Accuracy

Objective evaluation was performed using algorithmic or statistical methods and helped to determine whether an AMG model incorporated the intended affective musical features. A standard method of objectively evaluating a model was to use a classifier. Huang et al. (2021) proposed splitting classifiers into valence-half, arousal-half, and four-quadrant. These classifiers determined which half of each axis in the circumplex model of emotion (valence and arousal) inputted music was located, and the four-quadrant classifier determined the quadrant the music's emotion belonged to. Similarly, Bao and Sun (2022) incorporated the classifier used in their EBS algorithm and a pre-trained BERT annotator to classify music into 3 labels (positive, negative, and unlabeled).

Rather than comparing categorical data, Sulun et al. (2022) used a regression model. To determine the error, they measured the normalized L1 distance between the inputted valence/arousal values and the regression model's output. This allowed for the precise calculation of the model's emotional accuracy. Regression models are a valuable alternative to classifiers because they allow for direct comparison of continuous values.

In another study, Grekow and Grekow (2021) used a statistical analysis method instead of algorithm-based analysis. The average of 4 metrics (pitch range, # pitches used, pitch in C major rate, pitch in C minor rate) was calculated for 4 distinct emotions of both the generated sequences and the training dataset. The Kolmogorov-Smirnov (KS) statistic was then used to find how similar these metrics were for each emotion group.

Vol. 2025 (5) 393 – 401
ISSN 2688-3651 [online]

5. Discussion

This study reviewed the most up-to-date methods for developing and analyzing affective music generation systems. It compared different algorithms and models in their implementation of affective computing and categorized them based on the results of their evaluation methods.

The review found that deep learning models were the most commonly used approach to generating emotional music. More specifically, the most widely used models handled sequential data efficiently and accurately, such as transformers. Most studies used a dimensional model, specifically Russell's circumplex model of affect, to represent emotion, though some used discrete labeling and one-hot encoding.

Musical features such as tempo, pitch, note density, and mode were manipulated to create the feeling of emotion within the melody. Tempo and note density were found to have a direct correlation with arousal, while valence was related to pitch. Emotional accuracy and music quality values for each paper were determined, and the process in which studies obtained their results was described to give context to each value.

5.1 Barriers in Model Performance

The study found two main barriers in the advancement of AMG technologies – reliance on large datasets, and extensive training times. Transformers were seen to have the best results in emotional accuracy, due to their ability to retain the long term musical relationships necessary to imbue perceivable emotions in music. The transformer is also highly optimized in training time due to its ability to process data in parallel. However, transformer models require large datasets and high computational cost to be effective, which is a significant roadblock due to the lack of usable AMG training data currently available. Besides transformers, the study found that variations of RNNs were also widely used, due to their ability to handle sequential data. The classic RNN was less frequently used due to the vanishing gradient problem, which hindered its ability to handle long term dependencies. LSTMs used memory cells and gates to prevent the vanishing gradient problem, making them nearly as good as transformers in terms of long range dependencies. A limitation of LSTMs were their struggle with training time and inference due to its lack of compatibility with parallelization. To counteract LSTMs slow computational speeds, some studies used GRU, a simpler gating mechanism that could still retain long term dependencies. The simpler gating mechanism, however, struggled with more nuanced emotional expression. With access to considerable computational resources and large datasets, the study found that transformers have the greatest scalability, efficiency, and emotional accuracy for AMG of all models reviewed.

6. Conclusion

A significant limitation across all studies included in the review was the difficulty in objectively comparing different models due to the varying levels of complexity of both the models and their evaluation methods. Due to a lack of research, there is currently no clear standard for affective music generation systems. Therefore, comparing the raw results of models is difficult and inaccurate. To work around this issue, this review categorized studies by their model evaluation method and detailed the specific processes used to obtain a statistical result. Although an objectively optimal model was not possible to determine, this study gave in depth analysis on each AMG model's accuracy in context with how it was evaluated. As a future direction, multiple models considered the current state-of-the-art should be designed and evaluated in the same format. Specifically, this study proposes a two-faceted standardized evaluation metric, using both objective and subjective methods. For the subjective evaluation metric, the study suggests MOS with a range of 0-10, going from sad-neutral-happy. This framework will provide a reliable standardized metric for capturing the spectrum of human emotion perception. The participant sample size should be greater than 30 to avoid statistical outliers and ensure the generalizability of results. Future models should also be evaluated with 4-quadrant classifiers based on Russell's emotion circumplex. The 4-quadrant classifier is ideal because it provides an objective metric that is unaffected by the sampling errors of subjective data. Separation into quadrants based on valence and arousal is useful to understanding the overall emotional expression of generated music while leaving room for algorithmic error. The combination of these two evaluation methods allows for the effective conveying of emotional accuracy in terms of musical structure and human response.

Vol. 2025 (5) 393 – 401
ISSN 2688-3651 [online]

References

Agres, K. R., Dash, A., & Chua, P. (2023). AffectMachine-Classical: a novel system for generating affective classical music. *Frontiers in Psychology*, *14*. https://doi.org/10.3389/fpsyg.2023.1158172

Bao, C., & Sun, Q. (2022). Generating music with emotions. *IEEE Transactions on Multimedia*, *25*, 3602–3614. https://doi.org/10.1109/tmm.2022.3163543

Dash, A., & Agres, K. (2024). AI-Based Affective Music Generation Systems: A Review of Methods and Challenges. *ACM Computing Surveys*, *56*(11), 1–34. https://doi.org/10.1145/3672554

Grekow, J., & Dimitrova-Grekow, T. (2021). Monophonic music generation with a given emotion using conditional variational autoencoder. *IEEE Access*, *9*, 129088–129101. https://doi.org/10.1109/access.2021.3113829

Huang, R., et al. (2021). Melody Generation with Emotion Constraint. In *Proceedings of the 2021 5th International Conference on Electronic Information Technology and Computer Engineering*. Association for Computing Machinery. https://doi.org/10.1145/3501409.3501691

Madhok, R., Goel, S., & Garg, S. (2018). SentiMozart: Music Generation based on Emotions. In *Proceedings of the 10th International Conference on Agents and Artificial Intelligence - Volume 1: ICAART*. SciTePress. https://doi.org/10.5220/0006597705010506

Mehrabian, A. (1996). Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in Temperament. Current Psychological Research & Reviews, 14(4), 261–292. https://doi.org/10.1007/bf02686918

Miyamoto, K., Tanaka, H., & Nakamura, S. (2020). Music Generation and Emotion Estimation from EEG Signals for Inducing Affective States. In *Companion Publication of the 2020 International Conference on Multimodal Interaction*. https://doi.org/10.1145/3395035.3425225

Pangestu, M. A., & Suyanto, S. (2021). Generating Music with Emotion Using Transformer. In *2021 International Conference on Computer Science and Engineering (IC2SE)* (pp. 1-6). https://doi.org/10.1109/ic2se52832.2021.9791928

Posner, J., Russell, J. A., & Peterson, B. S. (2005). The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Development and Psychopathology*, *17*(03). https://doi.org/10.1017/s0954579405050340

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, *39*(6), 1161–1178. https://doi.org/10.1037/h0077714

Sulun, S., Davies, M. E. P., & Viana, P. (2022). Symbolic music generation conditioned on Continuous-Valued emotions. *IEEE Access*, *10*, 44617–44626. https://doi.org/10.1109/access.2022.3169744

Wallis, I., Ingalls, T., & Campana, E. (2008). Computer-Generating emotional music: The design of an affective music algorithm. In *Proceedings - 11th International Conference on Digital Audio Effects, DAFx*.

Wang, Y. (2021). Music composition and emotion recognition using big data technology and neural network algorithm. *Computational Intelligence and Neuroscience*, *2021*, 1–11. https://doi.org/10.1155/2021/5398922

Wiafe, A., & Fränti, P. (2023). Affective algorithmic composition of music: A systematic review. *Applied Computing and Intelligence*, *3*(1), 27–43. https://doi.org/10.3934/aci.2023003

Zhao, K., et al. (2019). An Emotional Symbolic Music Generation System based on LSTM Networks. In *2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)* (pp. 2039–2043). https://doi.org/10.1109/itnec.2019.8729266

Vol. 2025 (5) 393 – 401
ISSN 2688-3651 [online]

Zheng, K., et al. (2022). EmotionBox: A music-element-driven emotional music generation system based on music psychology. *Frontiers in Psychology*, *13*. https://doi.org/10.3389/fpsyg.2022.841926

Zheng, L., & Li, C. (2024). Real-Time Emotion-Based Piano Music Generation using Generative Adversarial Network (GAN). *IEEE Access*, *12*, 87489–87500. https://doi.org/10.1109/access.2024.3414673