AI-Based Affective Music Generation Systems: A Review of Methods and Challenges

ADYASHA DASH, National University of Singapore, Singapore, Singapore
KATHLEEN AGRES, National University of Singapore, Singapore, Singapore

Music is a powerful medium for altering the emotional state of the listener. In recent years, with significant advancements in computing capabilities, artificial intelligence-based (AI-based) approaches have become popular for creating affective music generation (AMG) systems. Entertainment, healthcare, and sensor-integrated interactive system design are a few of the areas in which AI-based affective music generation (AI-AMG) systems may have a significant impact. Given the surge of interest in this topic, this article aims to provide a comprehensive review of controllable AI-AMG systems. The main building blocks of an AI-AMG system are discussed and existing systems are formally categorized based on the core algorithm used for music generation. In addition, this article discusses the main musical features employed to compose affective music, along with the respective AI-based approaches used for tailoring them. Lastly, the main challenges and open questions in this field, as well as their potential solutions, are presented to guide future research. We hope that this review will be useful for readers seeking to understand the state-of-the-art in AI-AMG systems and gain an overview of the methods used for developing them, thereby helping them explore this field in the future.

CCS Concepts: • **Applied computing** → **Sound and music computing**; • **Information systems** → *Multimedia content creation*; • **Computing methodologies** → **Artificial intelligence**; • **General and reference** → **Surveys and overviews;**

Additional Key Words and Phrases: Affect, emotion, music, generative AI, automatic music generation, deep learning, machine learning

ACM Reference Format:
Adyasha Dash and Kathleen Agres. 2024. AI-Based Affective Music Generation Systems: A Review of Methods and Challenges. *ACM Comput. Surv.* 56, 11, Article 287 (July 2024), 34 pages. https://doi.org/10.1145/3672554

1 Introduction

Music can have a remarkable impact on listeners' emotional states [57]. Indeed, music is often used as a powerful medium for inducing and mediating the mood and emotional state of the listener [4, 19]. With the advent of a new era in computing, researchers are gaining interest in designing **artificial intelligence (AI)**–driven music generation systems [7, 42]. A set of these computational music composition systems focus on creating affective music. Such systems may be referred to as **AI-based affective music generation (AI-AMG)** systems. AI-AMG systems often have certain

Authors' Contact Information: Adyasha Dash (Corresponding author), National University of Singapore, Singapore, Singapore; e-mail: adyashadash90iitgn@gmail.com; Kathleen Agres, National University of Singapore, Singapore, Singapore; e-mail: katagres@nus.edu.sg.

ACM 0360-0300/2024/07-ART287
https://doi.org/10.1145/3672554

benefits compared with human-created music, such as the ability to skirt copyright issues, the computational means of blending genres/musical elements in novel ways, and, in the case of real-time music generation systems, the ability to flexibly tailor the generated music to aspects of the environment or changes in the listeners' physical or emotional state. In addition, AI-AMG systems are potentially capable of creating an infinite number of unique affective music compositions and composing music without any associated time constraint [112]. Due to these advantages, controllable AI-AMG systems are now rapidly gaining the attention of researchers as well as companies such as Google and Sony, which are actively pursuing the development of creative, interactive music generation methods.

AI-AMG systems have great potential to impact many fields, including, but not limited to, healthcare, co-creativity, and entertainment (gaming). By exploiting the power of affective music to induce/mediate/enhance different psychophysiological states in the listener, AI-AMG systems have been deployed in different sectors of healthcare [4]. Affective music is often useful to improve the mood state of patients suffering from anxiety [29] and depression[98] while also promoting self-expression [19]. Affective music can also be effective during rehabilitation to promote physical activity or rhythmic entrainment to the beat of the music. Furthermore, emotional music can be useful in uplifting patients' mood state, thereby promoting better adherence to their prescribed rehabilitation exercises [36]. Thus, music-based mood mediation techniques can be applied to patients suffering from neurological disorders such as stroke [24–26] while improving their participation in rehabilitation therapy. Another important area of application for AI-AMG is co-creativity. Computational co-creativity in affective music composition refers to the collaborative composition or improvisation of music by humans and computers. AI-based algorithms can support aspects of the creative process, such as mechanizing part of the music composition (e.g., the accompaniment), thereby sharing some of the creative burden with musicians [59, 75]. Entertainment is another area in which affective music may find various applications, such as music to accompany gaming and **virtual reality (VR)/augmented reality (AR)**–based story-telling scenarios. The use of affective music in VR-based games can enhance the user's sense of immersion while facilitating one's mediated presence [39]. In the case of story-telling scenarios, affective music and designated motifs can accompany certain characters and situations in the narration to enhance the experience and better capture the attention of the listeners. Therefore, this technology-based approach to generating affective music promises to be immensely helpful in multiple fields and can inspire composers and musicians through co-creativity.

Due to the rapidly growing interest in automatic AMG, it is prudent to take stock of existing systems and review the literature both to summarize the state-of-the-art and to help researchers working in the field gain a more thorough understanding of the most helpful techniques/methods in the area (e.g., what architectures seem most effective and what features lead to the greatest emotion induction). Yet, to our knowledge, no such recent review exists. To our knowledge, only one review article [113] has been written on this topic, which was published in 2015. Since that time, there have been huge advances in computational techniques, which have led to more advanced music generation systems using different computational methods. A later review article by Ji et al. [55], which focused on surveying deep music generation systems, briefly summarized the state-of-the-art deep neural network architectures used for generating affective music. However, a detailed survey on recent advancements in AMG systems has not yet been reported. Because this review aims to assist researchers working on state-of-the-art AMG, it also focuses more on the technical aspects of AMG systems as compared with [113]. In addition, we note that while numerous music generation systems exist, many overlook the emotional aspects of music despite the fact that conveying emotion is often one of the goals of music composition and performance. Historically, the majority of systems neither explicitly consider nor manipulate emotional content, nor do they evaluate the *affective* qualities of the generated music. However, a growing number of systems now *do* aim to create emotionally expressive music. Our focus here lies on reviewing controllable affective music generation systems.

In this review, we (1) summarize the literature on AI-AMG systems, (2) categorize these systems based on the core algorithm/method used for the music generation, (3) provide an extensive review of the different classes of AI-AMG systems, (4) identify existing challenges in state-of-the-art AI-AMG systems, and (4) explore potential directions for future research.

The remainder of the article is organized as follows. Section 2 presents the method adopted for conducting this review, followed by a background of AI-AMG in Section 3. Section 4 presents a detailed review of the literature. Section 5 presents the important musical features and methods used to manipulate them. Challenges and future work in the field are discussed in Section 6. The article is summarized with concluding remarks in Section 7.

2 Methodology

To review the literature, we conducted a systematic search of three different search engines: (1) Google Scholar, (2) Scopus, and (3) IEEE Xplore. These websites were mined using the following search queries: (1) (Affect OR Emotion OR Mood) AND (Synthetic OR Artificial) AND (Synthesis OR Generation) AND Music, (2) (Affective OR Emotion OR Mood) AND (Synthetic OR Artificial) AND (Synthesis OR Generation) AND Lead sheet, (3) (Valence OR Emotion OR Affect OR Mood) AND (Generation OR Synthesis) AND (Lead sheet OR Melody OR Rhythm), and (4) (Affect OR Emotion OR Mood) AND (Generation OR Synthesis OR Composition) AND (Music AND (Lead sheet OR Score)). During the initial shortlisting process, articles from topics such as (1) Emotion-based automatic playlist generation, (2) Emotion recognition from music, and (3) Pleasant sounding artificial music synthesis, and related fields that do not address the affective component of music generation, were excluded. The remaining relevant articles that include the design of a controllable computational affective music generation system/algorithm were shortlisted after examining the abstract. Furthermore, we looked into the references of the relevant articles in our initial search and included any additional articles published between 1990 and 2023 that met our criteria. This resulted in 63 articles in total. The shortlisted 63 articles were then critically reviewed. A detailed comparison of the articles is presented in this article. Out of these 63 articles, 38 articles were published after 2015.

Note that the scope of this literature review, as described above, is to identify papers describing controllable AI-AMG systems that explicitly consider emotion, affect, or mood. Part of the motivation behind these search criteria was (1) to take stock of how many controllable affective music generation systems are in the literature, and (2) to gain insight into the methods researchers have used to generate affective music and, in many cases, how they manipulate features in order to express different emotions (Section 5 of this article discusses the various methods employed to generate different emotions or levels of arousal/valence).

We note that, since 2023, a number of audio-based text-to-music systems have been developed. While these systems do not explicitly mention or evaluate their output in terms of target emotions, the text prompt input does allow the users to specify an input mood or emotion. These novel systems are typically diffusion models or autoregressive transformers combined with a **large language model (LLM)** [12, 21, 31, 74, 90]. While these systems can be instructed to generate a ‘sad song’ or ‘upbeat melody’ due to the open text input, the systems have not yet been explicitly evaluated in terms of their ability to express emotion. There is significant interest in this new field of text to music; we believe that it is only a matter of time before dedicated evaluation is conducted in terms of how well the systems are able to control and convey emotion and mood. However, according to the methodology of this survey article, no text-to-music systems were

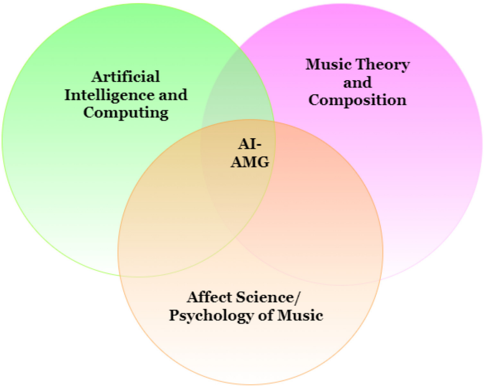
Artificial Intelligence and Computing
Music Theory and Composition
AI-AMG
Affect Science/ Psychology of Music

Fig. 1. Major underlying fields of AI-based affective music generation.

included in the literature discussed in Section 4, as these papers have not yet investigated emotion specifically.

There have also been impressive advances from industry[1], some of which are also recent text-prompt systems, such as Suno[2] and Udio.[3] For instance, the MusicLM music generation model from Google [1] can generate music from text using a hierarchical sequence-to-sequence modeling approach. However, again, systems that do not have a paper that explicitly mentions emotion, mood, or affect are not within the scope of this article, as they do not meet our systematic search criteria.

In the next section, we present a brief background of affective music generation systems.

3 Background: AI-Based Affective Music Generation

AI-AMG is an interdisciplinary field that requires knowledge of AI, music theory, and/or principles of music composition as well as the fundamentals of affective science. Figure 1 illustrates the main fields that interact to create AI-based affective music, i.e., computationally generated music that is meant to produce some perceived or induced emotion in the listener. Knowledge of AI is important for designing an algorithm to generate automatic music. In order to compose high-quality, real-sounding music, the algorithm requires knowledge of musical rules/structure that comes from music theory and composition. In addition, concepts from affective science are useful for understanding the emotional expression of music and how listeners respond emotionally to music. In short, AI-AMG requires interdisciplinary expertise that spans several fields, such as computing, music theory and composition, cognitive science/psychology, mathematical modeling, signal processing, and sometimes even areas such as physiology and neuroscience. This review covers the novel contributions from the main fields that contribute to the core design elements/components of an AI-AMG system and provides a comprehensive comparison between the existing systems that are designed to generate AI-based affective music. In the following section, we present a general overview of AI-AMG systems and provide a description of the major components of the system.

AI-AMG systems usually have three major components: (1) **Target Emotion Identification (TEI)**, (2) **Affective Music Generation (AMG)**, and (3) **Emotion Evaluation (EE)**. Figure 2 shows a block schematic diagram of an AI-AMG system and the interaction of its components.

[1] For example, see Google's Magenta (https://magenta.tensorflow.org), Meta's AudioCraft (https://ai.meta.com/resources/models-and-libraries/audiocraft/, and OpenAI's MuseNet (https://openai.com/research/musenet).

[2] Suno: https://suno.com/

[3] Udio: https://www.udio.com/

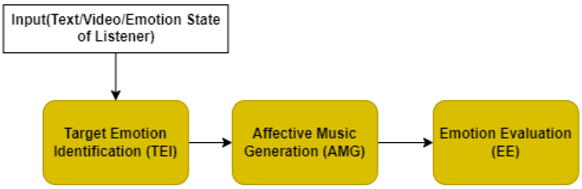
Input(Text/Video/Emotion State of Listener)
Target Emotion Identification (TEI)
Affective Music Generation (AMG)
Emotion Evaluation (EE)

Fig. 2. Components of an AI-AMG system.

The TEI component takes input from the user or input device and maps it to the emotion domain in a representation usable by the system. The AMG component then uses the emotion information provided by the TEI component and composes affective music accordingly. Lastly, the affective musical outputs (e.g., pieces or excerpts of music) are evaluated by the EE component (which can involve both computational and human evaluation) to examine their emotional expressiveness. A detailed description of each of these components is presented in the following subsections, followed by a description of (and considerations concerning) datasets used for training AI-AMG systems.

3.1 Target Emotion Identification

The aim of this component is to identify or select the target emotion for the AMG, i.e., the emotion(s) that the music aims to express or induce in the listener. This component takes the input from the user/input device in the form of text, image, video, sensor data, score, lead sheet, user input, or pre-determined target emotions, and provides the target emotion(s) as output, which is then used to drive the AMG. Target emotions can be provided as discrete emotions or may be indicated as valence-arousal coordinates (e.g., a point in continuous emotion space, which is typically a **2-dimensional (2D)** or **3-dimensional (3D)** plane) on an emotion map, such as the Circumplex model (described below). Similarly, the output data from the TEI component can either be presented as an element of a discrete emotion set or as a point in continuous emotion space, e.g., within an emotion map. The popular choices for discrete emotion sets include [happy and sad], and [happy, sad, calm, and angry]. Here, the elements of the discrete emotion sets can be visualized as categorical variables, and these categories of emotions are distinct from one another. In the case of an emotion map/continuous space, most often a circumplex model of emotion is used, such as that proposed by James Russell [88], to represent the target emotion(s) in multi-dimensional space. Most commonly, the circumplex model features a two-dimensional plane that includes the dimensions of valence (from negative/unpleasant to positive/pleasant) and arousal (the degree of energy or activation of the emotion). For example, the emotion “angry” would be considered low valence and high arousal, while “calm” would be high valence and low arousal. For visual examples of the 2-dimensional arousal-valence emotion plane, we recommend the reader consult Russell [88] and Yang et al. [116]. This dimensional representation of emotion has been used extensively in **music emotion recognition (MER)** and music generation and the field of **music information retrieval (MIR)** more generally. In this representation, the emotions are real-valued numbers that can be localized as a point on the map. The output of the TEI component can either be static, i.e., the TEI outputs a single discrete emotion or a point in emotion space, or dynamic, i.e., the TEI outputs a time-varying sequence of emotions or valence/arousal values.

The choice of input and output data for the TEI component depends on the application for which the affective music is being generated. For example, in order to generate a background music soundtrack for a gaming environment, the corresponding TEI component takes the video playback of the game as input. On the other hand, for some systems, the TEI component can take direct emotions or texts/emojis as input depending on the application. The type of output emotion also depends on the application. For instance, in an application such as the generation of affective music for treating depressed individuals, one may wish to include a discrete set of target output emotions such as calm and happy in the TEI component and the output emotion may be static. Conversely, applications such as generating background sound for multi-agent gaming environments will require the TEI component to output a time-varying sequence of target emotions depending on the performance of the player and the game states [51].

The TEI component often uses different types of feature extraction/classification algorithms to map the input data (text, video, etc.) to the space of emotions. Depending on the application, different algorithms are used to map the video or image files onto an emotion space. For instance, a **convolutional neural network (CNN)** architecture has been used to map input images onto a discrete emotion space (with 7 emotions) based on the emotional content in the image [70]. In some cases, the input to the TEI component is one or more emotions that come directly from the designer/user and require no algorithmic processing. In this case, the TEI component acts as a buffer and directly feeds the input emotions to the AMG component, which synthesizes the affective music based on this information.

3.2 Affective Music Generation: Approaches and Techniques

The aim of the AMG component is to compose affective music that can express or induce the target emotion(s). The input to this component is the target emotion as described above and the output is a piece, excerpt, or continuous stream of music intended to express the target emotion. In the AMG component, music is usually represented as a sequence of musical events in symbolic notation, in which each event is a combination of different musical features (tempo, note duration, chord sequence, melody, harmony, timbre, etc.) at different levels of hierarchy [73]. Briefly, a complex musical structure consists of low-level musical features, such as notes and chords, which are the basic building blocks for relatively high-level features, such as motives and phrases, and there exists a certain type of inter-dependency between these features in music [73]. Recently, researchers have also begun to explore how to use audio-based representations to generate affective music.

The representation of music in the AMG component can be broadly classified into two categories: (1) symbolic and (2) audio. In symbolic representation, the music is presented as a combination of discrete variables that capture musical features (such as pitch, duration, and chord) in an abstract form [67]. One very popular symbolic representation of music is the **Musical Instrument Digital Interface (MIDI)** format, a controlled protocol and interface between computer and musical instruments, in which MIDI events are used to realize various features of musical events. For example, in MIDI notation, a note or chord (types of MIDI events) would be represented using tuples of numbers that reflect different properties of the event, including pitch, velocity (amplitude), vibrato, and panning [67]. In addition to MIDI, symbolic representation can also be in formats such as LilyPond, Humdrum kern, and MusicXML. The advantage of using such music engraving representations is that they can capture the details of Western notation, including concepts such as notes, rests, keys, time signatures, articulation, ornaments, codas, and repetitions. Within actual AMG systems, these symbolic notations are translated into formats such as piano-roll or token-based representations. For an overview of symbolic representations used in music generation, see Le et al. [66].

In addition to a sequential representation in which the musical properties are represented as a **1-dimensional (1D)** sequence, a symbolic representation can also feature a 2D matrix in which the musical properties are represented at a particular temporal resolution [52, 117]. For instance, the piano roll representation requires two main pieces of information — the pitch and duration of the note — which can be represented as a 2D real-valued matrix. The rows of the 2D matrix represent the pitch of the notes, the number of columns reflects the temporal resolution, and each element of the matrix defines whether a note is to be played or not at a certain time [52, 117]. Researchers have used a 2D symbolic representation conditioned on emotion to generate affective music using a deep neural network architecture [117]. Apart from sequential and matrix-based representations, the music can also use text, an embedding (such as word2wav), or a graphical representation to represent music symbolically [52, 67] in AI-AMG systems. Symbolic representations have a loss of information, however, due to their low-dimensional nature (e.g., all of the spectral information in audio is not represented).

On the other hand, in audio representations, the raw music file is sampled and stored using the canonical **Waveform Audio File Format (WAVE)** and **Audio Interchange File Format (AIFF)**. Audio representations present the music as a continuous signal (a 1D waveform) in the time domain or as 2D spectrums (spectrogram or Mel-spectrogram, etc.) in the frequency domain [52]. In recent work describing a deep neural network-driven affective music composition system, the authors conditioned the input music, represented in 1D audio, on the target emotion to generate emotional music [95]. To the best of our knowledge, the use of 2D spectral representations has not yet been explored for *affective* music generation, although 2D spectral representations have been used for automatic music generation more generally [52]. The main advantages of working with audio rather than symbolic representations are the greater availability of training datasets (see Section 3.4 for more discussion on datasets) and the more realistic-sounding output of audio compared with, for example, MIDI. On the other hand, much more powerful **graphics processing units (GPUs)** would be required to handle large audio datasets for training and running models that use audio-based representations.

To generate affective music for given target emotion(s), the AMG component employs particular algorithms/methods to manipulate the musical features and their interactions conditioned on the target emotion at different hierarchical levels. Further, the AMG may be designed to operate in an open loop, in which the target emotion (either in static or dynamic form) is given directly to the system or may operate in a closed loop, in which the TEI is supplied or estimated based on user feedback, e.g., from users' real-time physiological states.

Various music generation algorithms and approaches may be used to compose the affective music. AI-AMG systems can be classified according to four broad types of algorithm/methods: (1) rule-based methods, (2) data-driven methods, (3) optimization methods, and (4) hybrid methods. We briefly describe each approach below. This classification is pictorially presented in Figure 3.

3.2.1 Rule-Based Methods. Rule-based methods define musical rules that capture the relationships between musical features used for the composition. These musical rules are represented in the AMG component either in the form of mathematical equations or logical statements/heuristic rules (if-else statements). The AMG component uses a set of rules to compose affective music that is meant to convey or induce the target emotion(s).

Depending on the relationship between the musical features and the emotion, the rules (whether simple or complex, as described in Section 4.1) can be described as feature-to-emotion rules and feature-to-feature rules. Feature-to-emotion rules define the relationship between the musical features and emotion dimensions. For example, Wallis et al. [109, 110] represent the relationship between the tempo, articulation, and roughness, with different levels of arousal (an emotion dimension). More recently, researchers have used this information to construct a more general representation of these rules in the form of parametric equations in which the musical features are parameterized by levels of valence and arousal (emotion dimensions) [28]. On the other hand, feature-to-feature rules depict the relationships between the features at different levels of the

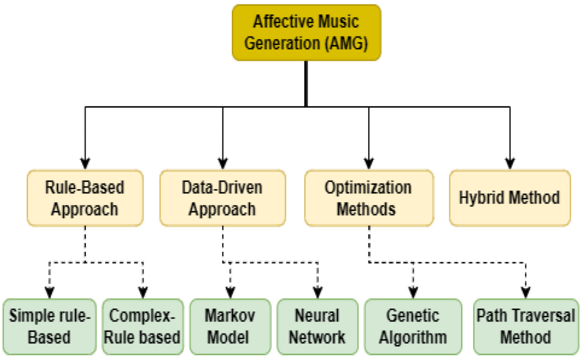

Fig. 3. Classification of AMG component based on the type of algorithm/method used in music generation.

musical hierarchy. One such example is the “rule for selection of a chord from a chord set to fit a given melody”. These rules can be encoded in the AI-AMG system in the form of a heuristic rule, conditional statements, or mathematical equations. Specifically, a mathematical formulation of rule-based modeling relies on exploiting the functional relationships between emotion parameters ($valence$ and $arousal$, $V, A \in [1, 0]$, both V and A are real numbers) and musical features such as tempo, rhythm, and more. For example, many systems define a linear relationship between tempo and arousal [3]; for example, the authors of [28] express this relationship mathematically by using arousal to dictate the duration in seconds of the smallest note value (an eighth note, in this case), as follows: $Note_{duration} = 0.3 - arousal * 0.15$. A number of these rules are grouped together to form the knowledge base of the AMG component, which then makes use of this knowledge to generate affective music. These rule-based approaches governed by a set of mathematical equations are efficient in terms of time and computation for generating music.

3.2.2 Data-Driven Methods. Data-driven methods used in the AMG component capture and leverage the patterns present in data to learn the innate structural information in music. Structural information is learned during the training stage and stored in the form of model parameters. Subsequently, the model parameters are used to generate artificial music during the generation process. This training and subsequent generation process can be conditioned on the target emotion to compose affective music. Depending on the structure of the model, data-driven methods are of two types: (1) Markov model–based approaches and (2) neural network–based approaches. In Markov model–based approaches, the model uses the syntactical/statistical information related to the occurrence of musical features and estimates the transition probabilities. Subsequently, the model is used to predict the feature values that fit the musical context. Researchers have used this method to predict features such as note value, chord movement, and octave movement/register for a given musical piece [85]. On the other hand, neural network–based approaches use popular deep neural network architectures for learning inherent dependencies between the features of music. Later, in the generation phase, these models are used to efficiently generate novel affective music. Neural networks such as seq2seq neural network architectures, **Long Short-Term Memory (LSTM)**, Transformer, **Variational Autoencoder (VAE)**, **Generative Adversarial Networks (GANs)**[38] and other well-known architectures are also used for composing affective music. These models either vary in the architecture (e.g., LSTM, Transformer) used or in the formulation of the loss function (e.g., variational loss and generative adversarial loss). A detailed discussion about the classification of these models is provided in Section 4.2.2. Despite the need to train the model on a large and often labeled dataset, the unified architecture, easy availability, smooth operation, and high accuracy are some of the qualities that have made these data-driven architectures a popular choice for affective music composition.

3.2.3 *Optimization Methods.*

A number of optimization methods have also been successfully deployed in the automatic composition of affective music. Broadly, these optimization methods can be put into two categories: (1) genetic algorithm and (2) tree/graph traversal optimization methods.

The genetic algorithm, a nature-inspired optimization method, has been leveraged for generating affective music. The algorithm can be used to optimize a cost function in both constrained and unconstrained environments. The affective music synthesis procedure may be modeled as a multi-objective constrained optimization problem in which the musical rules are stated as constraints. The genetic algorithm-based method can then be used to find the best-fit value for a music feature without violating the musical rules. For example, Scirea et al. used a genetic algorithm for selecting the melody for a given chord progression while not violating music theoretic rules [94]. Here, the evolutionary genome consists of a number of values (the number of notes to be generated). The generation process has objectives, e.g., a melody should approach and follow leaps larger than a second (musical interval) in a counter step-wise motion, as well as constraints, e.g., a melody should not have leaps between notes bigger than a fifth. The objectives constitute the fitness function that is used to generate a melody without violating the musical rules given the constraints. In this AMG system, named MetaCompose, alongside the melody, the chord progression, accompaniment, and so forth, are also tailored to collectively generate affective music. For details, please refer to [94].

Other optimization techniques based on path traversal, namely, (a) tree traversal and (b) graph traversal methods, have also been used in AI-AMG systems. Let us consider a graph $G(V, E)$ with vertices V and edges E. These methods select the path starting from a vertex $u \in V$ to a desired vertex $v \in V$ in the graph that optimizes a given cost function C. This graph $G(V, E)$ can be constructed by employing musical theory, in which vertices depict states of musical features and edges between these states are weighted based on their relatedness/closeness for a given context/emotion. After construction of the graph, the process of music generation simply involves traversing the path from the vertex u to vertex v as provided by the user. For example, Kuo et al. [64] designed a tree network for chord selection in which each vertex represents a chord name, and the path length between each linked pair of chords is weighted by the valence value of the emotion. The cost function aims to optimize the total path length by selecting the closest next chord in a sequence of chords given a valence/emotion value. Thus, here, C represents the sum of weights on the path from u to v. Similarly, graph traversal methods have also been used to select chords for composing affective music [94].

3.2.4 *Hybrid Methods.*

In addition to using the aforesaid methods separately, researchers have used a combination of these methods to compose affective music. In this case, the AMG component uses different methods (traversal methods, genetic algorithms, rules, and neural networks) to predict/select different musical features for a given emotion. Subsequently, these features are collectively presented to generate the composition. For instance, the authors of [64] used the combination of a graph-based traversal method, genetic algorithm, and a rule-based technique for selecting the next chord in a sequence of chords, generating a melody that fits the chord progression and features such as tempo and timbre, for a given emotion/affective state.

The above AI-AMG methods are used for composing music that can express the target emotion. Subsequently, the artificially generated music is evaluated for its efficacy in expressing the given emotion. This final step is overseen by the emotion evaluation component.

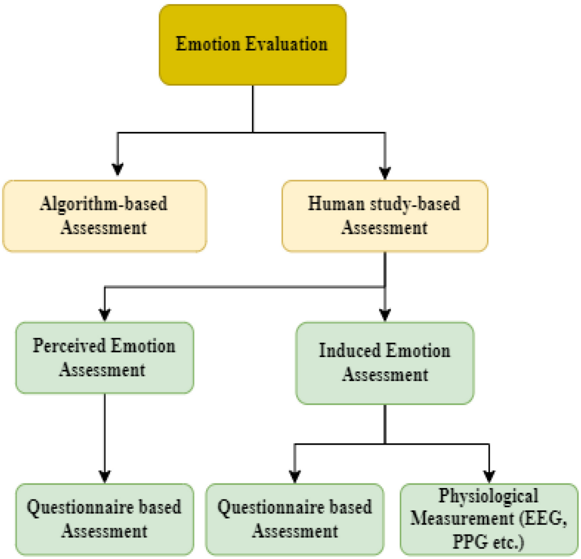
Emotion Evaluation
Algorithm-based Assessment
Human study-based Assessment
Perceived Emotion Assessment
Induced Emotion Assessment
Questionnaire based Assessment
Questionnaire based Assessment
Physiological Measurement (EEG, PPG etc.)

Fig. 4. Classification of emotion evaluation component.

3.3 Emotion Evaluation

The aim of this component is to evaluate the efficacy of artificially composed music for expressing the target emotion. Different approaches have been used by researchers for evaluating the emotional content of artificially generated music [2], which can be broadly classified into two categories: (1) **algorithm-based assessment (ABA)** and (2) **human study–based assessment (HBA)** (depicted in Figure 4). Evaluation is an extremely important, but seemingly undervalued, aspect of many AI-AMG systems. As argued by Sturm [99], it is often the case that the evaluation of many MER systems lacks validity, which then makes it difficult to arrive at accurate and meaningful conclusions. Further, many papers do not include any formal evaluation of their AI-AMG systems (43% of papers reviewed here), and several suffer from the general trend in AMG (see [2]) of including only subjective assessment on the part of the researcher, which is of course prone to bias. Out of the 63 articles reported in this survey, 36 have used assessment methods (1 used only ABA, 33 used only HBA, and 2 used both ABA and HBA methods) for evaluating the generated musical excerpts. These articles, along with the assessment method(s) employed, are presented in Table 1.

Algorithm-based assessment methods use an analytical approach for comparing certain properties of the generated affective music with a sample template to generate a measurement index. Based on the index value, the efficacy of the system for generating affect-specific music is determined. For example, Sergio and Lee [95] used an "emotional circle of fifths," in which chords are arranged in a circle based on their emotion-conveying capacity. The circle has 5 chords representing positive emotions (red color), 6 representing negative emotions (blue color), and 1 chord, F, that can be either positive or negative (purple color). The chords in a generated piece are then extracted, and the number of positive and negative emotions are computed. If the number of chords conveying positive emotions is higher than those conveying negative, then the excerpt is declared to represent a positive emotion (and vice versa). In another study, Makris et al. [71] manually assigned valence values to each chord type, drawing inspiration from the research conducted by Chase [11] and Scherer [89]. This unique approach was used to label a training dataset of lead sheets with emotions. Later, when assessing the emotional content of the generated music, the

Table 1. Summary Table of Articles Surveyed and Their Assessment Methods

Ref	Generation Model	Emotion Model	Features	Submethod	EEM	NDE
[58]	Rule-based	DE	Rhythm and Chord progression	CR	HBA	–
[100]	Rule-based	DE	Major/minor key, Tempo and Instrumentation	SR	–	–
[107]	Rule-based	DE	Mode, Dissonance/Consonance, Melodic pitch range and Tempo	CR	HBA	5
[44]	Rule-based	DE	Structure, Harmonic patterns, Motif patterns, Modes, Meters	CR	–	2
[22]	Rule-based	DE	Pulse salience and Kinesis of rhythm and diatonic mode	CR	HBA	4
[79]	Rule-based	DE	Key, Accompaniment (harmonic, rhythmic), Motif	CR	–	5
[86]	Rule-based	DE	Harmony, Melody, Rhythm and Volume	CR	–	2
[76]	Rule-based	V-A	Tempo, Rhythm, Loudness, Mode, and Pitch	CR	HBA	–
[28]	Rule-based	V-A	Tempo, Rhythm, Loudness, Pitch, Mode	CR	HBA	–
[110]	Rule-based	V-A	Pitch Register, Loudness, Roughness, Tempo, Articulation, Harmonic mode	CR	HBA	–
[48]	Rule-based	V-A	Tempo, Chord (major or minor) and Progression stability	CR	HBA	–
[35]	Rule-based	V-Energy	–	CR	–	–
[68]	Rule-based	V-A	Tempo, Rhythm, Loudness, Pitch, Mode	CR	–	–
[102]	Rule-based	–	Musical tension	CR	–	–
[117]	Data-driven	DE	–	NN	HBA	4
[70]	Data-driven	DE	–	NN	HBA	7
[47]	Data-driven	DE	–	NN	HBA	4
[95]	Data-driven	DE	–	–	HBA and ABA	2
[97]	Data-driven	DE	Chord sequence	HMM	–	2
[96]	Data-driven	DE	Chord sequence	HMM	–	2
[78]	Data-driven	DE	–	HMM	–	2
[34]	Data-driven	V-A	–	NN	HBA	–
[112]	Data-driven	V-A	Timbre, Key, Pitch spread, Tempo, and Envelope	NN	HBA	–
[23]	Data-driven	V-A	–	NN	–	–
[84]	Data-driven	V-A	–	NN	–	–
[71]	Data-driven	valence	–	NN	HBA and statistical comparison	–
[103]	Data-driven	valence	–	NN	–	–
[119]	OM	DE	–	GA	HBA	2
[8]	OM	romantic-Era Style	Harmonic tension, Formal regularities	GA	–	–
[114]	OM	V-A	Rhythm, Volume and Pitch range	–	HBA	–
[41]	OM	–	Musical tension	–	–	–
[51]	HM	DE	Harmony, Melody and Rhythm	–	–	–
[72]	HM	DE	Tempo, Chord, Note length and Octave	–	–	5
[64]	HM	DE	Tempo, Melody, Chord progression and Instrument configuration	–	–	12
[94]	HM	V-A	Chord sequences, Intensity, Timbre, Rhythm and Dissonance	–	HBA	–
[92]	HM	V-A	Chord sequences, Intensity, Timbre, Rhythm and Dissonance	–	HBA	–
[85]	HM	V-A	Tempo, Pitch range, Chord dominance, Note value, Octave and Chord	–	HBA	–
[13]	HM	V-A	Accompaniment, Pitch, Rhythm, Mode, Meter and Tempo	–	HBA	–
[10]	HM	–	Instrumentation, Harmonization, and Chord progression	–	HBA	–
[118]	Data-driven	DE	Mode and Tempo	NN	HBA	4
[27]	OM	DE	Scale, Tempo, Chord progression, Melody and Harmony	GA	–	2
[61]	Rule-based	V-A	Scale, Tempo, Pitch, Loudness, Keymode	CR	–	–
[115]	OM	DE	Scale, Tempo, Mode	GA	HBA	2
[111]	Rule-based	DE	Scale material, Timbre, Tempo, Sound level, Articulation, Time deviation	CR	HBA	4
[56]	HM	PAD	Pitch values, Duration between notes	–	HBA	–
[49]	Data-driven	V-A	–	NN	–	–
[83]	Data-driven	Energy-Tense	Duration, Vibrato	HMM	–	–
[46]	HM	V-A	Tempo, Mode, Pitch and Chords	–	–	–
[40]	Data-driven	DE	–	NN	–	4
[87]	Rule-based	–	Scale	–	–	–
[108]	Data-driven	–	–	NN	HBA	–
[77]	Rule-based	V-A	Tempo, Rhythm, Loudness, Pitch, Chord	–	–	–
[6]	Data-driven	DE	–	NN	HBA	4
[69]	Data-driven	V-A	–	NN	HBA	–
[101]	Data-driven	V-A	–	NN	–	–
[33]	OM	V-A	–	–	HBA	–
[53]	Data-driven	V	–	NN	HBA	5
[54]	Data-driven	V	–	NN	HBA	3
[82]	Data-driven	DE	–	NN	HBA	3
[50]	Data-driven	V-A	--	NN	HBA and ABA (classification)	–
[80]	Data-driven	V-A	–	NN	HBA	–
[32]	Data-driven	V-A	–	NN	HBA	–
[105]	Data-driven	DE	–	NN	HBA	2

EEM: Emotion Evaluation Method; NDE: Number of Discrete Emotions targeted; DE: Discrete Emotion; V-A: Valence-Arousal Dimensions; OM: Optimization Method; HM: Hybrid Method; CR: Complex Rule-based system; SR: Simple Rule-based system; NN: Neural Network; HMM: Hidden Markov Model; GA: Genetic Algorithm; CGA: Convectional Genetic Algorithm; PAD: Pleasure-Arousal-Dominance; "–": Not available; HBA: Human-Based Assessment; ABA: Algorithm-Based Assessment.

chords were extracted, allowing for the corresponding valence values to be computed. To gauge the emotional expression of AMG-generated music in comparison with compositions by human composers, the relative deviation (distance) between these emotion (valence) values was evaluated. Although algorithm-based evaluation methods are easy to implement and less time-consuming than human-based assessment, such methods do not directly capture human emotional responses, and often lack the ability to measure how "well structured" and natural the music sounds as well as the music's creativity (the system's ability to convey emotional content in a creative and not overly repetitive way) [2].

On the other hand, human study–based assessment methods [2, 34] rely on the listener's ability to rate the emotional content of the music. Emotion is a human phenomenon, and although emotion assessment by listeners can be noisy (as different individuals have different musical knowledge and preferences), human raters provide a direct assessment of perceived/induced emotion. In this approach, sample music pieces from the system are played to human listeners (individually or in a group) and the listeners are asked to provide their feedback/responses [34] regarding the emotional content of the music. A questionnaire is often used to collect subjective ratings or comments. This approach usually aims to quantify (through ratings) the emotion conveyed/expressed by the music (perceived emotion) or the emotion evoked in the listener by the music (induced emotion). Formally, *perceived emotion "refers to intellectual processing, such as the perception of an intended or expressed emotion"* [63], and *"induced emotion reflects the introspective perception of psychophysiological changes, which are often associated with emotional self-regulation"* [63]. Different scales, such as Likert scales, a Self Assessment Manikin [28], or **Mean Opinion Score (MOS)** [70], have been used for registering listeners' perceived and induced emotion responses. Thus, depending on the criteria of the emotion evaluation, human study–based assessment methods can again be subdivided into two categories: (1) perceived emotion assessment and (2) induced emotion assessment [110].

A conventional way of measuring *perceived emotion* is by collecting emotion ratings from the listeners in terms of valence-arousal levels or discrete emotion ratings. Here, the questions are framed to capture the expressed emotion in music. For example, "How positive/negative was the music?" or "Does the generated music fit the chosen mood?" [85] are some of the common example questions. *Induced emotion* assessment methods aim to estimate the evoked emotion in the listener during or after listening to a piece of music [100] and may be examined using questionnaires or through various physiological and sensor-based techniques [2]. For example, to capture induced emotion in the listener, the questions are framed around the mood state of the listener, i.e., "How happy/sad are you feeling now?". Further, physiological signals from **electroencephalography (EEG)**, facial **electromyography (EMG)**, and **photoplethysmography (PPG)** [100] may be monitored to identify and quantify the intensity of induced emotion in listeners [28, 100]. Such physiology-based assessment methods are often considered to be more reliable in the sense that these methods rely on involuntary, bias-free measures for "directly" quantifying induced emotion. Importantly, the choice of the emotion assessment method for an AI-AMG system should be made based on its potential application, namely, whether the AI-AMG system is used to express different emotions or is meant to induce emotions in the listener. For example, applications such as the therapeutic use of music for alleviating sadness, in which the goal is to mediate the mood of the listener, will aim to *induce* the desired target emotion(s) in the listeners. Thus, music produced by these systems should be evaluated on their emotion induction capabilities. On the other hand, applications such as background soundtrack generation for storytelling environments will usually focus on evaluating the *expressed* emotion of the generated music. Thus, the goal of the AI-AMG system dictates whether the systems should be evaluated based on their emotion induction or emotion expression (perceived emotion) capabilities. Even though 29 of the music generation systems reviewed here have been individually validated through assessment methods, the relative comparison between them has generally not been studied. This is mainly because of two factors: (1) the unavailability of musical pieces in the public domain to enable comparison (while those that *are* available are often "cherry-picked") and (2) variation in the tasks and input/outputs of the systems (e.g., some systems focus on a particular style, while others focus on matching the music to a visual or gaming environment).

Although existing AI-AMG systems are generally able to express the required target emotions in their generated music pieces, most researchers have not examined induced emotion in listeners, and their computational creativity has not yet fully matched the creative ability and quality of human-created music — there is room for improvement. The great majority of papers reviewed here focus on the ability of the system to either produce perceived or induced emotion in the listener or to try to minimize a cost function (for example), avoiding the issue of musical quality, and sometimes human ratings, altogether. Many systems claim to try to produce high-quality music, but very few ask listeners to rate the musical quality of the output or compare their system's generated music to human-composed music. Some exceptions are Ji and Yang [53], who included human ratings of naturalness (defined as "the degree of naturalness or authenticity of the music"), interestingness ("the extent to which the music is surprising or innovative"), and overall quality of the music; Makris et al. [71], who collected human ratings of naturalness (among other metrics) and compared between generated and human-composed music; and Neves et al. [80], who aimed to create "realistic" and "human-like" compositions, and had human raters assess the quality of the music.

3.4 Datasets

Training models for AMG often face a challenge due to the poverty of music datasets available with emotion labels, and the few datasets that are available are often regarded as noisy [15, 62]. Although more emotion labels exist for audio files (e.g., MP3s) than symbolic music files (e.g., MIDI tracks), the area of *audio* music generation is still fairly nascent. To date, most AI-AMG systems use, and are trained on, symbolic music notation.

Creating affective music datasets is very labor-intensive, as these datasets must be labeled with mood or emotion terms by hand. Further, the consistency in mood or emotion tags across raters can be poor due to the subjective nature of emotion perception in music and possible confounds regarding the listener's mood, familiarity with the music, and personal/cultural preferences [37, 62]. Indeed, a notorious challenge in the field of MER is the often poor to moderate level of inter-rater reliability in emotion labels, which makes settling on a ground truth difficult if not impossible [65, 99]. Further, in order to be distributed freely online, audio datasets such as the **Database for Emotional Analysis of Music (DEAM)** [5] must rely on royalty-free music, which removes most popular music (songs with widespread familiarity). This can also pose a challenge when wanting to train models on ecological, affective music. In other words, the training datasets may not be an accurate representation of what people actually hear in the real world.

Many of the labeled datasets that exist have arisen from the area of MER through the **Music Information Retrieval Evaluation eXchange (MIREX)** and **Multimedia Evaluation Benchmark (MediaEval)** competitions [5]). For example, one of the most well-known audio-based datasets in the field is the MediaEval DEAM, which contains 1802 excerpts and songs annotated with valence and arousal values [5]. For a more complete overview of music audio datasets with emotion labels, please refer to [16].

In terms of MIDI datasets with emotion annotations, which are needed for training most AI-AMG models for symbolic music generation, only a handful of options exist, such as the VGMIDI Dataset [34], which has 204 MIDI pieces with valence and arousal annotations, and the Panda et al. [81] dataset, which has 193 MIDI files with emotion labels (as well as additional labeled audio files). Recently, EMOPIA was released, a dataset of 387 piano music pop songs, containing audio and MIDI files annotated with perceived emotion ratings [50]. We expect to see more multimodal (audio + symbolic) emotion datasets emerging in the future. Note that these datasets contain on the order of 200 to 400 songs; by comparison, generative AI systems for creating visual artwork, such as OpenAI's DALL-E2[4], are trained on datasets with millions of examples. Due to this need for larger labeled MIDI datasets, Sulun et al. [101] recently released a dataset with 34,791 MIDI files with emotion labels based on the Lakh dataset. They also created a dataset with aligned MIDI and

[4] https://openai.com/dall-e-2

audio files (the Lakh-Spotify dataset). In the datasets by Sulun et al. [101], however, the arousal and valence annotations are taken from audio. Thus, they note that "because the Spotify features belong to the audio versions of each track, they can only be considered 'weak' labels for the MIDI versions" [101], e.g., because the MIDI files were not directly annotated and listeners' emotional responses may be due in part to timbre (etc.) not present in the MIDI.

In addition, although dimensional models of emotion (such as the circumplex model, using the dimensions of valence and arousal) are common to assess emotions perceived or induced by listeners during music listening, many datasets are only labeled with discrete emotion terms. A difficulty therefore arises when researchers want their system to seamlessly produce music based on varying levels of arousal/valence, but find that they are unable to accurately map discrete emotion terms to specific points or regions in arousal/valence space (however, see a workaround solution in [71] using a mapping from emotion terms to points in arousal-valence space inspired from [89]). More generally, given the overall scarcity of datasets in the field, models often have to rely on finding creative solutions, such as leveraging findings from music cognition and concepts from music theory in order to make connections between emotions/moods and music features.

In the next section, we present a detailed review of the state-of-the-art in AI-AMG systems. The papers below are briefly summarized to (1) capture which state-of-the-art methods are used and (2) allow the reader to assess which papers are most relevant to their research (to know where to direct further reading).

4 AI-Based Affective Music Generation Systems: State of the Literature

As discussed in the previous section, AI-AMG systems typically consist of three major components, (1) TEI, (2) AMG, and (3) EE. Among these three components, the core functionality of the music generation is performed in the AMG component. Thus, in this section, we present a discussion of the different AI-AMG systems based on the type of algorithm/method used for music generation.

This article reviews 63 research papers on controllable AMG systems. Of these articles, 28 adopted data-driven AMG methods, making the data-driven method one of the most used approaches. Amongst these, 4 employ a Markov model–based approach and the remaining 24 used a neural network architecture to generate affective music. In addition to data-driven methods, rule-based methods are also popular for AMG, with 18 articles in total. An optimization method is used in 7 articles and a hybrid method (employing multiple of the above strategies) is used in 10 articles. Table 1 presents a comprehensive list of these articles. Out of all articles reviewed here, only one uses an audio representation; the rest use symbolic representation.

4.1 Rule-Based Systems

Rule-based AI-AMG systems make use of musical rules (which define the relationship between musical features and emotion or the relationship between musical features at different levels of the musical hierarchy) to compose affective music. We categorize the musical rules informally into simple and complex rules based on how commonly they are used in the literature. Specifically, rules depicting the relationships between (a) 'tempo' and 'arousal' and (b) 'mode' and 'valence' are considered to be simple rules, as they cover basic relationships that are well accepted and have been used by almost all of the rule-based systems reviewed. Conversely, the complex rule set includes additional rules that define the relationships between emotions and pitch register, harmony, and chord progression in addition to the simple rules. Based on this categorization, the rule-based AMG systems can again be divided into two subcategories: (1) simple rule set–based systems, in which the rule set consists of only a small set of simple musical features such as tempo and mode (major/minor) to influence the emotional quality of the music, and (2) complex rule-based systems, in which the rule set consists of rules for other musical features (pitch register, harmony, chord progression, etc.) in addition to simple musical features such as tempo and mode. Both of these approaches are used in affective music composition.

4.1.1 Simple Rule-Based Systems. Some rule-based AI-AMG systems use rules that select the tempo and mode of the affective composition based on the target emotion. Tempo, measured as **beats per minute (BPM)**, is a simple feature that tends to have a direct relationship with the arousal component of emotion [109]. Similarly, the mode (major/minor) of the music typically affects valence [100]. These rules are widely accepted by the research community; thus, they are included in most rule-based AI-AMG systems.

Of all the papers summarized in this review, only one system has used a simple rule set for generating affective content in their musical compositions. Su et al. [100] have proposed a rule-based music generation system for improving the listener's affective state. The system is used to compare different emotion-induction strategies, namely, "Discharge," "Diversion," and "Discharge-to-Diversion," which all aim to induce positive affect in a listener compared with the listener's current emotional state. Briefly, "Discharge," "Diversion," and "Discharge-to-Diversion" are mood mediation strategies used for transferring the user from a sad/angry mood to a relatively happy mood by playing music expressing sad/angry emotions, pleasant emotions, and sad/angry to pleasant emotions, respectively. The system uses musical rules based on two features (major/minor mode and tempo) to generate musical events. Then, these events are played with different combinations of instruments (piano, guitar, cello, vibraphone, and kick drum). A horizontal re-sequencing method is then applied to the musical events in order to produce music for the three different strategies. The aim of the generated music is to create a happier emotional state in the listener; the emotional state is measured and quantified via the listener's facial expression. The results indicate that the Discharge-to-Diversion strategy is better at inducing positive affect compared with the other strategies.

A simple rule-based AI-AMG system is easy to design and implement. In addition to deploying simple musical rules based on tempo and mode, complex rule-based systems make use of musical rules that control other/additional musical features to generate music reflecting the target emotion. In the next section, we summarize the rule-based systems that use a complex rule set.

4.1.2 Complex Rule-Based Systems. Apart from using simple musical rules for tempo and mode, AI-AMG systems have also used more complex sets of rules to carefully manipulate musical features, such as harmony, melody, and rhythm, for creating an affective composition. Such AI-AMG systems are categorized as complex rule-based systems and summarized here. As mentioned above, these systems may be designed to operate in an open-loop mode in which the target emotion information is given to the system [87] or in a closed-loop mode in which target emotion information is decoded from users' physiological state in real time [28, 76].

First, we summarize the AI-AMG systems that were designed to operate in open-loop mode. Vieillard and colleagues [107] developed a system that manipulates mode, dissonance/consonance, melodic pitch range, and tempo for composing affective music. Later, Wallis and colleagues [110] developed a rule-based system that selects pitch register, loudness, rhythmic roughness, tempo, articulation, harmonic mode, and upper extensions for composing affective music with different levels of valence and arousal. Other researchers have developed a set of rules named KTH Music Performance Rules for composing affective music [35]. Musical rules can also be defined to manipulate musical features at different levels of hierarchy for composing affective music. For instance, Hoeberechts et al. presented a pipelined (sequential) architecture for synthesizing affective music by controlling the musical features at different levels of hierarchy [44]. Here, musical rules are defined to manipulate the high-level features, such as sections, blocks, and musical lines, as well as the low-level musical features, such as structure, harmonic patterns, motif patterns, modes, and meters, to synthesize happy and sad music. Recently, researchers have also tried to manipulate musical tension by changing the musical structure, i.e., varying the build-up and release of musical tension by violating musical structure (phrase violation and period violation) [102]. Varying the tension of computerized music can also be employed for designing AI-AMG systems.

In addition, AI-AMG systems have also been designed to operate in closed-loop mode. These systems are integrated with sensor-based platforms (EEG platforms, wearable sensors, etc.) for monitoring/decoding the real-time emotional state of the user. Subsequently, based on the decoded emotional information, these systems generate adaptive emotional music. In this line, researchers have developed AI-AMG systems for applications such as emotion mediation and emotion induction. Miyamoto and colleagues [76] developed a rule-based AI-AMG system in which the user can control the emotional music based on one's neural activation. The emotional information is decoded from real-time EEG activation of the user and mapped onto the valence-arousal plane. Subsequently, based on this emotional information, the system synthesizes affective music by collectively changing musical features such as tempo, rhythm, loudness, mode, and pitch using a set of rules. Similarly, Ehrlich et al. [28] have developed an AI-AMG system to support real-time emotion self-regulation in listeners by playing affective music based on the instantaneous EEG activation of the listener. The authors' AI-AMG system generates affective music by varying musical features, including tempo, rhythm, loudness, pitch, and mode. The AI-AMG system was embedded in a closed-loop **Brain–Computer Interface (BCI)** system, which leveraged EEG and neurofeedback to teach the listener to mediate one's own emotional states. The majority of the listeners were able to gain control of the music generation and self-induce the target emotion. Kirke and colleagues [61] and Agres et al. [3] have also developed an AI-AMG system that can potentially be integrated with an EEG module for user emotion mediation. In addition to EEG, wearable sensing technologies and motion/gesture sensing technologies have been employed for building a closed-loop AI-AMG system. In this endeavour, Wassermann et al. [111] have developed an affective music composition algorithm for designing an intelligent space called "Ada." Ada tracks users' emotions by monitoring gestures (video/image), audio (audio events), and body weight shifting profile (pressure load on the floor), which is then used to control the lighting and sound effects. The sound effect component of Ada makes use of affective music generation algorithms to generate mood-specific music by changing musical features such as scale, timbre, tempo, sound level, articulation, and time deviations using musical rules.

In addition, some researchers have tried integrating rule-based AI-AMG systems with multimodal platforms (narrative, virtual reality, storytelling, robotic platforms) to design systems for applications such as gaming and entertainment. In this endeavor, Kanno et al. [58] developed a rule-based AI-AMG system for generating affective music from input narratives. The authors first construct a musical rule base in the form of a dictionary linking a set of impression word pairs with two musical features: chord and rhythm. The impression word pairs, such as "Energetic - Calm" and "Fast - Slow," are used to represent the two extreme impressions or emotions. A narrative is presented to the system as input and the system changes the musical features (rhythm and chord progression) of the generated music using the rules, depending on the emotion and impression of the input narrative. A rule-based AI-AMG system has also been designed for storytelling applications [22]. In this computerized music generation system, musical features (pulse salience and kinesis of rhythm) and diatonic mode (major/minor) are manipulated to create emotionally expressive music. Similarly, the article by Nakamura and colleagues [79] describes a prototype system for composing affective background music and sound effects for a given animation. The input to the system is given in the form of mood parameters, music parameters, and motion parameters for the character in the scene. The mood parameters are glad, happy, sleepy, sad, angry, and tired, represented on a scale of 1 to 5. Music parameters are the musical motif, tempo, and timbre, and the motion parameters are composed of the actions of the character to be presented on the screen. Subsequently, these inputs are used to select different musical features. For example, based on the mood information, the system selects the musical key, the accompaniment (harmonic–rhythmic accompaniment), and the motif information (melody of the music) for generating the music. Interactive AMG has also been incorporated into an affect-sensitive robot [48]. The platform uses emotion-related information provided by the user to define the music parameters and the motion parameter for the robot. Specifically, the musical parameters were selected using a rule-based approach in which features such as tempo, chord (major/minor), and progression stability (chord changing rate) were determined based on the instantaneous valence and arousal values.

Most of these rule-based systems are able to reliably generate affective music pieces by encoding and then expressing the desired target emotion while using predefined musical rules. In the rule-based system by Wallis et al. [110], the perceived valence in an excerpt is dependent on the intended arousal level, whereas perceived arousal is not affected by the valence level. In addition to expressing certain emotions through music, many of these systems can also adaptively change the emotional expression in the music from one target emotion to another in real time. For example, the music can be dynamically changed depending on the scenes of a story-telling environment, the player's situation in a game, or the participant's current emotional state, not to mention applications in video gaming and mood mediation. Some of the articles covered in this review have also exploited the real-time transition capability of their AI-AMG systems and have integrated different sensor-based platforms (EEG platforms) to work in closed-loop mode or multimodal platforms (narrative/story-telling platforms, VR-based platforms) for creating adaptive AI-AMG systems. Adaptive systems (such as [28, 111]) are gaining popularity because of their potential for developing applications such as interactive games, music therapy, and computer music performance. Even though these systems are capable of being deployed in myriad applications, the design of these systems is often cited as critical. These systems often need skilled researchers with knowledge of music theory or composition to develop the rule set for the system. Furthermore, finding a convincing set of musical rules can be difficult, as the rules may differ for different genres/styles of music. Designing rules to tailor the musical feature for expressing different levels of valence has also been challenging [76] and needs further exploration. In spite of these limitations, rule-based systems are still a popular choice amongst researchers due to the ease of design, the requirement of fewer computational resources such as labeled datasets for training for implementation, and the ability to generate reliable, pleasant-sounding affective music.

4.2 Data-Driven Systems

Another popular way of generating affective music is by using data-driven approaches in which a database of music is used to train a model, i.e., learn the parameters of the model. Subsequently, these trained models are used to generate affective music to express a certain target emotion. The data-driven model may be a Markov model in which the model parameters are estimated or a deep learning model in which the model parameters are learned from a training dataset. The training of these model parameters can be conditioned on either discrete emotions or continuous emotions (e.g., valence and arousal) for affective music composition. Here, we discuss two broad classes of data-driven systems for AMG: (1) Hidden Markov model-based systems and (2) neural network-based systems.

4.2.1 Hidden Markov Model-Based Systems. AI-AMG systems have used **Hidden Markov models (HMMs)** for selecting chord progressions for composing affective music [78, 96, 97]. Specifically, researchers [96, 97] have used an HMM method for selecting a chord sequence to accompany a given melody for composing happy and sad music pieces. The transition probability matrices, linking chord groups (major/minor), are used to select the next chord in a sequence for a given emotion. The validation result from the user listening study in [96] demonstrated that chords assigned to melodies using the HMM receive similar subjective ratings as chords assigned manually by musicians. However, it is noteworthy here that both systems [96, 97] are designed and validated for their potential to compose music that can convey only two discrete levels of emotion (happy and sad).

Similarly, HMM-based approaches have also been incorporated into hybrid AI-AMG systems [10] in which the next chord is selected using an HMM-based chord transition model. This AI-AMG system is discussed in Section 4.5. Later, Park et al. [83] proposed a system that used **Hidden semi-Markov Models (HSMMs)** for manipulating the emotional content in a synthesized music/song. The system controls the emotion content, represented by an Energy-Tense model (2D plane), by changing the duration and vibrato parameters of the synthesized song. The results from the listening study show that the duration parameter is more effective than the vibrato parameters in representing an emotion.

Based on the human-based assessment results/listening studies mentioned in [96, 97], HMM-based methods seem to be efficient in encoding affective information in music and are particularly useful for selecting chords to express the desired affect. HMM-based methods can also be a potential option for designing hybrid AI-AMG systems. However, these HMM methods are only used to control the chord progression feature, and the efficacy of this method for handling/tailoring other musical features has not been fully explored. Secondly, designing an HMM for musical features is both time-consuming and computationally expensive, as it requires a large musical dataset for constructing the transition matrix for an HMM. In the next section, we summarize the AI-AMG systems that have used a neural network–based architecture to compose affective music.

4.2.2 Neural Network–Based Systems. Many AI-AMG systems have used a neural network–based architecture for generating affective music. The majority of these systems have used a (1) **recurrent neural network (RNN)**[30] (e.g., **gated recurrent unit (GRU)** [18], LSTM [43]), or a (2) Transformer-based architecture as the core architecture of the system. Further, in terms of the algorithm used for training, some of these systems have used (1) generative adversarial methods (e.g., GAN) or (2) variational methods (e.g., VAE [60]) to construct the loss function. In fact, some of the systems have even combined them, i.e., VAE and GAN, to develop VAE–GAN systems.

RNNs are sequential memory architectures [30], defined by weights ($\mathbf{w}$) and bias ($\mathbf{b}$), with recurrent connections used to model time-series data such as music (e.g., a melody as a sequence of notes). The input to an RNN is a musical sequence $\{\mathbf{x}_{t-n}\}_{n=0}^{N_i}$, where $\mathbf{x}$ represents the musical elements (e.g., notes), t denotes the current time index, and N_i denotes the input sequence length. The output of the RNN is $\{\mathbf{x}_{t+n}\}_{n=1}^{N_o}$, where N_o denotes the output sequence length. During training, the RNN adjusts its parameters ($\mathbf{w}$ and $\mathbf{b}$) to minimize a loss function, i.e., $L(\mathbf{w}, \mathbf{b})$, such that the probability $P(\{\mathbf{x}_{t+n}\}_{n=1}^{N_o} | \{\mathbf{x}_{t-n}\}_{n=0}^{N_i}, v)$, where v represents the emotion parameter, is maximized for a likely event and minimized for unlikely events in the sequence. LSTM architectures are a variant of RNNs designed to address the vanishing gradient problem by incorporating specialized memory cells and gating mechanisms, which preserves long-range dependencies in the time series. Thus, during training, along with $\mathbf{w}$ and $\mathbf{b}$, the activations of the memory cells are also being updated. A brief description of the systems with LSTM and an RNN are given below.

The use of the LSTM architecture [43], a specific type of RNN, has a long history of generating artificial music. The LSTM architecture has memory storing/recalling capabilities that are useful for the orderly generation of musical features in time. Thus, it has been used for generating music with a coherent long-term structure [17]. For example, Zhao and colleagues [117] have developed an AI-AMG system that uses a bi-axial LSTM architecture with a lookback module for tackling the problem of long-term structure in music. In the bi-axial model, the time-axis and pitch-axis are used to train the model on all of the possible pitches at each timestep, while the lookback module is introduced to reinforce the relationship between bars for yielding structural coherence in music. In another system, Madhok et al. [70] use a double-stacked LSTM architecture in conjunction with a **convolutional neural network (CNN)** for sentiment-specific affective music composition. In the first step, the CNN is used to extract sentiments from the input facial images. Subsequently, based on the sentiment information, the LSTM network composes affective music. In both of these systems, information related to emotion is fed to the network as a global condition in the form of emotional vectors (one-hot vectors) to compose affective music.

Rather than training an LSTM model using a matrix representation (e.g., piano roll), some LSTM-based AI-AMG models use a representation inspired by the domain of Natural Language Processing, i.e., music-coded tokens that represent musical events or a combination of musical features, such as tempo, chord, and duration [34, 71], in order to frame affective music generation as a language modeling problem. An AI-AMG system by Ferreira and Whitehead [34] uses a variant of the standard LSTM architecture, namely, **multiplicative long short-term memory (mLSTM)**, trained with music-coded texts. The trained network composes affective music based on the intended sentiment, represented as levels of arousal and valence. Similarly, Makris and colleagues [71] have developed an AI-AMG system for generating valence-conditioned lead sheets with an LSTM (as well as a transformer) sequence-to-sequence architecture. The sequence-to-sequence framework takes the musical features as input and "translates" high-level emotional features such as valence, along with time signature, grouping, and density, into lead sheets (chords and melody). The output from the AI-AMG system is a lead sheet whose chord progression reflects the desired valence.

LSTMs are a special form of RNNs optimized to avoid the vanishing gradient problem. Traditional RNN architectures have also been used as the core element of AI-AMG systems. RNNs have been used alone [118] or in combination with other models, such as a CNN [95], to generate emotion-specific affective music. For example, Zheng and colleagues [118] proposed an RNN-based AI-AMG system in which musical features such as pitch histogram and note density are used for training the network. These features can be considered to represent mode and tempo in music; altering these features can lead to a change in the emotional content of the music. Exploiting this concept, the AI-AMG system uses these features for training the network instead of using a large music dataset with emotion labels. Later, in the generation phase, the system could generate music to express four discrete emotions: happy, sad, tension, and peaceful. An RNN-based AI-AMG system has also been used in video-based applications [95], in which a CNN and RNN are used together in a sequence for generating emotion-carrying background music for videos with emotion annotations. The CNN is used to extract emotion-related visual features from the input video. Subsequently, the RNN (with GRU) is used to generate music based on the emotion information in the visual features.

Transformer-based architectures [45, 106] are another important method used for composing affective music. Transformer-based architectures enable parallel processing of time-series (only true during training for Transformers data, e.g., musical elements) and can be trained faster than RNNs/LSTMs. The transformer preserves the long-range dependencies in music by using an attention mechanism, in which the sequential data (e.g., contextual data such as music) is tokenized to create a database of tokens. The self-attention weighs the importance of different elements in a sequence when producing a representation of each element. It involves three sets of vectors: the query ($\mathbf{q}$), key ($\mathbf{k}$), and value ($\mathbf{v}$). The attention scores are calculated by measuring the similarity between the query and key vectors. The value vectors represent the information associated with each element, and they are weighted by the attention scores to produce the final output. This mechanism enables the model to focus on relevant parts of the input sequence, making it particularly effective for tasks involving long-range dependencies and capturing contextual information. Specifically, in the domain of music generation, the MIDI values are tokenized to form a database in which **k**, **q**, and **v** are then computed. In the case of affective music, the emotion elements are also tokenized as events in the database [82]. Transformer architectures have proven to be effective in AMG. Some of the state-of-the-art examples are discussed below.

The transformer architecture, for example, transformer-XL [82], has been used to compose affective music. The transformer-XL model has been used to generate music with controllable emotion in which the network takes MIDI-derived events (REMI), extracting chord information and MIDI events from music as input along with the emotional information (as tokens) [82]. Further, the work by Ji and Yang [53] used a vanilla transformer-based encoder-decoder architecture, exploiting a hierarchical latent representation to generate music according to valence specified at the piece level and bar level. In addition, Hung et al. [50] proposed a new multimodal (audio and MIDI) dataset named *EMOPIA*, and used a Compound Word Transformer (CP-transformer) architecture to compose affective music from scratch.

In addition to different architectures, different loss function formulations have been used for AMG, namely, variational methods and adversarial min-max game-based methods, popularly known as VAE and GAN methods, respectively. The variational method is realized through an encoder-decoder architecture, and minimizes **ELBO (evidence lower bound)** by minimizing the reconstruction loss and KL divergence term to minimize the difference between the data distribution in latent space and a standard distribution together. For example, Ji and Yang [54] approach emotion-conditioned melody harmonization by using a variational loss function that maximizes the conditional probability of a chord sequence (x) given the melody sequence (y) and emotion parameter (s). The GAN method is also used for affective melody generation [49]. The generative adversarial method uses a discriminator (D) and generator (G) model as well as a min-max loss function in which the D maximizes and G minimizes the loss, respectively. Mathematically, this loss function can be written as $\underset{G}{Min}\underset{D}{Max}\ \mathbb{E}_{X\sim P(X)}[logD(X)] + \mathbb{E}_{Z\sim P(Z)}[log(1 - D(G(Z))]$ where P(X) and P(Z) are the marginal distribution of actual data and noise, respectively, and $\mathbb{E}$ represents the expectation operator. The noise here is a Gaussian distribution and a required input for the generator G that is used to capture the diversity of the data distribution during training. We summarize the state-of-the-art methods using VAE and GAN next.

A popular neural network architecture used by the AI-AMG systems is the variational autoencoder and generative adversarial network (VAE-GAN) duo [47, 84]. The VAE-GAN networks together form a sequence-to-sequence architecture connecting an encoder and a decoder/generator in series. During the training phase, the model takes music pieces, annotated with the expression of emotion, as input. These emotion labels are fed to the network as the condition. Later, during the generation phase, the model generates affective music with the desired emotions. AI-AMG systems with a VAE-GAN model have been deployed to compose music with emotion tags from discrete emotion domains and different parts of the valence-arousal space.

In addition, standalone GAN and VAE architectures have been leveraged for AI-AMG systems. For example, Huang et al. [49] used a GAN for generating emotion-specific melodies. Similarly, the VAE architecture is also used independently in AI-AMG systems [40, 103]. Tiraboschi and colleagues [103] used the VAE model to design a system that can generate affective music in a closed-loop manner by sensing the user's emotional state through real-time EEG recording. Here, the user's emotions, decoded from the EEG signals, are used to drive the affective music generation process. Even simple architectures such as feed-forward neural networks have been used for composing affective music [112]. For example, Williams et al. [112] use a feedforward neural network model trained with musical features, namely, timbre, key, pitch spread, tempo, and envelope. The model is calibrated and updated by adjusting these musical features to maximize the spread of the emotional element of the generated music piece across the valence-arousal plane. The proposed system can generate affective music expressing a wide range of emotions. Later, this AI-AMG system was integrated into a closed-loop EEG system, in which the user's current emotional state was used for controlling the AMG process [23].

Evaluation of the efficacy of these neural network-based systems for composing affect-specific music may be conducted through user-based listening studies, by statistical comparison with existing music [71], or using classification methods [50]. The results from these studies and comparative analysis of the data confirm that the neural network-based AI-AMG systems are effective in generating affective music to express desired emotions. In fact, these systems are deployed in various fields, including human–computer interaction, AI movie creation, and emotion mediation for creating applications such as (1) design of EEG-integrated closed-loop music-based emotion mediation systems [112], (2) AMG in open-loop mode [80], (3) generation of music for emotional videos/games [95], (4) creation of audio accompaniment [70, 95], and (5) creation of X-reality [103]. However, the training of these systems requires huge computational resources and a large corpus of labeled musical data. Again, fine control of the emotion content in the music generated by these systems is often challenging. In video applications, it is also challenging for these systems to create music that can facilitate a smooth transition between different emotions. To mitigate the issue of the requirement of the large labeled dataset, researchers have tried to train the network with emotion-carrying musical features such as pitch histogram, and note density instead of training the network with a labeled musical dataset [118]. Even though the proposed method could limit the requirement related to data size to some extent, in the future, researchers could aim to devise more sophisticated approaches for handling this issue along with the other challenges. Some of the potential approaches for handling these issues are also mentioned in Section 6. In addition to rule-based and data-driven methods, researchers have deployed optimization-based methods for generating affective music, which are detailed in the next section.

4.3 Optimization Method–Based Systems

Systems using optimization methods for composing affective music can be categorized into two main types: genetic algorithm-based systems and tree/graph traversal optimization method-based systems. The detailed sub-categorization and the state-of-the-art are presented in the next section.

4.3.1 *Genetic Algorithm-Based Systems.*

AI-AMG systems have used a combinatorial optimization technique known as *genetic algorithms*. These systems formulate the task of AMG systems as an optimization problem that aims to find an optimal set of musical events or a set of generative musical rules (optimum weight for a set of musical rules) for composing emotion-specific music. Broadly, based on the approach used by these AI-AMG systems for updating the fitness function, the GA can be of two types. In the first type, the fitness function is a mathematical equation used for optimizing the musical rules. For instance, an AI-AMG system by de Azevedo Santos et al. [27] used the GA method (referred to as "Conventional GA" in this article) to compose monophonic piano music with an emotion profile that matches the emotion profile of the input music template. The fitness function used here computes a distance metric (mathematical equation) that represents the difference in the emotional content of composed and input music pieces. The system generates music by controlling features such as scale, tempo, chord progression, melody, and harmony. Even though the system can generate pieces of music to match the emotion profile of the input music, the efficacy of the system was tested for only two discrete emotions: happy and sad [27].

Another type of AI-AMG system uses a variant of a genetic algorithm, called an *interactive genetic algorithm*. An interactive genetic algorithm uses the listener's subjective evaluation information in the fitness function of the GA. Briefly, AI-AMG systems in [115, 119] have used the **interactive genetic algorithm (IGA)** method for composing happy and sad music. Here, the core algorithms use musical rules and subjective evaluation (IGA) to construct the fitness function of the GA during the training phase. Later, the updated rules are used to compose affective music. The authors have conducted listening studies to test the efficacy of these systems for generating affective music. According to the results of these studies [115, 119], the IGA method is capable of encoding/conveying the desired emotion in the generated music. However, the efficacy of these systems for expressing a more diverse set of emotions (e.g., more than two emotions) has not yet been tested.

Some AI-AMG systems have also been designed to use composition styles (e.g., a Romantic Era Classical style), instead of a target emotion, to generate affective music. For example, Brown [8] offers a system for composing Romantic Era–style music for computerized games using a genetic algorithm. The algorithm takes the game characters and the properties of the game environment (such as props and environmental features) as the input, which are used to select the "Leitmotivs." The system then uses the input composition (Leitmotivs) to build its composition by changing the musical features, such as harmonic tension and formal regularities.

The aforementioned AI-AMG systems have shown that GA-based methods are efficient in generating affective music and, more importantly, that the IGA method is a powerful way of including users' perception-related feedback in the AI-AMG, which ultimately improves the system's ability to compose affective music. While the use of an IGA may create a more effective system, as it requires recursive evaluation of the system by listeners, it is also time-consuming and can be exhausting for listeners in the training phase [104]. Also, only a few AI-AMG systems have used a GA method, and the reliability of this method for composing convincing affective music has only been tested for two discrete emotions. Thus, further investigation of the efficacy of such GA-based methods for generating affective music, especially for multiple emotions, is warranted.

4.3.2 Tree/Graph Traversal Optimization Method-Based Systems. In addition to genetic algorithms, researchers have used other combinatorial optimization techniques such as tree/graph traversal techniques and dynamic programming-based methods to design AI-AMG systems. In these optimization techniques, a search-based approach is used for selecting the value of a musical feature, for example, selecting the next chord in a progression. These combinatorial optimization techniques have been used in hybrid AI-AMG systems in which the emotion-specific chord progression feature is chosen using a graph traversal/tree traversal method [64, 93, 94]. For example, the hybrid AI-AMG system by Scirea and colleagues [93, 94] was designed to compose affective music for an interactive gaming environment and uses the graph traversal technique for selecting an optimal chord sequence depending on the emotion presented in the game. Similarly, Kuo and co-investigators [64] have used the tree traversal method for generating emotion-specific optimal chord sequences in a hybrid AI-AMG system in which the other features are selected using different methods (details of these hybrid systems and the information regarding selecting other features are summarized in Section 4.5). In both of these systems, the composed music is programmed to express different emotions on the valence-arousal plane and provide smooth transitions between these emotions.

Other combinatorial optimization techniques, such as dynamic programming-based optimization, have also been used to design an AI-AMG system for composing user-specified 'emotion flow'–guided musical accompaniment [114]. Given a melody, the system generates musical accompaniment through the optimal selection of a chord and accompaniment pattern (rhythm, volume, and pitch range) using dynamic programming-based constrained optimization. Even though the system is able to generate affective music to convey various levels of valence and some quantized arousal levels, the effectiveness of such systems in generating music from different points on the valence-arousal plane and transiting between them is yet to be verified. Variable neighborhood search has also been used in an AI-AMG system, called MorpheuS, for generating music based on a given tonal tension profile [41]. This method is effective in generating music with a specified tension profile and can potentially be used for expressing emotion.

The validation results for these AI-AMG systems help confirm the efficacy of optimization methods in generating affective music. The advantage of AI-AMG systems with tree/graph-based methods is that they use a probabilistic selection method (for musical features) that is particularly beneficial in generating non-monotonic and creative affective music. However, their major drawback is that they require skilled researchers with knowledge of music to design the tree/graph structure or objective function that constitutes the core algorithm. Also, these tree/graph-based methods can only select one feature at a time. For the selection of multiple musical features to form a composition, more graph networks/other methods would be needed. However, only the chord progression feature has been manipulated using tree/graph-based methods; hence, investigation is warranted for other musical features. In the next section, we present different AI-AMG systems that use hybrid methods for generating affective music.

4.4 Hybrid Systems

Hybrid methods fundamentally stand on the idea that different musical features can be manipulated more efficiently by combining different algorithms/methods. These hybrid AI-AMG systems use a combination of rule-based, data-driven, genetic algorithm–based, or graph/tree-traversal methods to compose affective music.

One group of hybrid AI-AMG systems combines HMM methods with rule-based methods [46, 72, 85]. In these AI-AMG systems, the HMM method is used for the selection of musical chords, note length, and the octave range (the probability of playing a note from the same/different octave), whereas other musical features such as tempo, mode, and pitch range are determined using musical rules. For example, the hybrid AI-AMG system proposed in [72] has deployed a combination of a Markov model and a rule-based approach to compose affective music for a textual narrative segment. The textual narratives are first processed using sentiment analysis methods (deep learning models, support vector machine, naïve Bayes classifier) to extract emotion/mood information. Subsequently, this sentiment information is used as the input condition by the hybrid AI-AMG to compose sentiment-specific affective music. In another group of hybrid AI-AMG systems, instead of Markov models, different graph/tree traversal methods are used in combination with rule-based approaches [64, 92, 94]. In these systems, graph/tree traversal methods are used for selecting chord progressions in a polyphonic affective composition. For example, a hybrid system by Scirea and colleagues [92] has used a graph traversal procedure to select the chord progression and a rule-based approach to determine intensity, timbre, rhythm, and dissonance in the generated music. Specifically, for realizing the melody component in music, the system has deployed an evolutionary algorithm (named FI-2POP) which uses a multi-constraint optimization method for minimizing the violation of musical rules during the generation of a melody. The resulting AI-AMG system can generate musical pieces for an interactive gaming platform that conveys different emotions on the valence-arousal plane. The aforementioned hybrid AI-AMG systems are used to generate polyphonic affective music from scratch.

In addition to composing affective music from scratch, hybrid AI-AMG systems have been designed to create affective music based on a pre-composed melody [10, 13]. In this case, the pre-composed melody is fed into the system as input along with the emotion-related information. The system then creates the affective composition by generating the accompaniment for the given melody. For example, the hybrid AI-AMG in [10] adopted an HMM for chord selection and a nearest neighbor algorithm for harmonization and setting the instrumentation parameters to compose affective music based on a given melody. The resulting composition is created by combining mood-specific instrumentation, melody, harmonization, and chord progression.

Hybrid AI-AMG systems have also used neural network architectures in combination with rule-based methods [51] or genetic algorithms [56]. For instance, Hutchings and McCormack developed an AI-AMG system to compose adaptive affective music for a multi-agent game environment. Depending on the properties of the game environment (called *context*), the system first retrieves a pre-composed melodic theme from a dataset [51]. This context information and the emotion information are then used to manipulate musical features such as harmony, melody, and rhythm over the pre-composed melodic theme to create affective music. Based on the context, an RNN is used to select the harmonic and rhythmic content of the music, and the musical rules are used to select the melody of the composition. In a similar fashion, the AI-AMG developed in [56] combined a genetic algorithm method with a neural network architecture as well as musical rules to compose affective music by manipulating musical features such as pitch values, duration, and melody.

In general, hybrid approaches have proven effective for generating affective music. Based on the evidence collected from human-based assessment, these systems can generate music to reliably express different emotions (either discrete emotions or specified on the valence-arousal plane). These AI-AMG systems are used for applications such as adaptive music generation for gaming [51], accompaniment generation [13], and physiology-sensitive AMG [46]. Even though hybrid methods for generating affective music are well accepted in the research community, there are some limitations to this approach. One of the main drawbacks of these systems is that the rationale behind the choice of different methods for manipulation/selection of different musical features is not clearly stated. This challenge is known as the problem of “hybridization, “which is detailed in Section 5 with some possible solutions. The second major drawback in designing hybrid systems is the need for skilled researchers with knowledge of music theory as well as generative models. Specifically, as a part of the hybrid system, designing graph/tree traversal architectures, HMM and musical rules demands a deep understanding of music theory to properly construct the models.

In the next section, we provide a comprehensive overview of the algorithms used to create affective music, focusing on the musical features they employ to control and generate the affective music.

5 Recommendations for Designing Controllable Affective Music Generation Systems

In this section, we will briefly summarize the features that have been shown to be most reliable for encoding and expressing emotion information in music. We will also discuss the procedures that can be potentially employed for tailoring them. The general procedure for creating an AI-AMG system includes four main steps: (1) Establish research goals and functional application(s) of the system; (2) based on (1) and the capabilities of the research team, select appropriate methods and the AI-AMG system architecture; (3) implement the AI-AMG system, which should manipulate musical features in order to achieve the goals of (1); and (4) evaluate the efficacy of the system. Thereafter, the researchers should update and revise the system as needed according to the results of the evaluation. This review has focused on presenting the various architectures used for AMG, and discusses evaluation at some length in Section 3.3. We therefore focus on (3), selecting features for the AI-AMG implementation, in the remainder of this section.

5.1 Selecting Features for Affective Music Generation

Table 2 summarizes the important musical features for affective music generation and the aspects of emotion (valence and arousal) that they most commonly impact according to the literature. Various musical features have an impact on emotions. Different algorithms can harness (all of)

Table 2. Musical Features, Their Impact on Different Dimensions of Emotion (Valence and Arousal), and the Methods that Most Often Manipulate Each Feature for Affective Music Generation (Based on the Literature)

Feature	Method most frequently adopting the feature to control affect	Emotion component
Tempo	Rule-based/Neural Network/Interactive Genetic Algorithm	Arousal
Mode/Scale (major or minor)	Rule-based	Valence
Chord progression/Sequence	Tree traversal/Graph traversal/Hidden Markov model method	Valence
Instrument Volume	Rule-based	Valence and arousal
Rhythm	Rule-based	Arousal/Valence and arousal
Pitch Characteristics	Rule-based	Valence/Arousal

Note that while many non-rule-based systems manipulate rhythm, for example, this feature is not always explicitly utilized to influence the *affective* nature of the music in other approaches.

these features, but some algorithms may be able to more effectively control certain features than others and are therefore used by researchers more frequently. In order to present the features and their impact on emotion dimensions in a concise manner, we consider each feature independently and present their dependence on valence, arousal, or both, i.e., "valence" or "arousal," "valence and arousal." The relationship between individual features and discrete emotions is not included in the table because the information related to discrete emotions is often encoded into an affective composition by manipulating multiple features of the composition collectively (not individually).

We will now discuss the main features that are tailored by the majority of researchers to embed emotion in music. One often manipulated feature is tempo (beats per minute). Rule-based approaches are effective in harnessing the ability of tempo to convey emotional information [28, 107, 109]. The tempo of a composition [28, 109] is normally controlled based on the arousal value of the target emotion, in which the mapping between tempo and arousal is governed by an algebraic equation. Nevertheless, methods such as neural networks [112] and interactive genetic algorithms [115] have also effectively used this important feature. Moreover, in all of these systems, the rule of thumb for manipulating tempo is that faster tempos are associated with higher arousal.

Another very important feature in affective music composition is the mode of the composition, e.g., the scale (typically major/minor) in which the music is composed, which is strongly linked to emotional valence. A rule-based approach can be used for selecting the mode of an affective composition [100, 118], in which positive valence is associated with the selection of major chords and negative valence is associated with minor chords. A combination of mode and tempo has also been used by various researchers (e.g., [107, 112]) for encoding different combinations of valence and arousal in an AI-AMG system. Apart from selecting the scale/mode, the selection of a sequence of chords, or *chord progression*, is often cited as an important feature in affective composition, and the selection of the chord progression is usually based on valence [64]. Probabilistic approaches such as HMMs [96] and tree traversal/graph traversal methods [64, 94] have shown promise in selecting the chord progression to elicit particular emotional responses as well. It is noteworthy that these approaches are not only effective in selecting the next chord for the composition but also in potentially bringing some amount of variation into the composition due to their probabilistic nature.

The pitch/pitch register in music is also crucial for expressing particular emotions through music and can be defined by the selection of a certain pitch register or the allowed range for the pitch registers of multiple instruments. Pitch register can be selected based on valence [109] or arousal [114]. The literature shows that rule-based approaches for selecting the probability of a certain pitch register depend on valence levels [109], although a rule-based method for defining the allowed range of pitches has also been successfully deployed based on arousal [114]. Such methods can be potentially considered in the future for selecting pitch characteristics in affective music composition.

Instrument Volume (dynamics) is also considered to be a critical parameter for AMG, and can carry relevant information about both valence and arousal [64]. Rule-based methods have proven to be effective in selecting the instrument volume in a composition [64].

Rhythm is another important feature that can encode and convey affect in music composition. Rhythm is defined by relative note durations and their temporal organization [22]. Rhythms in a composition may be selected based on arousal levels alone [109] or combinations of valence and arousal levels [91]. In order to modulate rhythmic components, a simple rule-based approach (e.g., [109]) such as a probabilistic note-onset rule for given arousal values is used. In addition, a complex rule-based approach for generating Euclidean rhythms has been effectively used as the rhythmic component tailored to both valence and arousal values in an affective composition [94].

Above we discussed some potential features of affective music composition that can be tailored independently based on given emotion components (valence and/or arousal). However, the majority of affective composition systems are designed to compose music by simultaneously modifying multiple features for a given valence/arousal value or discrete emotion. In order to simultaneously modify multiple features, a combination of different approaches has been considered in the literature [64, 94]. Hybrid methods may be a powerful choice, as they can provide the flexibility of tailoring different features using their most suitable algorithms, and then generating the music by combining all of these features. This approach can be pursued further in the future with the aim of finding a suitable and feasible combination of features along with their tailoring algorithms for composing affective music. Apart from hybrid methods, data-driven (specifically neural network) approaches have also proven to be efficient and effective in handling multiple features at a time to compose affective music. In particular, sequential architectures such as RNN models have proven to be effective. Even though such architectures have worked well for generating music with a desired emotion, a great deal of work remains in terms of precisely controlling the emotional content of the music generated in this process [9]. This is often referred to as the “challenge of controlling the affect” of artificial music composition (a detailed discussion is in Section 6.2).

In the next section, we discuss the major challenges in designing computationally creative systems, with a focus on AI-based AMG systems, and will also discuss possible directions for improvement.

6 Discussion

6.1 Bottlenecks in State-of-the-Art Systems

In the previous sections, we discussed state-of-the-art methods used for AMG and summarized the AI-AMG systems that employ these methods. An overview is presented in Table 1. These AI-AMG systems have represented emotions on a valence-arousal plane or as discrete emotion sets. We care to mention here that there is a lack of comparability possible across systems because of the disparity in the selection of the (1) number of discrete emotions (and comparison of systems that use discrete vs. continuous emotions) and (2) the listener group in the validation study (if human validation was conducted). In addition, different types of music are generated across AI-AMG systems, which does not make comparison between systems straightforward. Further, demos of the music are not always available, and when examples are tested, they may be cherry-picked for evaluation [2].

The disparity in the choice of the number of discrete emotions can be referred to as “variability in the number of discrete emotions.” For example, some AI-AMG systems have targeted four discrete emotions while other AI-AMG systems tested only two (happy and sad). Systems capable of expressing more discrete emotions are arguably better at precisely tailoring the musical features; thus, they may be seen as superior to their counterparts. On the other hand, for the emotion

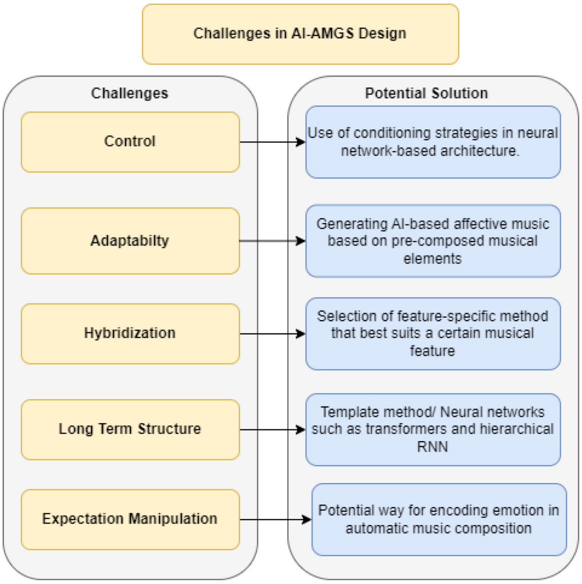
Challenges in AI-AMGS Design
Challenges
Potential Solution
Control
Use of conditioning strategies in neural network-based architecture.
Adaptabilty
Generating AI-based affective music based on pre-composed musical elements
Hybridization
Selection of feature-specific method that best suits a certain musical feature
Long Term Structure
Template method/ Neural networks such as transformers and hierarchical RNN
Expectation Manipulation
Potential way for encoding emotion in automatic music composition

Fig. 5. Challenges in designing AI-AMG systems.

evaluation process, the majority of these AI-AMG systems have relied on human study–based assessment methods. While human study–based assessments are advantageous due to their inherent ability to capture human perception (of emotion and creativity in music), there are often few criteria used to select the listener group (and human evaluation of musical emotion can be noisy). The listener groups used in validation studies can be characterized by high variability in the number of listeners, demographics of the listeners (age, gender, etc.), and prior musical training. This variation in the kinds of listeners used for assessment may be referred to as "variability in the listener group," This variability may have a significant confounding effect on the validation of AI-AMG systems. In the future, listening groups should be more standardized across studies. The comparisons presented in Table 1 are susceptible to the effects of these variabilities (selection of a discrete emotion set and the listener group used for validation).

In the next section, we discuss the major challenges and open research questions in designing a reliable AI-AMG system that is capable of generating realistic-sounding affective music.

6.2 Challenges

In this section, we discuss the open challenges in designing and implementing AI-based automatic AMG systems. These challenges are presented in Figure 5. We hope that this discussion will help readers who aim to design AI-based AMG systems that are capable of producing real-sounding, affective music using computational creativity.[5]

One of the major challenges in creating AI-AMG systems is "Control," which in this case refers to allowing the user to specify the desired emotional content of the generated music and the ability of the system to precisely control the musical features so that the resulting music exhibits the desired affect [9]. Unlike human composers, who are privileged to adapt musical patterns/ideas [9], AI-AMG systems only rely on the given parameters and information gained during training

[5] Computational Creativity is a multidisciplinary field that uses computational means to try to exhibit or simulate behavior that would be deemed creative in humans [9, 20].

to compose the music (the latter is especially true for data-driven systems/neural network systems). Thus, for such systems, controlling the features to produce desired output often becomes difficult due to limited transparency in the input, output, and their interdependencies [7]. In order to improve controllability, researchers have, for example, developed strategies that use conditional architectures. For instance, the data-driven architecture/neural network model may have an additional condition imposed during the training phase. Makris et al. [71], for example, proposed an AI-AMG system with high-level conditional information that is fed into the encoder of a sequence-to-sequence architecture for generating valence-specific affective music in a more controllable manner (e.g., the music is generated to match a profile of valence values provided by the user). However, a challenge in this direction is that datasets labeled with reliable affective information are scarce, although some researchers are working on this limitation [16]. Apart from using a conditional architecture, another potential approach to improve controllability is to use the fundamentals of reinforcement learning to train the neural network models [7]. Briefly, the idea is to frame the music generation process as a reinforcement learning problem, in which the objective function of the model is a combination of the objective function of the recurrent network along with some user-defined constraints. Such methods have proven to be beneficial in improving the controllability of automatic music generation systems [7] and can be considered in the future to improve the controllability of the emotional content of the generated music.

The second major challenge is "Adaptability," also called "Narrative Adaptability," which refers to the capability of AI-AMG systems to generate a coherent piece of music that can adaptively change based on a given narrative/sequence (emotional requirement of the narration) of a story. One of the potential applications for AI-AMG systems is to compose music that is able to match the dynamically changing events in the narrative/story-telling by adaptively controlling the emotion expressed in the generated music. It is often difficult, however, for the AI-AMG system to control/tailor the musical features to reliably convey the transitions between different events in the narrative. It is worth mentioning here that this task of generating adaptive music for narratives in real time cannot be accomplished using human-composed music. Thus, it can be seen as an interesting challenge for AI-AMG systems [9]. One potential solution to this problem is co-creativity (using a combination of human- and machine-generated music). For example, an AI-AMG system named "Mezzo" [8] developed by Daniel Brown composes romantic era–style music for video gaming. Here, a theme (Leitmotiv) related to an in-game scenario that involves particular characters and situations is composed by humans, and these themes are given as input to the system. Later, when the event is encountered in the game, the AI-AMG system interactively composes music by blending the human-composed theme with the computationally generated music in order to express the appropriate in-game situation and emotion. Another possible way of improving adaptability is to first understand the interplay between features with respect to different emotions, study how each specific feature can be independently tailored to convey a particular emotion, and then adaptively change a single (most reliable) feature based on the narration while keeping others constant. Such an approach, also proposed by [9], can allow AI-AMG systems to carefully adapt their emotional expression.

Other pertinent challenges in designing AI-AMG systems are hybridization, long-term structure, and manipulation of the listener's expectations. Hybridization refers to combining more than one technique for music generation in a single AI-AMG system. Even though researchers have deployed many different combinations of techniques (details in Section 4.5) for designing AI-AMG systems, the rationale behind the choice of these techniques and combining them is often unclear. In this regard, a more systematic approach can be helpful, in which the selection of different techniques for different musical features can be optimized based on their efficacy in generating affective music. The problem of long-term structure refers to the difficulty in generating longer excerpts that have an adequate amount of musical structure, such as repetitive patterns, in order to sound musically coherent across the composition. This issue of long-term structure is a general problem for automatic music generation systems and is not only limited to affective automatic music generation systems. In order to tackle this problem, researchers have, for example, developed an optimization-based approach that enforces generated pieces to follow a template for repeated patterns. This method is efficient in generating automatic music with long-term structure [41]. Other neural network approaches, such as transformer and transformer XL models, have also been shown to be more capable of preserving long-term structure. Such approaches may be helpful when translated and implemented into AI-AMG systems.

In addition, the manipulation of musical expectation in a composition can be seen as a viable method for eliciting different emotions in a listener. More precisely, the process of emotion induction via music can be seen to be fundamentally governed by a reward mechanism in the brain that responds to the realization and violation of musical expectations over time [14]. Therefore, another possible way of encoding emotion in music would be by algorithmically manipulating musical expectations in the piece. This could create a new avenue of research in the field of AI-AMG systems. Thus, finding a feasible algorithmic approach for explicitly manipulating the expectation in music can be posed as a future research direction.

7 Conclusion

In this article, we have presented a comprehensive review of controllable AI-AMG systems. The main components of AI-AMG systems were detailed, and different approaches used for designing these components were presented. Subsequently, we reviewed and summarized the state-of-the-art methods that have been deployed to develop reliable AI-AMG systems. For the reader's interest, we presented a set of important features to include in such systems, their relationship with arousal and valence, as well as previously used methods for controlling them. Finally, we summarized the challenges and open questions for the development of reliable affective automatic music generation systems. We hope that this review will be useful for readers who seek to understand the different AI-AMG systems that have been developed and to acquire an overview of the methods used for developing them, thereby aiding future exploration of this field. We hope that this review is also helpful to researchers entering this field as they frame their research questions for developing new AI-AMG systems.

Acknowledgments

This work was supported by the RIE2020 Advanced Manufacturing and Engineering (AME) Programmatic Fund (No. A20G8b0102), Singapore.

References

[1] Andrea Agostinelli, Timo I. Denk, Zalan Borsos, Jesse Engel, Mauro Verzetti, Antoine Caillon, Qingqing Huang, Aren Jansen, Adam Roberts, Marco Tagliasacchi, Matt Sharifi, Neil Zeghidour, and Christian Frank. 2023. MusicLM: Generating music from text. *arXiv preprint arXiv:2301.11325* (2023).

[2] Kat Agres, Jamie Forth, and Geraint A. Wiggins. 2016. Evaluation of musical creativity and musical metacreation systems. *Computers in Entertainment (CIE)* 14, 3 (2016), 1–33.

[3] Kat R. Agres, Adyasha Dash, and Phoebe Chua. 2023. AffectMachine-Classical: A novel system for generating affective classical music. *Frontiers in Psychology* 14 (2023), 1158172.

[4] Kat R. Agres, Rebecca S. Schaefer, Anja Volk, Susan van Hooren, Andre Holzapfel, Simone Dalla Bella, Meinard Müller, Martina de Witte, Dorien Herremans, Rafael Ramirez Melendez, Mark Neerincx, Sebastian Ruiz, David Meredith, Theo Dimitriadis, and Wendy L. Magee. 2021. Music, computing, and health: A roadmap for the current and future roles of music technology for health care and well-being. *Music & Science* 4 (2021), 2059204321997709.

[5] Anna Aljanaki, Yi-Hsuan Yang, and Mohammad Soleymani. 2017. Developing a benchmark for emotional analysis of music. *PloS One* 12, 3 (2017), e0173392.

[6] Chunhui Bao and Qianru Sun. 2022. Generating music with emotions. *IEEE Transactions on Multimedia* (2022).

[7] Jean-Pierre Briot and François Pachet. 2020. Deep learning for music generation: Challenges and directions. *Neural Computing and Applications* 32, 4 (2020), 981–993.

[8] Daniel Brown. 2012. Mezzo: An adaptive, real-time composition program for game soundtracks. In *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, Vol. 8.

[9] Filippo Carnovalini and Antonio Rodà. 2020. Computational creativity and music generation systems: An introduction to the state of the art. *Frontiers in Artificial Intelligence* 3 (2020), 14.

[10] Heather Chan and Dan Ventura. 2008. Automatic composition of themed mood pieces. In *Proceedings of the 5th International Joint Workshop on Computational Creativity*. Citeseer, 109–115.

[11] Wayne Chase. 2006. *How Music Really Works!: The Essential Handbook for Songwriters, Performers, and Music Students.* Roedy Black Pub.

[12] Ke Chen, Yusong Wu, Haohe Liu, Marianna Nezhurina, Taylor Berg-Kirkpatrick, and Shlomo Dubnov. 2024. MusicLDM: Enhancing novelty in text-to-music generation using beat-synchronous mixup strategies. In *ICASSP 2024—2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 1206–1210.

[13] Pei-Chun Chen, Keng-Sheng Lin, and Homer H. Chen. 2013. Emotional accompaniment generation system based on harmonic progression. *IEEE Transactions on Multimedia* 15, 7 (2013), 1469–1479.

[14] Vincent K. M. Cheung, Peter M. C. Harrison, Lars Meyer, Marcus T. Pearce, John-Dylan Haynes, and Stefan Koelsch. 2019. Uncertainty and surprise jointly predict musical pleasure and amygdala, hippocampus, and auditory cortex activity. *Current Biology* 29, 23 (2019), 4084–4092.

[15] Keunwoo Choi, György Fazekas, Kyunghyun Cho, and Mark Sandler. 2018. The effects of noisy labels on deep convolutional neural networks for music tagging. *IEEE Transactions on Emerging Topics in Computational Intelligence* 2, 2 (2018), 139–149.

[16] Phoebe Chua, Dimos Makris, Dorien Herremans, Gemma Roig, and Kat Agres. 2022. Predicting emotion from music videos: Exploring the relative contribution of visual and auditory information to affective responses. *arXiv preprint arXiv:2202.10453* (2022).

[17] Ching-Hua Chuan and Dorien Herremans. 2018. Modeling temporal tonal relations in polyphonic music through deep networks with a novel image-based representation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 32.

[18] Junyoung Chung, Caglar Gulcehre, KyungHyun Cho, and Yoshua Bengio. 2014. Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555* (2014).

[19] Amy Clements-Cortés. 2004. The use of music in facilitating emotional expression in the terminally ill. *American Journal of Hospice and Palliative Medicine* 21, 4 (2004), 255–260.

[20] Simon Colton, Geraint A. Wiggins, et al. 2012. Computational creativity: The final frontier?. In *ECAI*, Vol. 12. Montpelier, 21–26.

[21] Jade Copet, Felix Kreuk, Itai Gat, Tal Remez, David Kant, Gabriel Synnaeve, Yossi Adi, and Alexandre Défossez. 2024. Simple and controllable music generation. *Advances in Neural Information Processing Systems* 36 (2024).

[22] Ricardo Miguel Moreira Da Cruz. 2008. I-Sounds: Emotion-based Music Composition for Virtual Environments. Msc Thesis, Instituto Superior Técnico, Lisbon, 2008.

[23] Ian Daly, Asad Malik, James Weaver, Faustina Hwang, Slawmoir J. Nasuto, Duncan Williams, Alexis Kirke, and Eduardo Miranda. 2015. Identifying music-induced emotions from EEG for use in brain-computer music interfacing. In *2015 International Conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE, 923–929.

[24] Adyasha Dash, Anirban Dutta, and Uttama Lahiri. 2019. Quantification of grip strength with complexity analysis of surface electromyogram for hemiplegic post-stroke patients. *NeuroRehabilitation* 45, 1 (2019), 45–56.

[25] Adyasha Dash and Uttama Lahiri. 2019. Design of virtual reality-enabled surface electromyogram-triggered grip exercise platform. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 28, 2 (2019), 444–452.

[26] Adyasha Dash, Anand Yadav, and Uttama Lahiri. 2019. Physiology-sensitive virtual reality based strength training platform for post-stroke grip task. In *2019 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*. IEEE, 1–4.

[27] L. Rocha de Azevedo Santos, Carlos Nascimento Silla Jr, and M. D. Costa-Abreu. 2021. A methodology for procedural piano music composition with mood templates using genetic algorithms. (2021).

[28] Stefan K. Ehrlich, Kat R. Agres, Cuntai Guan, and Gordon Cheng. 2019. A closed-loop, music-based brain-computer interface for emotion mediation. *PloS One* 14, 3 (2019), e0213516.

[29] Dave Elliott, Remco Polman, and Richard McGregor. 2011. Relaxing music for anxiety control. *Journal of Music Therapy* 48, 3 (2011), 264–288.

[30] Jeffrey L. Elman. 1990. Finding structure in time. *Cognitive Science* 14, 2 (1990), 179–211.

[31] Zach Evans, Julian D. Parker, C. J. Carr, Zack Zukowski, Josiah Taylor, and Jordi Pons. 2024. Long-form music generation with latent diffusion. *arXiv preprint arXiv:2404.10301* (2024).

[32] Lucas Ferreira, Levi Lelis, and Jim Whitehead. 2020. Computer-generated music for tabletop role-playing games. In *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, Vol. 16. 59–65.

[33] Lucas N. Ferreira, Lili Mou, Jim Whitehead, and Levi H. S. Lelis. 2022. Controlling perceived emotion in symbolic music generation with Monte Carlo tree search. In *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, Vol. 18. 163–170.

[34] Lucas N. Ferreira and Jim Whitehead. 2021. Learning to generate music with sentiment. *arXiv preprint arXiv:2103.06125* (2021).

[35] Anders Friberg. 2006. pDM: An expressive sequencer with real-time control of the KTH music-performance rules. *Computer Music Journal* 30, 1 (2006), 37–48.

[36] Takako Fujioka, Jon Erik Ween, Shahab Jamali, Donald T. Stuss, and Bernhard Ross. 2012. Changes in neuromagnetic beta-band oscillation after music-supported stroke rehabilitation. *Annals of the New York Academy of Sciences* 1252, 1 (2012), 294–304.

[37] Juan Sebastián Gómez-Cañón, Nicolás Gutiérrez-Páez, Lorenzo Porcaro, Alastair Porter, Estefanía Cano, Perfecto Herrera-Boyer, Aggelos Gkiokas, Patricia Santos, Davinia Hernández-Leo, Casper Karreman, et al. 2023. TROMPA-MER: An open dataset for personalized Music Emotion Recognition. *Journal of Intelligent Information Systems* 60, 2 (2023), 549–570.

[38] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative adversarial nets. *Advances in Neural Information Processing Systems* 27 (2014).

[39] Alessandra Gorini, Claret S. Capideville, Gianluca De Leo, Fabrizia Mantovani, and Giuseppe Riva. 2011. The role of immersion and narrative in mediated presence: The virtual hospital experience. *Cyberpsychology, Behavior, and Social Networking* 14, 3 (2011), 99–105.

[40] Jacek Grekow. 2021. Generating musical sequences with a given emotion. In *2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, 1941–1946.

[41] Dorien Herremans and Elaine Chew. 2017. MorpheuS: Generating structured music with constrained patterns and tension. *IEEE Transactions on Affective Computing* 10, 4 (2017), 510–523.

[42] Dorien Herremans, Ching-Hua Chuan, and Elaine Chew. 2017. A functional taxonomy of music generation systems. *ACM Computing Surveys (CSUR)* 50, 5 (2017), 1–30.

[43] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural Computation* 9, 8 (1997), 1735–1780.

[44] Maia Hoeberechts, Ryan J. Demopoulos, and Michael Katchabaw. 2007. A flexible music composition engine. *Audio Mostly* (2007).

[45] Wen-Yi Hsiao, Jen-Yu Liu, Yin-Cheng Yeh, and Yi-Hsuan Yang. 2021. Compound word transformer: Learning to compose full-song music over dynamic directed hypergraphs. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 35. 178–186.

[46] Chih-Fang Huang and Yajun Cai. 2017. Automated music composition using heart rate emotion data. In *International Conference on Intelligent Information Hiding and Multimedia Signal Processing*. Springer, 115–120.

[47] Chih-Fang Huang and Cheng-Yuan Huang. 2020. Emotion-based AI music generation system with CVAE-GAN. In *2020 IEEE Eurasia Conference on IOT, Communication and Engineering (ECICE)*. IEEE, 220–222.

[48] Chih-Fang Huang and Wei-Po Nien. 2013. A study of the integrated automated emotion music with the motion gesture synthesis via ZigBee wireless communication. *International Journal of Distributed Sensor Networks* 9, 11 (2013), 645961.

[49] Renjie Huang, Yin Li, Da Kang, Yujie Chen, Chunyan Yu, and Xiu Wang. 2021. Melody generation with emotion constraint. In *Proceedings of the 2021 5th International Conference on Electronic Information Technology and Computer Engineering*. 1598–1603.

[50] Hsiao-Tzu Hung, Joann Ching, Seungheon Doh, Nabin Kim, Juhan Nam, and Yi-Hsuan Yang. 2021. EMOPIA: A multi-modal pop piano dataset for emotion recognition and emotion-based music generation. In *International Society for Music Information Retrieval Conference, ISMIR 2021*. International Society for Music Information Retrieval.

[51] Patrick Edward Hutchings and Jon McCormack. 2019. Adaptive music composition for games. *IEEE Transactions on Games* 12, 3 (2019), 270–280.

[52] Shulei Ji, Jing Luo, and Xinyu Yang. 2020. A comprehensive survey on deep music generation: Multi-level representations, algorithms, evaluations, and future directions. *arXiv preprint arXiv:2011.06801* (2020).

[53] Shulei Ji and Xinyu Yang. 2023. EmoMusicTV: Emotion-conditioned symbolic music generation with hierarchical transformer VAE. *IEEE Transactions on Multimedia* (2023).

[54] Shulei Ji and Xinyu Yang. 2023. Emotion-conditioned melody harmonization with hierarchical variational autoencoder. *arXiv preprint arXiv:2306.03718* (2023).

[55] Shulei Ji, Xinyu Yang, and Jing Luo. 2023. A survey on deep learning for symbolic music generation: Representations, algorithms, evaluations, and challenges. *Comput. Surveys* (2023).

[56] Minjun Jiang and Changle Zhou. 2010. Automated composition system based on GA. In *2010 IEEE International Conference on Intelligent Systems and Knowledge Engineering*. IEEE, 380–383.

[57] Patrik N. Juslin and John Sloboda. 2011. *Handbook of Music and Emotion: Theory, Research, Applications*. Oxford University Press.

[58] Saya Kanno, Takayuki Itoh, and Hiroya Takamura. 2015. Music synthesis based on impression and emotion of input narratives. In *Sound and Music Computing Conference (SMC2015)*. 55–60.

[59] Anna Kantosalo and Hannu Toivonen. 2016. Modes for creative human-computer collaboration: Alternating and task-divided co-creativity. In *Proceedings of the 7th International Conference on Computational Creativity*. 77–84.

[60] Diederik P. Kingma and Max Welling. 2013. Auto-encoding variational Bayes. *arXiv preprint arXiv:1312.6114* (2013).

[61] Alexis Kirke, Eduardo Miranda, and Slawomir J. Nasuto. 2013. Artificial affective listening towards a machine learning tool for sound-based emotion therapy and control. In *Proceedings of the Sound and Music Computing Conference*. Citeseer, 259–265.

[62] En Yan Koh, Kin Wai Cheuk, Kwan Yee Heung, Kat R. Agres, and Dorien Herremans. 2022. MERP: A music dataset with emotion ratings and raters' profile information. *Sensors* 23, 1 (2022), 382.

[63] Gunter Kreutz, Ulrich Ott, Daniel Teichmann, Patrick Osawa, and Dieter Vaitl. 2008. Using music to induce emotions: Influences of musical preference and absorption. *Psychology of Music* 36, 1 (2008), 101–126.

[64] Ping-Huan Kuo, Tzuu-Hseng S. Li, Ya-Fang Ho, and Chih-Jui Lin. 2015. Development of an automatic emotional music accompaniment system by fuzzy logic and adaptive partition evolutionary genetic algorithm. *IEEE Access* 3 (2015), 815–824.

[65] Elke B. Lange and Klaus Frieler. 2018. Challenges and opportunities of predicting musical emotions with perceptual and automatized features. *Music Perception: An Interdisciplinary Journal* 36, 2 (2018), 217–242.

[66] Dinh-Viet-Toan Le, Louis Bigo, Mikaela Keller, and Dorien Herremans. 2024. Natural language processing methods for symbolic music generation and information retrieval: A survey. *arXiv preprint arXiv:2402.17467* (2024).

[67] Elad Liebman and Peter Stone. 2020. Artificial musical intelligence: A survey. *arXiv preprint arXiv:2006.10553* (2020).

[68] Steven R. Livingstone, Ralf Mühlberger, Andrew R. Brown, and Andrew Loch. 2007. Controlling musical emotionality: An affective computational architecture for influencing musical emotions. *Digital Creativity* 18, 1 (2007), 43–53.

[69] Gang Luo, Hao Chen, Zhengxiu Li, and Mingxun Wang. 2022. Music generation based on emotional EEG. In *2022 the 6th International Conference on Innovation in Artificial Intelligence (ICIAI)*. 143–147.

[70] Rishi Madhok, Shivali Goel, and Shweta Garg. 2018. SentiMozart: Music generation based on emotions. In *ICAART (2)*. 501–506.

[71] Dimos Makris, Kat R. Agres, and Dorien Herremans. 2021. Generating lead sheets with affect: A novel conditional seq2seq framework. *arXiv preprint arXiv:2104.13056* (2021).

[72] Mehak Maniktala, Chris Miller, Aaron Margolese-Malin, Arnav Jhala, and Chris Martens. 2020. MINUET: Procedural musical accompaniment for textual narratives. In *International Conference on the Foundations of Digital Games*. 1–7.

[73] Brian McFee, Oriol Nieto, Morwaread M. Farbood, and Juan Pablo Bello. 2017. Evaluating hierarchical structure in music annotations. *Frontiers in Psychology* 8 (2017), 1337.

[74] Jan Melechovsky, Zixun Guo, Deepanway Ghosal, Navonil Majumder, Dorien Herremans, and Soujanya Poria. 2023. Mustango: Toward controllable text-to-music generation. *arXiv preprint arXiv:2311.08355* (2023).

[75] Gianluca Micchi, Louis Bigo, Mathieu Giraud, Richard Groult, and Florence Levé. 2021. I keep counting: An experiment in human/AI co-creative songwriting. *Transactions of the International Society for Music Information Retrieval* 4, 1 (2021).

[76] Kana Miyamoto, Hiroki Tanaka, and Satoshi Nakamura. 2020. Music generation and emotion estimation from EEG signals for inducing affective states. In *Companion Publication of the 2020 International Conference on Multimodal Interaction*. 487–491.

[77] Kana Miyamoto, Hiroki Tanaka, and Satoshi Nakamura. 2022. Online EEG-based emotion prediction and music generation for inducing affective states. *IEICE TRANSACTIONS on Information and Systems* 105, 5 (2022), 1050–1063.

[78] Dan Morris, Ian Simon, and Sumit Basu. 2008. Exposing parameters of a trained dynamic model for interactive music creation. (2008).

[79] Jun-Ichi Nakamura, Tetsuya Kaku, Kyungsil Hyun, Tsukasa Noma, and Sho Yoshida. 1994. Automatic background music generation based on actors' mood and motions. *The Journal of Visualization and Computer Animation* 5, 4 (1994), 247–264.

[80] Pedro Neves, Jose Fornari, and João Florindo. 2022. Generating music with sentiment using Transformer-GANs. *arXiv preprint arXiv:2212.11134* (2022).

[81] Renato Eduardo Silva Panda, Ricardo Malheiro, Bruno Rocha, António Pedro Oliveira, and Rui Pedro Paiva. 2013. Multi-modal music emotion recognition: A new dataset, methodology and comparative analysis. In *10th International Symposium on Computer Music Multidisciplinary Research (CMMR 2013)*. 570–582.

[82] M Aqmal Pangestu and Suyanto Suyanto. 2021. Generating music with emotion using transformer. In *2021 International Conference on Computer Science and Engineering (IC2SE)*, Vol. 1. IEEE, 1–6.

[83] Younsung Park, Sungrack Yun, and Chang D. Yoo. 2010. Parametric emotional singing voice synthesis. In *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 4814–4817.

[84] Zhaolin Qiu, Yufan Ren, Canchen Li, Hongfu Liu, Yifan Huang, Yiheng Yang, Songruoyao Wu, Hanjia Zheng, Juntao Ji, Jianjia Yu, and Kejun Zhang. 2019. Mind band: A crossmedia AI music composing platform. In *Proceedings of the 27th ACM International Conference on Multimedia*. 2231–2233.

[85] Adhika Sigit Ramanto and Nur Ulfa Maulidevi. 2017. Markov chain based procedural music generator with user chosen mood compatibility. *International Journal of Asia Digital Art and Design Association* 21, 1 (2017), 19–24.

[86] Judy Robertson, Andrew de Quincey, Tom Stapleford, and Geraint Wiggins. 1998. Real-time music generation for a virtual environment. In *Proceedings of ECAI-98 Workshop on AI/Alife and Entertainment*. Citeseer.

[87] A. Rory and Zbigniew W. Ras. 2007. Rules for processing and manipulating scalar music theory. In *2007 International Conference on Multimedia and Ubiquitous Engineering (MUE'07)*. IEEE, 819–824.

[88] James A. Russell. 1980. A circumplex model of affect. *Journal of Personality and Social Psychology* 39, 6 (1980), 1161.

[89] Klaus R. Scherer. 2005. What are emotions? And how can they be measured? *Social Science Information* 44, 4 (2005), 695–729.

[90] Flavio Schneider, Ojasv Kamal, Zhijing Jin, and Bernhard Schölkopf. 2023. Moûsai: Text-to-music generation with long-context latent diffusion. *arXiv preprint arXiv:2301.11757* (2023).

[91] Marco Scirea. 2013. Mood dependent music generator. In *International Conference on Advances in Computer Entertainment Technology*. Springer, 626–629.

[92] Marco Scirea, Peter Eklund, Julian Togelius, and Sebastian Risi. 2017. Can you feel it? Evaluation of affective expression in music generated by MetaCompose. In *Proceedings of the Genetic and Evolutionary Computation Conference*. 211–218.

[93] Marco Scirea, Julian Togelius, Peter Eklund, and Sebastian Risi. 2016. Metacompose: A compositional evolutionary music composer. In *International Conference on Computational Intelligence in Music, Sound, Art and Design*. Springer, 202–217.

[94] Marco Scirea, Julian Togelius, Peter Eklund, and Sebastian Risi. 2017. Affective evolutionary music composition with MetaCompose. *Genetic Programming and Evolvable Machines* 18, 4 (2017), 433–465.

[95] Gwenaelle Cunha Sergio and Minho Lee. 2021. Scene2Wav: A deep convolutional sequence-to-conditional SampleRNN for emotional scene musicalization. *Multimedia Tools and Applications* 80, 2 (2021), 1793–1812.

[96] Ian Simon, Dan Morris, and Sumit Basu. 2008. MySong: Automatic accompaniment generation for vocal melodies. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 725–734.

[97] Amani Indunil Soysa and Kulari Lokuge. 2010. Interactive machine learning for incorporating user emotions in automatic music harmonization. In *2010 5th International Conference on Information and Automation for Sustainability*. IEEE, 114–118.

[98] Joanna Stewart, Sandra Garrido, Cherry Hense, and Katrina McFerran. 2019. Music use for mood regulation: Self-awareness and conscious listening choices in young people with tendencies to depression. *Frontiers in Psychology* 10 (2019), 1199.

[99] Bob L. Sturm. 2013. Evaluating music emotion recognition: Lessons from music genre recognition?. In *2013 IEEE International Conference on Multimedia and Expo Workshops (ICMEW)*. IEEE, 1–6.

[100] David Su, Rosalind W. Picard, and Yan Liu. 2018. AMAI: Adaptive music for affect improvement. In *ICMC*.

[101] Serkan Sulun, Matthew EP Davies, and Paula Viana. 2022. Symbolic music generation conditioned on continuous-valued emotions. *IEEE Access* 10 (2022), 44617–44626.

[102] Lijun Sun, Li Hu, Guiqin Ren, and Yufang Yang. 2020. Musical tension associated with violations of hierarchical structure. *Frontiers in Human Neuroscience* 14 (2020), 397.

[103] Marco Tiraboschi, Federico Avanzini, and Giuseppe Boccignone. 2021. Listen to your mind's (He)art: A system for affective music generation via brain-computer interface. In *Sound and Music Computing Conference*. SMC, 146–153.

[104] Nao Tokui and Hitoshi Iba. 2000. Music composition with interactive evolutionary computation. In *Proceedings of the 3rd International Conference on Generative Art*, Vol. 17. 215–226.

[105] Bo-Wei Tseng, Yih-Liang Shen, and Tai-Shih Chi. 2021. Extending music based on emotion and tonality via generative adversarial network. In *ICASSP 2021—2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 86–90.

[106] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in Neural Information Processing Systems* 30 (2017).

[107] Sandrine Vieillard, Isabelle Peretz, Nathalie Gosselin, Stéphanie Khalfa, Lise Gagnon, and Bernard Bouchard. 2008. Happy, sad, scary and peaceful musical excerpts for research on emotions. *Cognition & Emotion* 22, 4 (2008), 720–752.

[108] P. Vishesh, A. Pavan, Samarth G. Vasist, Sindhu Rao, and K. S. Srinivas. 2022. DeepTunes-music generation based on facial emotions using deep learning. In *2022 IEEE 7th International Conference for Convergence in Technology (I2CT)*. IEEE, 1–6.

[109] Isaac Wallis, Todd Ingalls, and Ellen Campana. 2008. Computer-generating emotional music: The design of an affective music algorithm. *DAFx-08, Espoo, Finland* 712 (2008), 7–12.

[110] Isaac Wallis, Todd Ingalls, Ellen Campana, and Janel Goodman. 2011. A rule-based generative music system controlled by desired valence and arousal. In *Proceedings of 8th International Sound and Music Computing Conference (SMC)*. 156–157.

[111] Klaus C. Wassermann, Kynan Eng, Paul F. M. J. Verschure, and Jônatas Manzolli. 2003. Live soundscape composition based on synthetic emotions. *IEEE MultiMedia* 10, 4 (2003), 82–90.

[112] Duncan Williams, Alexis Kirke, Eduardo Miranda, Ian Daly, Faustina Hwang, James Weaver, and Slawomir Nasuto. 2017. Affective calibration of musical feature sets in an emotionally intelligent music composition system. *ACM Transactions on Applied Perception (TAP)* 14, 3 (2017), 1–13.

[113] Duncan Williams, Alexis Kirke, Eduardo R. Miranda, Etienne Roesch, Ian Daly, and Slawomir Nasuto. 2015. Investigating affect in algorithmic composition systems. *Psychology of Music* 43, 6 (2015), 831–854.

[114] Yi-Chan Wu and Homer H. Chen. 2016. Generation of affective accompaniment in accordance with emotion flow. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 24, 12 (2016), 2277–2287.

[115] Bin Xu, Shangfei Wang, and Xian Li. 2010. An emotional harmony generation system. In *IEEE Congress on Evolutionary Computation*. IEEE, 1–7.

[116] Yi-Hsuan Yang, Yu-Ching Lin, Ya-Fan Su, and Homer H. Chen. 2008. A regression approach to music emotion recognition. *IEEE Transactions on Audio, Speech, and Language Processing* 16, 2 (2008), 448–457.

[117] Kun Zhao, Siqi Li, Juanjuan Cai, Hui Wang, and Jingling Wang. 2019. An emotional symbolic music generation system based on LSTM networks. In *2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)*. IEEE, 2039–2043.

[118] Kaitong Zheng, Ruijie Meng, Chengshi Zheng, Xiaodong Li, Jinqiu Sang, Juanjuan Cai, and Jie Wang. 2021. EmotionBox: A music-element-driven emotional music generation system using Recurrent Neural Network. *arXiv preprint arXiv:2112.08561* (2021).

[119] Hua Zhu, Shangfei Wang, and Zhen Wang. 2008. Emotional music generation using interactive genetic algorithm. In *2008 International Conference on Computer Science and Software Engineering*, Vol. 1. IEEE, 345–348.

Received 16 January 2023; revised 3 May 2024; accepted 27 May 2024