Limitations of Emotional Expression in AI-Generated Popular Music Creation

Ming Ni, Qingyu Yao *

School of Design and Art, Jiangxi University of Finance and Economics, Nanchang, Jiangxi, China

* Corresponding Author: Qingyu Yao

Abstract. The usage of artificial intelligence (AI)-generated music technology in popular music production has gained higher popularity. It has been noted though that there is a difference in the intensity of emotional conveying of music generated by AI as compared to the human creation of music. The human beings express the nuanced moods and sentiments in the context of popular music creation through various elements in which human beings strive to create the song structure in ways that are conducive to their tastes, the layers of texture, and the choice of various timbers, thus carrying out the deep meanings of the popular music. Conversely, music AI-generated does not do so in most cases. The paper is dedicated to the constraints of emotional expression in AI-generated popular music and adopts the research method based on the combination of musicological theory and analysis to disclose the problems of the lack of emotional expression methods, low adaptability to context, and inability to adjust cultural element in AI-generated popular music. This study intends to offer musicological theoretical explanation of maximization of the emotional expression of the AI popular music generation technology and encourages improvement of the humanistic aspects of the artificial intelligence in the domain of music production.

Keywords: Music Artificial Intelligence; Music Creation; Emotional Expression; Creation Limitations.

1. Introduction

Music Artificial Intelligence Music Artificial Intelligence (MusicAI) (Source: https://baike. baidu. com/ item/AI/9180?fr=ge_ala) is a type of computer-based system, which uses the technology of artificial intelligence, and can mimic human music-generation abilities. The methods mostly employed in this technology include machine learning and deep learning to learn by studying a large amount of music data based on human feedback and through self-iteration and constant training can be trained to eventually produce music on its own or critique and assess the existing music independently. This is the fundamental process of MusicAI, as the original approach to simulating and extending human musical creativity, which includes the systems of knowledge related to a multitude of disciplines, such as musicology, computer science, and artificial intelligence, encompassing such key areas as music theory, digital signal processing, machine learning, and the deep learning. Regarding music production, the use of MusicAI has shown extensive opportunities in use, including automatic work on melodies, harmonies, and arrangements, seeking to make music creation more efficient and productive in the quality of work and offer a fresh methodological approach to music creation and analysis. [1]

2. Artificial Intelligence in the Sphere of Human Popular Music Creation

The essence of the popular music composition consists in the synthetic synthesis of various instruments, timbre materials and sound effects to form a timbre peculiarity and comparatively homogeneous musical ambiance. Recording creators usually assemble the temporal arrangement and the amplitude of the melody were forming several rhythmic parts (drum sets, bass, guitar, and others), and then forming individualized timbre textures and mastery layers, using synthesizers, samplers, and other sound filters. It is important at this point to combine melodic lines with the use of chosen timbres, instruments, and effect elements in the design of the arrangement to increase its expressiveness and the sound content of the music and introduce peculiar artistic coloration to the piece.

Once the arrangement phase is accomplished, the creation portion progressively moves to to the stages of recording, mixing, and mastering. Sound collection is performed with the help of professional recording equipment, and then the process of mixing is performed to provide the vocal balance and space shaping and the final step of the process is the mastering to improve the professional listening experience and quality of release of the whole audio. It is worth marking that the execution order of every link of popular music creation is to some extent flexible, and it can be varied dynamically, depending on the creative habit of a creator and the real needs. These connections are not independent of each other and depend on one another affecting each other, as they compose the prototypical workflow of the popular music production. In this concept, an artist will be able to advance his or her creative potentials by constantly practicing artistic and technical approaches and creating more attractive musical works. [2]

The essence of AI music generation is in the creation of a complicated computation system, which can be used to recreate the whole process of the popular human song development. The process is fundamentally a multi-level, multi-module system engineering (iterative, collaborative) and expressed as an essential architecture as a closed-loop creative pipeline with its simplistic structure.

To begin with, during the first input stage, the system has to undertake creativity requirement analysis and setting of goals. According to natural language processing technology, the system is able to recognize the creative will of the users via text command or dialogic interactions, and also, it can incorporate the massive market data analysis to explain critical parameters like work style orientation, the tone of emotion, target audience, and the application scenario.

Secondly, the system goes to the data preparation and preprocessing phase. This stage is based on large-scale high quality music databases and includes structured MIDI data (data on pitch, duration, and chords) and audio waveform data. In case it is related to the creation of lyrics, some text processing tasks like word segmentation, emotional annotation, and prosodic analysis cutoff should be done on text data to facilitate further semantic and style modeling.

The third step is the model selection and architecture design. The system may be configured to suit various algorithmic paradigms depending on the task of interest. As an illustration, massive AI models such as ChatGPT are capable of identifying and analyzing certain features of creators, and hence, producing lyric pieces that satisfy the needs. The patterns of such interaction corroborate dialogue process with more accomplished lyricists, in which the user is able to develop creative ideas and taste in greater detail by asking questions, and the system automatically refines the output generated in response. This system has the capability not only to assist creative inspiring works of professional artists but also to create high-quality text lyrics, an essential feature of AI to comprehend and produce syntactic in music composition.

The fundamental creative computational processes in AI music generation are melody and lyric generation. During the melody generation stage, the models usually create melody motives, which follow musical grammar rules (like scales and tonality) by methods that may include sampling, interpolation or optimization in the latent space, and expand them into structurally complete musical phrases. Lyric producing has to strike a balance between literary expression and formal strictness, semantic coherency, and emotional customization and counterbalancing rhyming patterns, syllable counts, and corresponding requirements with the melodic contours to generate phonological coordination.

the arrangement and orchestration module is supposed to deepen the voices of the music and the expression of the timbre. Its work is to automatically set harmonic progressions of melodies depending on preset format labels and to pick the right virtual instruments (drum set, bass, piano, etc.) to play to make the music more expressive and stylistic. During the last optimization and export phase, the system or humans undergo a refined processing on the output work such as fine-tuning and assessment of the harmonic richness, rhythmic dynamics, and mixing effects, and export the finished work as a standard audio format file.

The whole process represents the methodological features of the data-driven and knowledge-guided integration, the core intention of which is the systematic externalization of the implicit knowledge (in human music creation) and its automated replication by mathematical modelling.

3. The Analysis of Comparative Semantics of Emotions in the Creation of Popular Music.

Yi Fu Zhi Ming (***In the Name of the Father***) is a seminal piece of works in the musical life of ***Jay Chou***, having taken a significant place in the history of the Chinese popular music because of its high artistic level and complexity. Created by Jay Chou, with the words by ***Huang Junlang***, it is a smash of dramatic movie narration, classical opera samples, and a gloomy magnificence of composition, the piece greatly disrupting the traditional artistic paradigms of the Chinese popular musique during that period of time. Not only is it the culmination of the personal style of the Jay Chou as the style of a dark aesthetic but raises the expressive level of the popular songs in musical texture and depth of feelings due to the forward-looking artistic conception.

When it was released, the song premiered in more than fifty radio channels throughout Asia and was named ***JAY Z*** by the listeners as a sign of the social acceptance and impact within the industry. The iconic importance of ***"Yi Fu Zhi Ming"*** is that it was able to combine comparatively niche artistic experimentation with pop culture, pushing the limits of the musical expression meanwhile proving that even commercial art may be very ideologically rich and artistically meaningful. It is against this reason that the work has remained popularly considered to be a representative/research worthy classic as far as the Chinese popular music sector is concerned.

This below discussion will use the ***Yi Fu Zhi Ming*** as a case study, through an analytical approach to musicology, to discuss how its musical language is involved and constitutes the deep emotional semantics under the presence of such aspects as the song structure (e.g., introduction, verse, chorus, etc.), textural layers (e.g., interactive relationships between main melody, counter-melody, harmonic layer, rhythmic layer, and bass layer), timbral configuration, and dynamic processing.

3.1 Macro Structure: Movie Story Mondrain.

The song does not rely on a conventional versechorus cycle format of popular music, but instead follows a dramatic building format of a five act play: Prelude-Narration-Conflict-Transcendence-Emotional Resonance. The prelude that is 1 minute and 26 seconds of time is a full story: the sound of the pipe organ and violin strings together with the textual background of the Latin formula: Lord pray with me, is the Gothic religiously solemn composition. The high register (E5 C6) operatic soprano arias at 1 minute and 3 seconds abruptly combine with the timbers of the synthesizers imitating the sound of police sirens creating a deadly acoustic clash between, on the one hand, sacred space and on the other hand, the criminal scene. This juxtaposition of tones and dramatic contrast method does not only produce a considerable level of auditory tension, but also supports dramatic shifts in the story. The song suddenly ends as they have a 5 minutes and 34 seconds of pure silence at the end of the song. This emotional resonance treatment does not only expand the imaginative space on the work but also raises the gangster narrative, in particular, to metaphysical ideas of fate and redemption, enhancing the tragic connotations of the whole work. [3]

3.2 Three-Dimensional Texture: Multiple Narration Polyphonic.

The arrangement has a polyphonic form of construction based on advanced soundfield design, penetrated with three layers of imagery: "religious narrative," "gang conflict," and "inner monologue." The verse section presents the pronounced and sustained rap soundtrack of ***Jay Chou*** with melodious aria performed by the operatic soprano, and it is a counterpoint, the conflict between human passion and divine will, a fine intuitively charting about the moral conflict and questioning of belief of the protagonist. The left-right channel separation mixing effect is also used in the work in the section of the coda, The mottled family crest (4:50 5:20): in the former section, the left channel

is focused on the development of the plot musical discussion, and the channel is compatible with an inner monologue, which is characteristic of the fragmented psychological state of the character at the audio level in the latter.

The rhythmically and harmonic sustained bass of the pipe organ (C2 pitch all along the piece) is the irrevocable fatalistic background that adds to the tragic colour of the work as well as its integrity. The rhythmic patterns of triplets and dense eight notes used alternately are also effective in controlling the tension in music and the speed of the narration and give a balance of dynamics between the tension building and the discharge of emotions to the whole piece, respectively, this is also attributed to the high dramatic and auditory integrity within the piece.

The song develops the relationships of the voices by creating multifaceted dialogic relations, depending on the polyphonic technique. The primary melody and the accompaniment voices do not merely have a hierarchical character but each of them has their own narrative functions and are bonded and reverberated with each other. This stratified musical sound is not only an enrichment of the auditory imagery of the piece, but also gives more substantial coloring of emotional expressiveness. As an example, in part, the bass line progresses slowly, which indicates the time flow and the absence of a rewinding of the fate, and in another part, the inner anxiety is conveyed by quick ornamental lines performed by the treble voice. Their conflict creates an active balance, which helps the audience to understand the conflicting nature of the character and his multidimensional feelings. Besides, the high-level polyphonic framework provides the song with higher narrative logic, in the sense of every voice layer presenting independent tale, that eventually fuses into a whole musical spectacle.

3.3 Symbolic Timbre: Aesthetics of Cross Cultural Collage.

This collage art and timbre archaeology across time are successfully realized in the arrangement of Hong Jingyao. The core narrative of the whole song usages the classical timbers: the low sounds of the German Klais church organ denote the power of Godfathers, the strumming of Flamenco guitar with the help of a method called Rasgueado foreshadows the motives of revenge, and the polyphony of the baroque piano raises the question of the ability to be redeemed. Electronic sound effects create the listening experiences in film industry, and they total audialization of the visual signifiers of gangster films via the methods of music concrete. This collage of timbre across cultures is as much a show of how the arranger knows how to utilize the variety of musical languages but essentially a reverberation of power, violence and prejudice of gangster movies. The utilization of the narrativeness of the traditional tools with the creation of the atmosphere of the modern electronic sound effects helped the work to merge the boundaries of classical and popular, eastern and western, creating a very original sounding. In such a way, it maintains the seriousness of classical music in addition to introducing the appeal of popular music, which enables the listeners to experience the art defamiliarization effect in the familiar musical environment.

3.4 Precise Dynamics: Figuring Emotional Arc.

With its accurate dynamic composition and sound shaping, the whole song creates a subtle psychological development process of stifling whispers to absolute bursting out. In the verse section, the vocals are purposefully subdued and the so-called proximity effect of the microphone is also skillfully applied to make one feel the presence, as they overhear confidential faces in an confession booth; when it comes to the chorus part, the so-called parallel compression technology finds an immediate way to push the vocal, strings and drums dynamic levels (pun intended) about -9dBFS and brighten deliberately at the frequency band of 5kHz, and the main text of the song is the following: When you are in the hide-out, the song There are as accurate the entry and exit timing of instrumental arrangements, after gun shots the pipe organ fades out in a linearly decaying manner, which is symbolical of the gradual disintegration of belief systems; there are also trembling long notes in reverberant tails, which creates an ethereal and sad mood and everyone is returned to silence. This enables layer dynamic processing which does not only result in a high level of control in audio

production but also full narrative narrative of suppression to release in feelings expression. In particular, during major transitional passages of emotional fluctuations, with subtle use of automation and frequency moulding, it provides a sense of tension and anticipation to the atmosphere of the song that makes the listeners feel that they are sucked into the emotional whirlwind of the inner world of the character.

All the inclusiveness of the notes, the orchestration of the sounds, and the dynamic variations, which are present in Yi Fu Zhi Ming, only service their monumentalization of narrative and philosophical guesswork. Not only does it become the culmination of Jay Chou, as an individual effort in his artistic aesthetics, but also lays a model piece combining technical rationality and artistic maintainability to Chinese popular music with its exquisite craftsmanship and rich expressivity of emotions. A landmark case of AI music commercialization happened in February 2025 when the single "Seven-Day Lover" written by programmer Yang Ping (addressed online as Yapie Programmer Brother) using AI tools achieved a hit in NetEase Cloud Music. The song did not even take a week to enter the NetEase Cloud Music Rising Chart TOP 10 upon release, and ended up going up to more than 2 million plays in addition to more than 4,600 comments and the offer of copyright priced a five-figure bid. It was popularized with comparatively advanced prompt engineering (users had to ask it to give a simulation of the Jay Chou style and a seven-day love story, the text produced contained golden sentences such as Ice Americano is too bitter, need to add a half-spoonful of my compromise). It took just 2 hours, disrupting the old music production paradigm and demonstrating that even the most music foundation-deprived people can also provide music with the help of AI and the entry into the music industry has been reduced tremendously.

Nevertheless, the novel ***Seven-Day Lover*** is certainly an eye-opener to the emotional constraints of the creation of AI. First of all, its lyrics are criticized as being based on "canned images" (as moonlight on the clothesline like wind chimes), not based on actual experience and originality of human creators, implying that AI is just a formulaic rearrangement of existing databases, as opposed to actual expression of human feelings. Secondly, despite its technical completion, which was awarded the Gold Award at the "Born for Audio AI" competition, most of its comments revolved around the topic of AI creation, not the music itself (with 72 percent of them discussing the subject matter of the AI creation), a factor of the communication attributes of concentrating more on the technical gimmicks than the musical changes in response (with respondents also noting that it had strong attributes of an internet bubblegum pop). Moreover, AI constraints were even more vivid in the cases of controversies within the industry: several musicians criticized AI as depriving music of human warmth and afterwards, when an imitation hit the scene, namely Eight-Day Lover, the data reduced drastically, proving that even the technical stacking of AI cannot keep the audiences permanently on their feet. The artistic depth needs to be incorporated into the human creativity and that is the design of narrative structures with the disintegration and synthesis of natural human emotions.

The emergence of AI music can mean the music industry will change, but the example of the ***Seven-Day Lover*** shows that the present AI is not able to access the bottleneck of genuine human emotions. It is produced by algorithmic blending of the styles already in existence, not true artistic creativity, even as the Grammy clause that AI works must show evidence of meaningful human authorship, technology must come to bond with mankind to move beyond tool qualities and attain the jump between viral hit and classic work.

The dynamic variations in music composition are essential mediums of expressing emotions. Although AI is accurate on how to compute shifts in volume, rhythm, and frequency, it is not very good at attributing these new shifts with the logic of nature. To illustrate, during the crescando processing of the chorus departments or emotional changes in bridges the works produced by AI tend to sound too mechanical without transitions and sense of hierarchy. This drawback makes the listeners unable to immerse themselves in understanding the emotional aura that the music generates. Moreover, the creation of emotional arcs using AI is more of a template rather than the adaptable strategy depending on the actual conditions, which also brings the AI and the human creator even

further apart. In a bid to have more emotional coloring, AI has to transcend technical simulation and venture into more humanistic creative routes. [4]

4. Limitations Analysis and Future Directions

4.1 A. Technical Limitations in Expression of feeling.

These shortcomings of AI-generated music in expressing emotions can be examined using two aspects of technical simulation and cognitive experience. Technically, AI is capable of re-creating the surface emotional attributes by parameter changes, including the dynamics and tempo changes, but it can hardly make the creator of the micro-trim of emotional subtext within a human voice or performance. The more basic bottleneck is cognitive: according to the Book of Music, all music is created by the heart of a human, having made it clear that music is, in fact, a projection of human feelings and life experience. Devoid of real emotional experiences such as love, loss and joy, the creation of AI is limited exclusively to the analysis of data patterns that leads to the creation of works that are beautiful, but empty. As an example, when humans compose love song they incorporate emotions and life experiences of the love and include intricate emotional undertones, but AI generated versions combine generic words and tunes to imitate human emotion in a robotic way. On April 4, 2023, a study published on the official website of the University of York confirmed that AI-generated music has a much lower score on aesthetic pleasure metrics than human ones. The result of this emotional insufficiency is evident in the audience acceptance about 70% of the respondents are ready to give AI music a chance, and only 40% of them think that this kind of music may express real feelings.

4.2 B. Cultural Elements Adaptation Barriers.

The field of artificial intelligence has gone a long way in creation of popular music, especially the formulaic building of melodies, harmonies, and rhythms where artificial intelligence can replicate hits in high degree. Nevertheless, when it comes to the issue of historical implementation and cultural identity, the limits of AI are made very noticeable. The cause of the problem is in the fact that the cultural elements are not expressed in measurable data but dynamic emotions and definite meanings that lie in the specific social-historical circumstances. When dealing with these true cultural features, there is an inevitable challenge faced by artificial intelligence of the following:

4.3 Data Reliance and Situational Unavailability.

The concept of AI learning completely is based on the present data, but cultural factors exist in the context of the world of beyond the data. To illustrate, AI is able to learn the iconic reggae offbeat, deep basslines, and even produce works that are very close in timbre and structure; however, it cannot comprehend the religious zeitgeist of Rastafari that spawned a rhythm or the burden of the colonial past; AI-generated reggae is an empty rhythmic shell with no moving humanistic information and credibility attached to it, without support in the geographical-cultural context. These products without spirit are culturally anaemic and they do not appeal to particular groups of people.

Here is another example of traditional Chinese music: AI is able to imitate pentatonic scales, pipa tones or gongche notation, but fails to understand aesthetic principles that are based on the Eastern philosophy such as art of negative space and vital energy. it is unable to grasp how blindness and social storm of Abing has changed into the desolating vibrato of ***Erquan Yingyue***, nor can it consciously unite the pure, subtle, calm, distant, spirit sought in literature music. This cultural gap in the manifestation of the culture usually dooms AI works to the fallacy of symbolic stratification-apparently culturally typical, but disconnecting profound links between sound, history, and personal destiny, and reducing to rootless simulations.

4.4 The Cultural Elements are Misinterpreted and Simplified.

The intervention of AI in the present popular music production has elicited renewed consideration of the adaptability of cultural symbols. In popular music, identity, atmosphere, and emotional attitudes of humans are often expressed through cultural symbols: dialects, slangs, ethnic stereotyped tones of instruments, pentatonics, or opera. Although AI is able to rapidly recognize these symbols in question using the pattern recognition method, it fails to draw value and insight into cultural nuances as well as the inner heart of human feelings and emotions, which frequently results in abuse and inappropriate collage. One example, it may unnaturally jazz up the supposedly heartfelt erhu sounds with the rhythms of the electronic dance music, or introduce the rough dialect in futuristic structures of arrangement, creating an auditory and cultural logical disruption. Even though it is technically novellish, this type of combination seems culturally dishonest, fake, and even offending to the viewers with different cultures, thus forming what is referred to as the so-called plastic-like aesthetic boundary. Adaptation to culture requires delicacy and accuracy, but the approach to big data mixing is not always authentic by AI since it ignores context and feelings behind symbols.

4.5 Future Directions

In perspective, it must be noted that the development trajectory of AI in the domain of popular music creation will potentially be conceived as human-machine cooperation instead of seeking completely autonomous content generation. Its core value is the fact it is used as an auxiliary means of motivating creativity, efficiency, and creative dimensions. In this regard, technical development should start with creation of more advanced and multimodal databases, not only audio samples, but structured cultural background texts, images, analyses of social emotions, etc, to give models greater richness in cultural contextual knowledge.

Based on this, AI systems might aid creators with suggestions of the elements of their style that they utilize it in their use, the explanation of the historical origin, and even the warning thematic of the taboo of using the cultural elements without providing it with the source, yet in the process of creating art, the creators will be more aware of its cultural components and the intensity of expressionism.

Moreover, ethical standards that are shared between developers and platforms should be used to encourage unified use of AI in music making. Cultural annotation and identification of training materials and signals of the origin of cultural styles in AI-made works might be taken specifically, as evidence of respectful attitude towards other cultures; at the same time, minorities and endangering cultures might be modernized and promoted through AI algorithms which would promote cultural exchange of exchange instead of elimination of disagreements.

Finally, the basic feature of AI in the process of producing music is instrumental. Its meaning is not a substitute of human beings in the process of culture creation but rather an accelerator and amplifier of culture expression. Indeed, in a perfect human-machine, collaborative model: AI processes the big data and pattern recognition, and human creators are concerned with the injection of cultural knowledge, emotional sensitivity, and critical thinking. It is only at that point that popular music would have real emotions and soul of the times and its final compelling force would be traced to the human soul that possesses the cultural perception and experience of life.

Acknowledgments

The above paper is from the "Chinese Music History and Appreciation of Masterpieces" course alliance of Jiangxi Province.

References

[1] Xu, C. L. (2024). Research on AI Training and Application in Chinese Style Popular Music Composition. Cultural Industry, (32), 142-147.

[2] Zhou, Y. X. (2024). Research on the Application of Artificial Intelligence Technology in Popular Music Creation. Grand View (Forum), (12), 27-29.

[3] Jiang, R. J. (2024). Research on the Application of Music Artificial Intelligence in Popular Song Creation (Doctoral dissertation, Nanjing University of the Arts). DOI:10.27250/ d. cnki. gnjyc. 2024.000251.

[4] Edimakor Mac V4.2.0 Ignites a New Era of AI Music Creation with ElevenLabs Integration[J].M2 Presswire, 2025.