**ISSN** 1816-6075 (Print), 1818-0523 (Online)
**Journal of System and Management Sciences**
Vol. 14 (2024) No. 2, pp. 467-481
DOI:10.33168/JSMS.2024.0229

# Emotion-Driven Music Composition using AI and User Feedback

Hyeong Gyun Kim

Professor, Department of Software, Kookmin University, Korea

*multikim@kookmin.ac.kr*

**Abstract.** This study developed an AI music composition system that generates compositions based on users' emotional responses to artworks. An online art class platform enabled users to input emotions via a keypad while viewing artworks. The AI system analyzed the emotion inputs, selected appropriate musical templates, and generated MIDI music scores. Evaluation using METEOR and BLEU metrics on 50 test compositions showed average scores of 0.385 and 0.305, indicating appropriate alignment with reference compositions.

**Keywords:** AI composition, emotion-keypad, automatic composition, personalized music

# 1. Introduction

The representation of emotions in the field of art and music is important because it shapes the user's experience and reaction to creative expression. Various studies have been conducted on artistic practices related to the process of collecting and communicating these emotions (Kim, 2023; Clancy, 2023). Prior studies have explored emotion recognition and representation in music (Vuoskoski and Eerola, 2011) but few have focused on generating compositions based on user emotion inputs.

In recent years, the advent of Artificial Intelligence (AI) has presented a variety of studies on the integration of emotion, art, and music (Briot et al, 2020; Roig-Francolí, 1995). This paper studied AI-based composition and focused on how to create music based on user emotions from art experiences (Carnovalini and Rodà 2020).

AI-based composition offers a unique opportunity to bridge the gap between different art disciplines (Civit et al, 2022). Developing an AI composition system that can interpret and translate the emotions evoked by art as musical expressions can improve the way individuals express their experiences and reactions to art creations (Chaudhary et al, 2021). This paper sought to study the relationship between emotion, art, and music in order to develop an AI composition system capable of generating music that reflects the user's emotional response to art.

This paper proposes the integration of AI technology with an online art teaching environment. Through this platform, users participate in the art appreciation process and express their emotions in real time using an emotion keypad. The AI composition algorithm analyzes and interprets the transmitted emotion signals to select the most appropriate musical template, and generates music that collects the essence of the expressed emotion through an integration and coordination process between templates. This approach creates a harmonious fusion of art, emotion, and music, providing a customized experience for users participating in the system (Kim, 2023).

The aim was to develop an AI composition system that can interpret user emotions from art experiences and generate corresponding musical scores, and evaluate its performance using METEOR and BLEU metrics.

Throughout this dissertation, the relevant literature on emotion, art, and music, were explored. Their interconnections were examined and the existing AI composition techniques were investigated. Empirical evaluations were conducted to assess the performance and efficiency of the proposed AI-based composition system, utilizing existing metrics such as METEOR and BLEU (B. Harini and N.Thirupathi Rao, 2019). The results of these evaluations will contribute to the ongoing development and improvement of the system.

# 2. Related Research

## 2.1. Music and emotion

Music has long been recognized for its effect on evoking human emotions (Simões et al, 2022). Research in this area focuses on understanding the mechanisms by which music evokes emotional responses, the relationship between musical features and emotions, and the subjective characteristics of the emotional experiences that music evokes (Carnovalini and Rodà, 2020).

Music has the unique ability to evoke and influence emotions and has been used throughout human history as an important medium to express and communicate emotions. Therefore, it is important to understand the relationship between music and emotion in order to develop AI composition systems that can effectively collect and communicate emotional responses through music (Simões et al, 2022).

Many researchers have studied the mechanisms by which music elicits emotional responses (Kim, 2022). Musical characteristics such as tempo, rhythm, melody, harmony, and dynamics play an important role in shaping emotional experiences. For example, fast, lilting music with vibrant rhythms and key tones tends to evoke feelings of happiness, excitement, and energy. And slow, melancholic melodies with subtle tonalities often elicit feelings of sadness, longing, and reflection (Carnovalini and Rodà, 2020).

The emotional impact of music is not only determined by its acoustic characteristics, but also by personal factors such as personal taste, cultural influences, and past experiences. Different individuals may have different emotional responses to the same music, depending on their unique backgrounds and associations. Cultural context also shapes the emotional interpretation of music, as certain musical styles or patterns may relate to specific emotions in other cultures (Clancy, 2023).

The subjective nature of music-induced emotional experiences adds complexity to the task of creating AI composition systems that accurately collect and represent emotions (Harini and Thirupathi Rao, 2019). Researchers have studied individual differences in emotion recognition while investigating factors such as personality traits, mood states, and cognitive processes that influence the interpretation of music-induced emotions (Clancy, 2023).

Music has the ability to convey emotions and allow listeners to immerse themselves in a world of other emotions. It can evoke memories, trigger thought, and amplify or regulate emotions (Simões et al, 2022). Music also plays an important role in social and cultural contexts that facilitate the expression and sharing of emotions among individuals or within communities (Kim, 2022).

Understanding the relationship between music and emotion can help design AI composition systems that generate music that empathizes with and communicates to specific emotional states. By analyzing patterns in musical structure and fusing emotional cues, AI composition systems can generate compositions that effectively convey desired emotional responses, providing a powerful tool for individuals to express, explore, and communicate their emotions through music (Kim, 2022).

## 2.2. AI composition

Developments in AI and machine learning are opening up new possibilities for music production. AI systems can analyze vast musical databases, learn patterns, and create unique compositions. Research in this area presents a variety of AI techniques used in music composition, including rule-based systems, Generative Adversarial Networks (GANs), and deep learning approaches. Understanding these techniques is important in developing AI composition systems that can effectively collect and represent emotional responses to art (Harini and Thirupathi Rao, 2019).

In music composition, AI uses computer algorithms and machine learning techniques to analyze musical patterns, learn existing compositions, and create new musical material. The algorithms allow the system to analyze vast amounts of musical data, identify patterns, and make predictions based on the learned patterns. This allows the AI composition system to create compositions that match a given musical genre and style or are inspired by a particular composer (Harini and Thirupathi Rao, 2019).

Rule-based systems are one approach used for AI composition. Such systems rely on a predefined set of rules, such as music theory principles or compositional techniques, to generate music. According to these rules, AI composition systems can create compositions that conform to specific stylistic constraints or structural guidance. However, rule-based systems lack flexibility and may have difficulty collecting the nuances and complexity of musical expression (Harini and Thirupathi Rao, 2019).

GANs are commonly used in AI composition. A GAN consists of two neural networks: a generator and a discriminator. The generator generates new musical sequences and the discriminator evaluates the generated sequences and provides feedback. The iterative process of training and feedback allows the GAN to create compositions that exhibit compelling musical characteristics and mimic the style of the training data (Roig-Francolí, 1995).

LSTM networks (Hochreiter and Schmidhuber, 1997) can learn sequential patterns in data and generate consistent musical sequences. By analyzing the relationship between musical elements and structure, deep learning models can demonstrate musical consistency and generate music that follows established compositional patterns (Fomicheva and Specia, 2019).

One of the challenges of AI composition based on user emotion is integrating emotional aspects into the generated music. While AI systems are good at learning patterns and creating structurally consistent compositions, collecting and expressing emotions through music is more complex. Recent research has focused on integrating emotion recognition algorithms into AI composition systems. By analyzing features related to emotion expression, such as intensity, timbre, and melodic contour, AI systems can generate compositions that reflect specific emotional states (Roig-Francolí, 1995).

AI composition based on user emotion is about creating music that not only follows predetermined rules and patterns, but also elicits an emotional response and creates empathy for the listener. This requires a deep understanding of the interplay between musical structure and emotional experience. By developing an AI composition system that can effectively analyze and integrate emotional signals, songs that collect and express emotional responses to art can be created, allowing users to communicate and share their emotional experiences through art and music.

The motive is the smallest unit that has independent value among the components of music, and a four-word composition consisting of two motives is called a phrase. The musical section in which the two phrases are combined is called a PERIOD. The period is the smallest unit of completion in a piece. In the one-period, two-period, and three-period forms, each form means that a song is composed of one period. The structure of the two-part form consists of two large bad verses of eight verses (16 verses) (Roig-Francolí, 1995).

Figure 1 is the score of "The Goose" in two period forms. It can be seen that the large musical section A consists of small musical sections a and a', and the large section B consists of small sections b and a'.

In this study, the 2-period form was used as the basic structure of a song for the automatic composition function utilizing the emotional key.


Fig. 1: Composition of the song 'Goose'

## 2.3. Evaluation metrics

This paper proposed the BLEU (Bilingual Evaluation Understudy) and METEOR (Metrics for Evaluation of Translation with Explicit Ordering) algorithms as methods for evaluating compositions generated through artificial neural networks.

BLEU is a typical algorithm for evaluating the quality of machine translated content (Evtikhiev, M. et al, 2023). It evaluates the quality of translation by comparing machine-translated sentences with human-translated sentences. It has a nodal value between 0 and 1, with closer to 1 meaning more similar to the reference sentence. The measurement is based on n-grams, is not language-bound, and has the advantage of high calculation speed.

METEOR, like BLEU, is an algorithm for evaluating machine-translated content with the sentence as the basic unit (Fomicheva and Specia, 2019). A machine-translated sentence is called a Hypothesis and a human-translated sentence is called a Reference. METEOR first creates an alignment by concatenating unigrams that match between the Hypothesis and the Reference. At this time, a unigram in Hypothesis can be linked to at most one unigram in Reference. The alignment with the fewest number of connection crossings is selected among the connected alignments generated in this way.

Evaluation metrics are needed to quantify and compare the performance of AI compositions. Music composition involves the structure and flow of music, and METEOR is used to evaluate the order agreement between a translated sentence and a reference sentence. Music composition can also be evaluated by comparing the structure and flow of composed and reference music fragments.

BLEU can be compared by measuring the degree of agreement between a musical work composed in music composition and a reference musical work. While BLEU began with a metric that evaluated textual sentence structure and lexical selection, music has structure and patterns. The matching of musical structures and patterns between music composition AI products and reference music can be evaluated with BLEU.

# 3. AI composition system based on user emotion

## 3.1. Configuration of the AI composition system

Figure 2 shows the concept of the user emotion-based AI composition system proposed in this study. An online art class is provided through a web server, where a commentator introduces an art work. The user would listen to the commentary and input in real-time the emotions he/she recalls about the query presented using an emotion keypad. The result of the user's emotional expression will be turned into music by the AI server via an automatic music composition function and provided to the user.

The Web server will manage content consisting of artwork descriptions and queries, and will serve as a relay between the user and the AI server. The user uses an emotional keypad connected to a smart device with Wi-Fi support. The device would load artwork commentary content from a web server.

The user would listen to the commentary about the artwork and select an emotional adjective key in response to a situational question. When the commentary is finished, the emotion adjective key values selected for each situation are sent to the AI server via the web server.

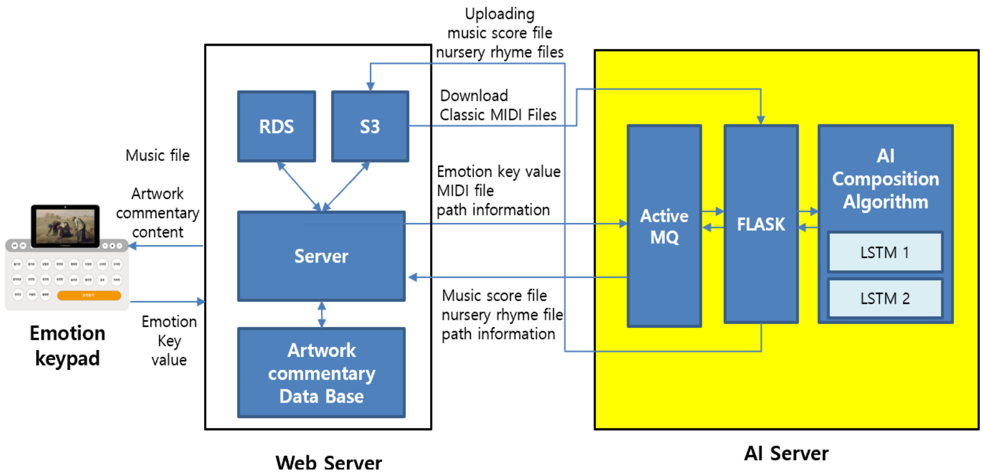


Fig. 2: Construction of an AI composition system based on user emotions

### 3.2. Structure of art commentary content

The artwork commentary content provided by the AI server is composed in the form of a video consisting of an image of the artwork, background music, commentary, and query audio, as shown in Figure 3.

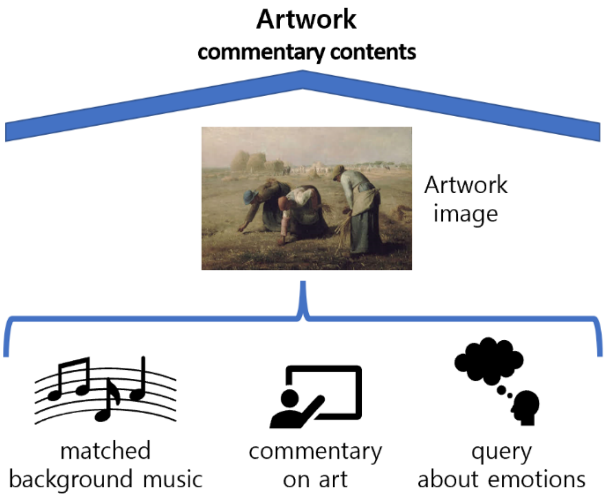


Fig. 3: Elements of art commentary content

The query content for the artwork commentary and emotion was composed in four steps to have a relatively simple "introduction, development, turn, and conclusion" structure to be utilized for the compositional information for automatic composition.

Table 1 shows the structure of the questions regarding sentiment for each stage of artwork commentary. The artwork commentary was divided into four stages, and four inquiries were composed for each stage. This was done in order to organize the structure of the piece in the form of a headnote.

Table 1: Structure of a query by step-by-step commentary

| commentary stage | query number | query content |
|---|---|---|
| Introduction | 4 | Commentary and questions about the historical background of artworks |
| Development | 4 | Commentary and questions about the artist's influence on art history |
| Turn | 4 | Commentary and questions about external elements such as colors and lines of artworks |
| Conclusion | 4 | Commentary and questions about the meaning and metaphor of artworks |

In the Introduction part, the commentary on the historical background of the work is provided in four stages, with each stage presenting a query about the content of the commentary, and the learner is asked to select his or her feeling about the query.

In the Development section, the artist's influence on art history in general and the artist's life history at the time of the work's creation are explained in four stages, and questions about the explanatory content are posed at each stage, and the learner chooses his or her feelings about the questions. The learner chooses the emotion for the question.

In the Turn part, the artist's work is described in four stages, each stage being followed by a question about the content of the description, and the learner is asked to choose the emotion he or she feels in response to the question.

In the final section, "Conclusion," learners are asked to consider the meaning of the work and to think about what the metaphor is.

## 3.3. Emotion-keypad design

For AI composition based on user emotion, the user's emotion must first be entered. In this study, the user listens to a description of a work of art and uses the emotion - keypad to select emotion adjectives for situational questions.

The emotion adjectives used here were composed as shown in Table 2, based on an analysis of adjectives mainly used in art and music (Kim, 2022).

Table 2: Emotional adjectives commonly used in art and music

| **music and art** | | | | **music** | |
|---|---|---|---|---|---|
| **adjective** | **emotion key** | **adjective** | **emotion key** | **adjective** | **emotion key** |
| lively | Liv | windless | Wnl | nectarous | Nec |
| joyful | Joy | amazing | Amz | magnificent | Mag |
| intense | int | doubtful | Dou | pious | Pio |
| relaxed | Rex | fear | Fer | | |
| happy | Hap | boring | Bor | | |
| fond | Fon | static | Stc | | |
| occult | Ocu | gloomy | Glo | | |
| elegant | Ele | unpleasant | Unp | | |

Sixteen of the emotional adjectives used here are commonly used in art and music, with three emotional adjectives used only in music: sweet, grand, and pious. The emotional adjectives composed in this way were designed on the emotional keypad as shown in Figure 4. The user connects a smart device to the emotion keypad to listen to the work description and select the emotion adjective key for each situational query (Kim, 2022).

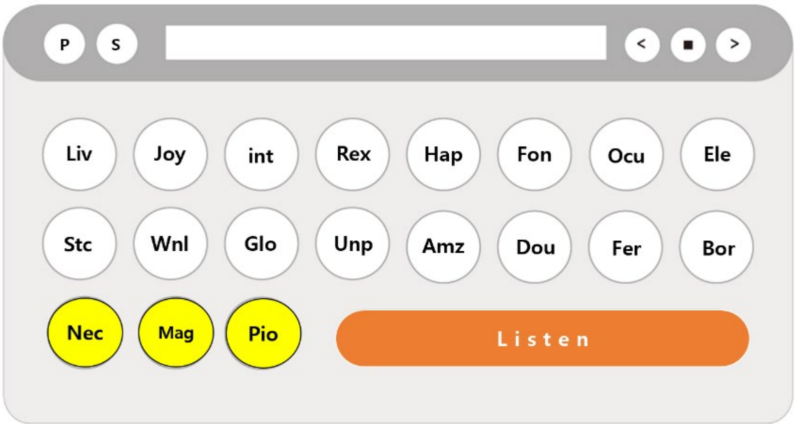


Fig. 4: Emotion-keypad

## 3.4. AI composition algorithm

The AI composition algorithm using user emotions proposed in this study proceeds as shown in Figure 5. First, the AI server learns pre-stored classical MIDI (Musical Instrument Digital Interface) files and generates templates for each type of emotion. MIDI files were used provided by MusicNet.

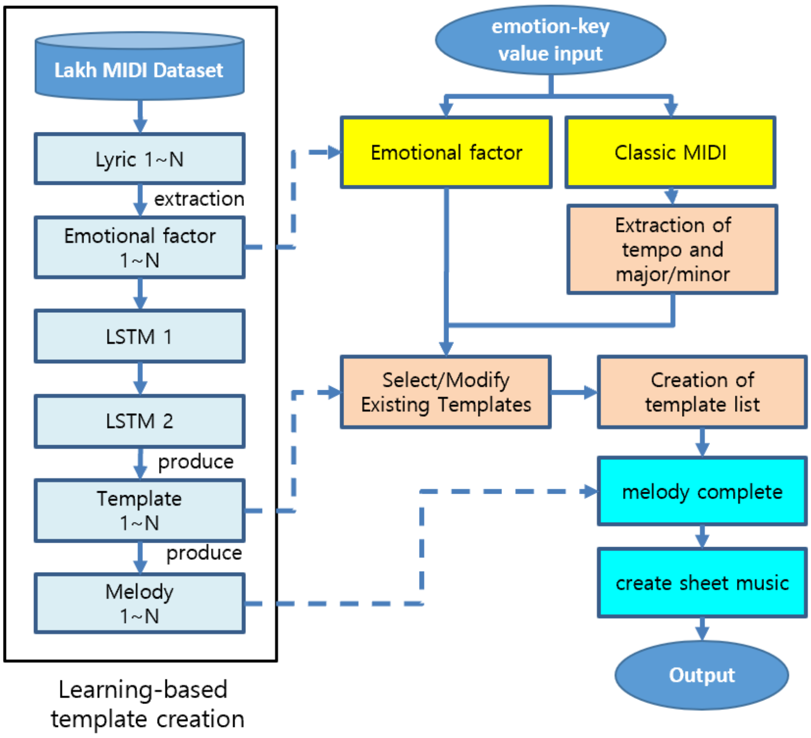


Fig. 5: Flowchart of AI composition algorithm

MusicNet is a collection of 330 freely-licensed classical music recordings, together with over 1 million annotated labels indicating the precise time of each note in every recording, the instrument that plays each note, and the note's position in the metrical structure of the composition. The labels are acquired from musical scores aligned to recordings by dynamic time warping. The labels are verified by trained musicians; we estimate a labeling error rate of 4% (Simões et al., 2022).

For mapping between emotional keywords and music templates, a music emotional classification model using the XGBoost artificial intelligence model was designed as shown in Figure 6.

MFCC features were extracted from the music data collected in step 1 and the music sections were automatically divided. Next, data labeling was performed using music experts on the segmented section data. Finally, a music emotion classification model was created using the XGBoost algorithm.

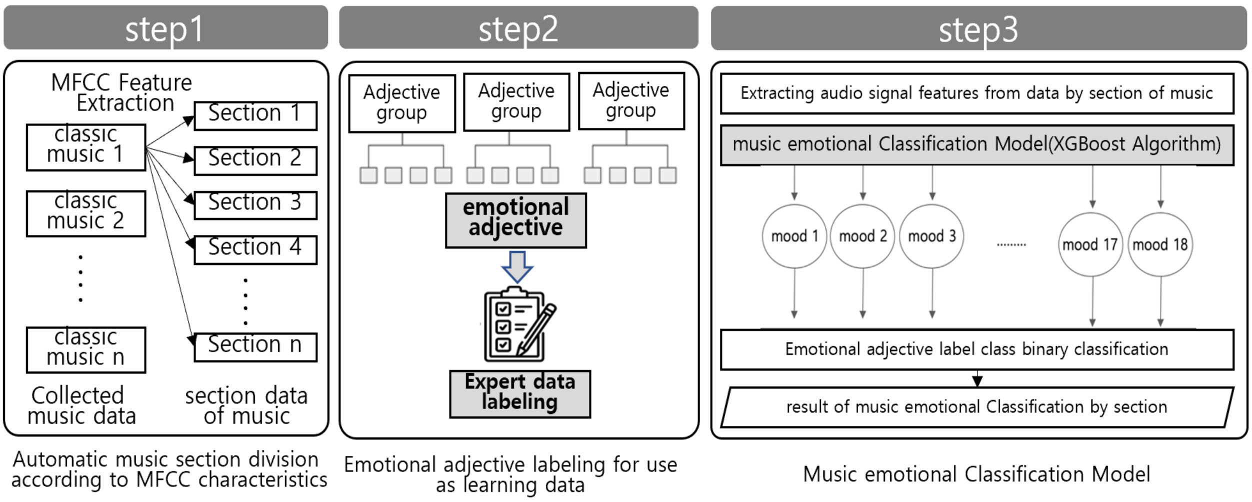

Fig. 6: Music emotional classification model using XGBoost artificial intelligence model

Two LSTM layers were used as the circulating neural network. The user views a work of art and communicates the emotions felt to the AI server through the emotion keypad. The AI server extracts templates according to the emotion key values and generates a musical score by naturally concatenating them while sequentially generating midi notes. In this case, the emotional elements and templates closest to the input emotion key values are selected, and the melody is composed by modifying the selected templates according to classical MIDI information.

Figure 7 shows the process of the part of the AI composition algorithm that corresponds to automatic template generation. This process involves reading the MIDI file of the original song, analyzing the pattern, generating the chord, and then generating a template in JSON file format for use in automatic composition. For this purpose, the researchers looked for the most commonly used Chord in the MIDI file of the original song, with phrases of 4, 8, 12, and 16 verses in length, with the first verse beginning with the first chord of tonality. The phrase patterns with the phrase lengths determined to be most appropriate are then sorted in their original order. At this point, the clause length with the most repetitions of similar patterns will be identified, and a template for each of those clauses will be created.

Figure 8 shows the partial contents of the “template.json” file generated by the automatic template generation algorithm.

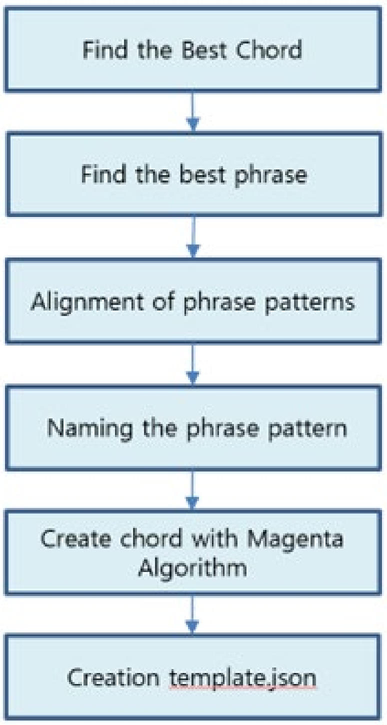


Fig. 7: Block diagram of automatic template generation

```
{
"para structure": ["A", "B", "C", "D", "X"],
"phrase structure": [["pre"], ["a", "b1", "a", "b2"], ["c", "d", "a", "b1"], ["e", "e", "a", "b1"], ["a", "b2"]],
"dev structure": {
"pre": ["motive|dev|repeat|dev", 1],
"a": ["motive|dev|repeat|dev", 1],
"b1": ["motive|dev|repeat|dev", 1],
"b2": ["motive|dev|repeat|dev", 1],
"c": ["motive|dev|repeat|dev", 1],
"d": ["motive|dev|repeat|dev", 1],
"e": ["motive|dev|repeat|dev", 1]},
"chord progress": {
"pre": ["G:m7 G:m7 G:m7 G:m7 G:m7 G:m7 G:m7 G:m7"],
"a": ["G:m7 D:m D:m D:m D:maj7 A: A: A:"],
"b1": ["A: A: F:maj7 C: G:m7 D:m D:m D:m"],
"b2": ["A: A: G:m7 C: D:m D:m D:m D:m"],
"c": ["D:m D:m A:m7 E:dim E:dim A:m7 G:m7 D:m"],
"d": ["D:m D:dim D:dim C: E:m7b5 D:m D:m D:m"],
"e": ["D:m D:m D:m D:m D:m D:m D:m D:m"]}
}
```


Fig. 8: Example content of “template.json” file

Next, name the phrase patterns, giving them the same name if the chord patterns are the same, or naming the patterns by b1, b2, etc. if the chord patterns are slightly different. Then, for each pattern generated, a chord is generated using the magenta algorithm and a template.json file is generated.

### 3.5. Learning of music data

The target music for learning music data for automatic composition had to be converted to text. In this study, the music21 library in Python was used to extract the necessary information from a MIDI (Musical Instrument Digital Interface) file and convert it to text. Figure 9 shows an example of outputting music data in the form of sheet music using the music21 library. The note information used the numeric notation scheme defined in music21, and the negative height information was represented as fixed numbers via the pitch.Pitch.midi object.

```
from music21 import *
test = converter.parse("""tinynotation : 4/4
          c4_c4 d_d e_e f_f g_g a_a b_b r4_R4
          c'4_c'4 d'_d' e'_e' f'_f' g'_g' a'_a' b'_b'
          r4_R4""")
test.show()
```


Fig. 9: Sheet music output using music21 library

## 4. Experiments and Analysis

### 4.1. Experimental Results

Figure 10 illustrates the process invoked by the AI server after the emotion values entered by the user using the emotion keypad are sent to the Web server. The structure of the data tagged with emotion values consists of emotion values for the start and end times of the interval in which the emotion key was pressed.

```
A006 : [['00:00', '1:22', '슬픔'], ['1:23', '2:12', '분노'], ['2:13', '2:56', '기쁨'], ['2:57', '3:57', '분노'], ['3:58', '4:41', '사랑'], ['4:42', '5:03', '놀람'], ['5:04', '5:50', '분노'], ['5:51', '6:19', '슬픔'], ['6:20', '6:41', '기쁨'], ['06:42', '07:37', '공포'], ['07:38', '08:23', '놀람']]

A006
 템포 변화 ------------------------------
 52.000006933334255 BPM at 0 ticks
 120.0 BPM at 69120 ticks
 132.00042240135167 BPM at 69120 ticks
 132.00071280384915 BPM at 851040 ticks
 150.0 BPM at 862800 ticks

 max_tick : 958802

 슬픔 : [0, 82] --> [0.0, 68224.00909653454]
 분노 : [83, 132] --> [69056.0092074679, 172445.89249885597]
 기쁨 : [133, 176] --> [174557.8992572776, 183909.2357999998]
 분노 : [177, 237] --> [267486.19662782917, 330309.2357999998]
 사랑 : [238, 281] --> [396318.6088915484, 435909.2357999998]
 놀람 : [282, 303] --> [489246.90626210003, 488709.2357999998]
 분노 : [304, 350] --> [535711.0549473758, 601509.2357999998]
 슬픔 : [351, 379] --> [634975.3725931923, 671109.2357999998]
 기쁨 : [380, 401] --> [696223.5685874195, 723909.2357999998]
 공포 : [402, 457] --> [742687.7172726953, 858309.2357999998]
 놀람 : [458, 503] --> [860960.1175686348, 968709.2357999998]
```

Fig. 10: AI server loading process of emotion key value

Figure 11 shows the process of reflecting emotion tag data in phrases. The template.json file with the most similar pattern to the user's emotion value tagged data is selected, where the process of reflecting emotion tag data into phrases is completed.

Fig. 11: The process of reflecting emotion tag data to phrases

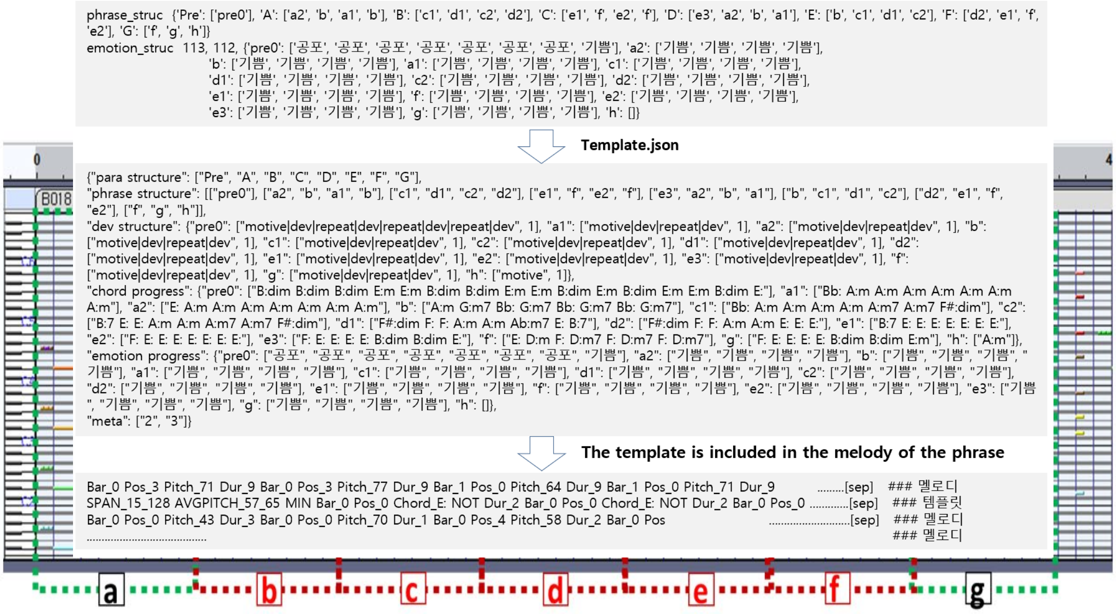


Figure 12 shows the phrase structure of a song generated by the automatic composition function, reflecting the emotion key values entered by the user.

Fig. 12: Composition of songs automatically composed by reflecting emotional key values

## 4.2. Analysis of experimental results

In this study, METEOR and BLEU, commonly used measures of machine learning model performance, were used to evaluate performance against AI-based automatic composition (Evtikhiev et al, 2023; Fomicheva and Specia, 2019).

The METEOR score, ranging from 0 to 1, represents the degree of similarity between the musical composition generated by the AI and the referenced training model composition. Higher METEOR scores indicate greater alignment between the AI-based automatic composition and the referenced training model in terms of musical elements such as melody, harmony, rhythm, and structure.

Models or systems with higher METEOR scores may have performed better in generating musical compositions that were very similar to the reference composition in terms of musical characteristics.

In the case of BLEU, a score between Figure 0 and 1 represents the result, with higher scores indicating greater similarity between the two compositions with respect to musical patterns, motifs and other structural elements.

For this study, the researchers would like to determine that test results METEOR and BLEU scores between 0.2 and 0.6 are appropriate generation results. A low score indicates that the AI-based automatic composition does not match well with the musical elements, structures, or stylistic characteristics in the referenced training model, while a high score means a high degree of similarity with the reference piece. Therefore, a range between 0.2 and 0.6 was specified as the range where differentiation from the reference piece is maintained and some similarity in musical structure is present.

## 4.3. Results

Figure 13 shows the results of the performance evaluation of the AI infrastructure for automatic composition. There were 50 songs generated for testing, and the METEOR and BLEU results are shown for each song. The average value for METEOR is 0.377, with a 96% probability of success in the proper range, and the average value for BLEU is 0.302, with a 84% probability of success in the proper range. This test represented the case of one experiment; these tests were performed a total of 10 times.

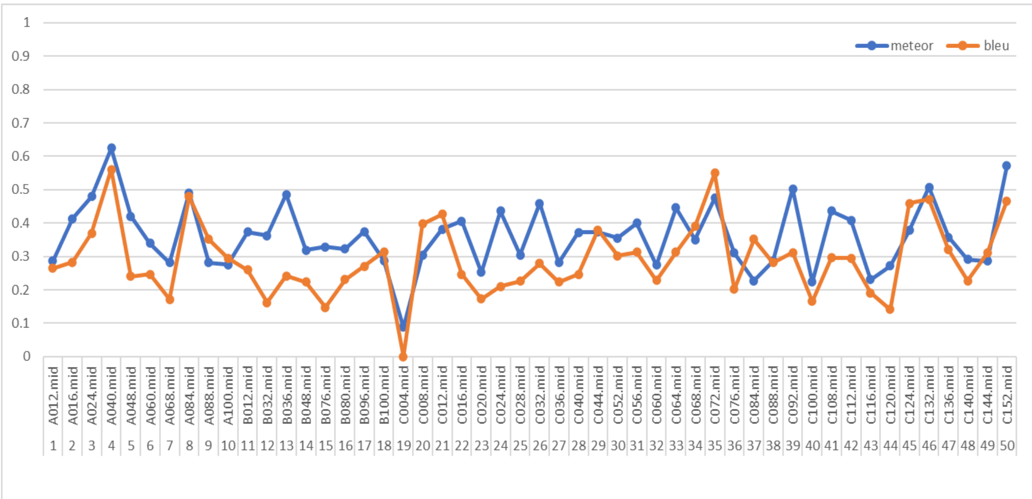


Fig. 13: Performance evaluation of AI-based automated composition

Table 3 shows the results of the 10 runs of the test shown in Figure 9. This test is the result of comparing 50 songs generated by selecting the same 50 learning models referenced in the experimental conditions and specifying different emotion key values each time.

In the case of Round 1, the METEOR and BLEU average values were 0.385 and 0.294, confirming that the appropriate range success probabilities are 92.0% for METEOR and 74.0% for BLEU. The mean of the METEOR and BLEU values per each experiment relative to the proper range success probability is 94.8% and 82.6%, indicating standard deviations of 0.02 and 0.057, respectively.

The average METEOR and BLEU values per 10 tests were 0.385 and 0.305, confirming that the average METEOR and BLEU values relative to the proper range success probability are 94.8% and 82.6%, respectively.

Table 3: Results of 10 tests

| number | meteor | bleu | meter success rate | bleu success rate |
|---|---|---|---|---|
| Round 1 | 0.385 | 0.294 | 92.0% | 74.0% |
| Round 2 | 0.377 | 0.302 | 96.0% | 84.0% |
| Round 3 | 0.372 | 0.307 | 100.0% | 84.0% |
| Round 4 | 0.393 | 0.309 | 94.0% | 88.0% |
| Round 5 | 0.385 | 0.304 | 94.0% | 88.0% |
| Round 6 | 0.393 | 0.303 | 94.0% | 76.0% |
| Round 7 | 0.381 | 0.316 | 94.0% | 90.0% |
| Round 8 | 0.375 | 0.294 | 96.0% | 78.0% |
| Round 9 | 0.384 | 0.296 | 94.0% | 76.0% |
| Round 10 | 0.402 | 0.321 | 94.0% | 88.0% |
| average | 0.385 | 0.305 | 94.8% | 82.6% |
| std | 0.009 | 0.009 | 0.020 | 0.057 |

## 5. Conclusion

This paper focuses on the synthesis of emotion, art, and music by an AI composition system to collect users' emotional responses to art and generate music reflecting such emotions. The proposed AI composition system allows users to express their emotions in real-time using an emotion keypad in an online art lesson based on user emotions. The system converts the expressed emotion into simple music for the user via an automatic music composition function.

The AI server learns from pre-stored classical MIDI files and generates templates for each type of emotion. Two LSTM layers were used as the circulating neural network. The user views a work of art and communicates the emotions felt to the AI server through the emotion keypad. The AI server extracts templates according to the emotion key values and generates a musical score by naturally concatenating them while sequentially generating midi notes. It would select the emotion element and template closest to the input emotion key value, and then compose the melody by modifying the selected template according to the classical MIDI information.

The performance of AI-based automatic composition was evaluated using commonly used metrics such as METEOR and BLEU. Test results showed satisfactory results with high success probabilities within the appropriate range, with a mean of 94.8% and 82.6% for the probability of success and standard deviations of 0.02 and 0.057, respectively.

Future research directions would include the development of segmented emotional templates, personalization, and some adaptive systems. By pursuing this direction, the AI composition system can be further improved to allow improved personalization, creative expression, and emotional experience in the art and music fields.

## References

Briot, J., Hadjeres, G., & Pachet, F. (2020). Deep learning techniques for music generation (1st ed. 2020. ed.). Cham: Springer International Publishing. DOI:http://dx.doi.org/10.1007/978-3-319-70163-9 Retrieved from http://hdl.handle.net/2078/ebook:151952.

Carnovalini, F., & Rodà, A. (2020). Computational creativity and music generation systems: An introduction to the state of the art. *Frontiers in Artificial Intelligence,* 3, 14. DOI:http://dx.doi.org/10.3389/frai.2020.00014.

Chaudhary, D., Singh, N. P., & Singh, S. (2021). Development of music emotion classification system using convolution neural network. *International Journal of Speech Technology*, 24(3), 571-580. DOI:http://dx.doi.org/10.1007/s10772-020-09781-0.

Civit, M., Civit-Masot, J., Cuadrado, F., & Escalona, M. J. (2022). A systematic review of artificial intelligence-based music generation: Scope, applications, and future trends. *Expert Systems with Applications, 209*, 118190. DOI: http://dx.doi.org/10.1016/j.eswa.2022.118190.

Clancy, M. (2023). Artificial intelligence and music ecosystem (1st ed.). London ; New York: Routledge. doi:10.4324/9780429356797 Retrieved from http://bvbr.bib-bvb.de:8991/F?func=service&doc_library=BVB01&local_base=BVB01&doc_number=034147187&sequence=000001&line_number=0001&func_code=DB_RECORDS&service_type=MEDIA.

Cupido, C., Venter, W., & Meyer, J. (2022). Music, art and emotion AOSIS. DOI:http://dx.doi.org/10.4102/aosis.2022.BK333. Retrieved from https://directory.doabooks.org/handle/20.500.12854/94244.

Evtikhiev, M., Bogomolov, E., Sokolov, Y., & Bryksin, T. (2023). Out of the BLEU: How should we assess quality of the code generation models? *The Journal of Systems and Software*, 203, 111741. DOI:http://dx.doi.org/10.1016/j.jss.2023.111741.

Fomicheva, M., & Specia, L. (2019). Taking MT evaluation metrics to extremes: Beyond correlation with human judgments. *Computational Linguistics - Association for Computational Linguistics, 45*(3), 515-558. DOI:http://dx.doi.org/10.1162/coli_a_00356.

Ha, S. & Lee, S. (2023). Feature extraction and analysis of AI-based speech signals for auditory rehabilitation. *Asia-pacific Journal of Convergent Research Interchange*, KCTRS, ISSN: 2508-9080 (Print); 2671-5325 (Online), 9(5), 105-116. DOI:http://dx.doi.org/10.47116/apjcri.2023.05.09.

Harini, B. & Rao, N. T. (2019). An extensive review on recent emerging applications of artificial intelligence. *Asia-pacific Journal of Convergent Research Interchange, SoCoRI*, ISSN: 2508-9080 (Print); 2671-5325 (Online), 5(2), 79-88. DOI:http://dx.doi.org/10.21742/apjcri.2019.06.09.

Han, X., Chen, F., & Ban, J. (2023). Music emotion recognition based on a neural network with an inception-GRU residual structure. *Electronics (Basel)*, 12(4), 978. DOI:http://dx.doi.org/10.3390/electronics12040978.

Horvat, M., Jović, A., & Burnik, K. (2022). Investigation of relationships between discrete and dimensional emotion models in affective picture databases using unsupervised machine learning. *Applied Sciences, 12*(15), 7864. DOI: http://dx.doi.org/10.3390/app12157864.

Hyeong-Gyun Kim. (2022). Study on non-face-to-face art appreciation system using emotion key. *Journal of Digital Convergence*, *20*(2), pp.57-62. DOI:http://dx.doi.org/10.14400/JDC.2022.20.2.057.

Kim, B. –H. (2023). a study on the perception of artificial intelligence education and the artificial intelligence teaching efficacy in secondary school technology teacher. *Asia-pacific Journal of Convergent Research Interchange, KCTRS*, ISSN: 2508-9080 (Print); 2671-5325 (Online), 9(4), 517-527. DOI:http://dx.doi.org/10.47116/apjcri.2023.04.41.

Juslin, P. N. (2011). Handbook of music and emotion. Oxford: Oxford University Press. DOI:http://dx.doi.org/10.1093/acprof:oso/9780199230143.001.0001 Retrieved fromhttp://digital.casalini.it/9780191621970.

Micchi, G., Gotham, M., & Giraud, M. (2020). Not all roads lead to rome: Pitch representation and model architecture for automatic harmonic analysis. *Transactions of the International Society for Music Information Retrieval*, *3*(1), 42-54. DOI: http://dx.doi.org/10.5334/tismir.45.

Rohrmeier, M. (2022). On creativity, music's AI completeness, and four challenges for artificial musical creativity. *Transactions of the International Society for Music Information Retrieval*, 5(1), 50-66. DOI:http://dx.doi.org/10.5334/tismir.104.

Roig-Francolí, M. A. (1995). Harmonic and formal processes in ligeti's net-structure compositions. *Music Theory Spectrum*, 17(2), 242-267. DOI: http://dx.doi.org/10.1525/mts.1995.17.2.02a00050.

Sehar, U., & Naseem, M. L. (2022). How deep learning is empowering semantic segmentation. *Multimedia Tools and Applications, 81*(21), 30519-30544. DOI:http://dx.doi.org/10.1007/s11042-022-12821-3.

Simões, L. V., Simões, A. R. M., Komati, K. S., & Andrade, J. O. (2022). Automatic musical transcription applying fine-tuning by composer in neural networks for the MusicNet database. *International Journal of Engineering Trends and Technology, 70*(12), 328-337. DOI:http://dx.doi.org/10.14445/22315381/IJETT-V70I12P231.

Stoller, D. (2020). Deep learning for music information retrieval in limited data scenarios. Retrieved from https://qmro.qmul.ac.uk/xmlui/handle/123456789/68513.

Van den Oord, A., Dieleman, S., Zen, H., Simonyan, K., Vinyals, O., Graves, A., Kavukcuoglu, K. (2016). WaveNet: A generative model for raw audio. (). Ithaca: Cornell University Library, arXiv.org. DOI: http://dx.doi.org/10.48550/arXiv.1609.03499 Retrieved from Publicly Available Content Database Retrieved from https://search.proquest.com/docview/2080352104.

Yu, H., Taube, H., Evans, J., & Varshney, L. (2020). Human evaluation of interpretability: The case of AI-generated music knowledge. Ithaca: Cornell University Library, arXiv.org. DOI: http://dx.doi.org/10.48550/arxiv.2004.06894 Retrieved from Publicly Available Content Database Retrieved from https://search.proquest.com/docview/2390350870.