感情に応じて動的に音楽を生成するゲームシステムの設計

Design of a Game System that Dynamically Generates Music According to Player Emotions

毛 碩[1]　角 薫[1]

MAO Shuo [1], SUMI Kaoru [1]

[1] 公立はこだて未来大学大学院 システム情報科学研究科

[1] Graduate School of Systems Information Science,
Future University Hakodate

Abstract: This study proposes and implements a game system that generates background music based on estimated player emotion categories. While recent advances in AI-based music generation enable flexible music creation, limited research has explored how emotion-related information can be incorporated into interactive game systems and how such generated music may influence player experience. The proposed system integrates a facial expression analysis module and an AI-based music generation module within a game environment developed in Unreal Engine. During gameplay, the system estimates the player's emotional category (happy, sad, or angry) and generates corresponding music. The generated music is then presented in a subsequent scene experience and compared with fixed background music. A within-subject user study was conducted with university students across three emotional scenes. Questionnaire results and qualitative responses indicate that participants generally perceived the generated music as matching the scene and influencing their emotional experience. Changes in background music were also reported to affect immersion and scene interpretation. These findings suggest that incorporating emotion-category-based music generation into game systems may influence player experience and provide a foundation for further research on adaptive interactive music design

1. INTRODUCTION

Music plays an important role in shaping the game experience. Appropriate background music can enhance emotional expression, improve player immersion, and guide players' perception of the game environment. However, most existing games still rely on pre-composed, fixed music tracks. These tracks cannot respond to changes in the player's emotional state and therefore may not always remain consistent with the player's actual experience.

Music in interactive media has been shown to influence not only emotional response but also cognitive appraisal and narrative interpretation. Rather than functioning solely as background accompaniment, music may shape how users interpret and emotionally respond to visual content. From this perspective, music can be considered a potential affective modulation mechanism within interactive environments.

With the advancement of AI-based music generation technologies such as MusicGen, Stable Audio, and Suno, it has become possible to automatically generate music based on specific conditions. These technologies provide new opportunities for developing emotion-category-based music systems. However, research that integrates player emotion estimation with music generation in game environments and systematically examines its influence on player experience remains limited.

To explore a more personalized and emotion-category-based game experience, this study proposes a game system that generates music based on estimated player emotion categories. The system detects the player's emotional state during gameplay and combines this information with the player's music preferences to generate music aligned with the estimated emotional category, which is then played during subsequent gameplay.

The objective of this study is to investigate whether music generated based on estimated emotion categories is perceived as matching the emotional tone of game scenes

and how such music may influence players' subjective experience, including immersion and scene interpretation. To examine this, a user study was conducted comparing generated music with fixed background music conditions across multiple emotional scenes.

2. LITERATURE REVIEW

This study involves multiple research areas, including emotion recognition, AI-based music generation, and adaptive music in games. This chapter reviews relevant studies in these areas and analyzes their limitations as well as the contributions of the present study.

2.1 Emotion Recognition in Interactive Systems

Emotion recognition technology serves as a fundamental component for emotion-driven interactive systems. In recent years, with advances in computer vision and machine learning, facial expression–based emotion recognition methods have been widely applied in human–computer interaction systems.

DeepFace is a deep learning–based facial analysis framework capable of recognizing various emotional categories, including happiness, sadness, and anger. Previous research has demonstrated that DeepFace achieves high accuracy in real-time emotion recognition tasks and can be applied to interactive systems for automatic emotion detection (Live Emotion Detection Using DeepFace) [8].

In the gaming domain, researchers have begun exploring the use of emotion recognition to enhance player experience. For example, Enhancing Game Experience with Facial Expression Recognition as Dynamic Balancing proposed a dynamic difficulty adjustment system based on facial expression recognition [1]. This system detects the player's facial expressions in real time and adjusts game difficulty according to the player's emotional state. Experimental results showed that the system significantly improved player experience.

Additionally, the GAMEEMO study constructed an emotion recognition dataset based on electroencephalogram (EEG) signals collected during gameplay, providing an important foundation for emotion recognition research in games [14].

Furthermore, Violent video game play impacts facial emotion recognition demonstrated that gameplay itself can influence players' emotion recognition abilities [15], further highlighting the close relationship between emotion and game experience.

These studies indicate that emotion recognition technology can effectively detect player emotional states and can be used to improve gameplay experience. However, most existing research primarily uses emotion recognition to adjust game difficulty or gameplay content, while relatively few studies focus on dynamically generating music based on the player's emotional state.

2.2 AI-based Music Generation

In recent years, significant progress has been made in AI-based music generation due to advances in deep learning technologies.

MusicGen proposed a Transformer-based music generation model capable of generating high-quality music from text descriptions while providing strong controllability (Simple and Controllable Music Generation) [7].

MusicLM introduced a hierarchical music generation approach that can generate structurally coherent music from textual input while maintaining high audio quality (MusicLM: Generating Music From Text) [6].

Stable Audio is based on diffusion models and can generate long-duration audio from text prompts with high efficiency and audio quality (Fast Timing-Conditioned Latent Audio Diffusion) [5].

These studies demonstrate that AI models can generate music that matches specific emotional and stylistic conditions based on textual descriptions, providing a technical foundation for dynamic music generation.

However, most of these studies focus primarily on improving music generation models themselves, while relatively few explore how generated music can be applied in game environments and dynamically adapted based on player emotional states.

2.3 Adaptive Music in Games

Adaptive music refers to music that dynamically changes according to game conditions to enhance the gameplay experience.

Adaptive Music Composition for Games proposed an adaptive music system capable of dynamically generating music based on game states, thereby improving player immersion [9].

The study Music Matters showed that adaptive music enhances emotional experience and increases player immersion compared to fixed music [10].

An adaptive music generation architecture for games based on the deep learning Transformer model proposed a Transformer-based music generation system capable of generating music based on game states [4].
Progressive-Adaptive Music Generator introduced a procedural music generation method that dynamically adjusts musical structure according to game states [13].
These studies demonstrate that adaptive music can improve player experience. However, most existing research focuses on adapting music based on game states rather than dynamically generating music based on player emotional states.

2.4 Relationship Between Music and Emotion

There is a strong relationship between music and emotion. The study Music and Emotion showed that music can induce emotional changes in listeners and influence both psychological and physiological states [11].
Music: A Link Between Cognition and Emotion demonstrated that musical emotion is closely related to cognitive processes [12].
The study Exposure to music and cognitive performance showed that different types of music can influence cognitive performance by affecting emotional states [3].
Additionally, Effects of individualized music on agitation demonstrated that music selected based on individual preferences can significantly improve emotional states [2].
These studies indicate that music can influence emotional experience, and personalized music can further enhance this effect.

2.5 Summary and Research Gap

Previous studies have demonstrated that:
Emotion recognition technologies can detect player emotional states;
AI-based music generation technologies can generate high-quality music;
Adaptive music can improve player experience;
Music can influence player emotional states.
However, several limitations remain in existing research:
Most adaptive music systems adjust music based on game states rather than player emotional states;
AI music generation research primarily focuses on generation techniques rather than their application in game environments;
Few studies integrate emotion recognition, AI music generation, and game systems to dynamically generate music based on player emotional states.
To address these limitations, this study proposes a game system that dynamically generates music based on the player's emotional state. The system integrates emotion recognition and AI-based music generation technologies to generate music tailored to the player's emotional state and applies it within a game environment.
This study aims to evaluate the impact of this system on player experience.

2.6 Position of This Study

Although previous studies have made significant progress in adaptive music, emotion recognition, and AI-based music generation, several limitations still remain.
First, in the field of adaptive music for games, previous research has proposed systems that dynamically adjust music based on game states or player behavior. These studies have shown that adaptive music can improve player immersion and emotional experience. However, most existing systems rely on switching between pre-composed music tracks rather than generating new music based on detected or estimated player emotion categories..
As a result, their ability to provide personalized and emotionally adaptive music remains limited.
Second, in emotion recognition research, various approaches have been developed to detect player emotions using facial expressions, physiological signals, and behavioral data. These emotion recognition methods have been applied to game difficulty adjustment and interface adaptation. However, few studies have directly utilized emotion recognition results to generate music dynamically.
Furthermore, recent advances in AI music generation, such as MusicGen, MusicLM, and Stable Audio, have enabled the generation of high-quality music from textual descriptions. These technologies provide a strong foundation for dynamic music generation. However, most existing research focuses on improving music generation models themselves, while relatively few studies have explored integrating these models into interactive game systems to generate music based on estimated player emotion categories.
Based on these observations, this study makes the following contributions:
(1) Emotion-category-based music generation framework incorporating facial expression estimation
This study uses DeepFace to detect players' emotional states in real time and dynamically generates music based on the detected emotions, enabling personalized emotional

adaptation.

(2) Integration of AI music generation into a game system

This study integrates AI music generation models, such as Stable Audio, into a game system to dynamically generate and play music based on the player's emotional state.

(3) Experimental validation of the impact of generated music on player experience

This study conducts a user experiment comparing fixed music and dynamically generated music, evaluating their effects on immersion, emotional experience, and music-scene consistency.

By integrating emotion recognition, AI music generation, and game systems, this study proposes an emotion-driven music generation framework and experimentally validates its effectiveness.

The results contribute to the development of emotionally adaptive game design and interactive music systems.

3. Experimental System

This study designed and implemented a game system that dynamically generates music based on the player's emotional state. During gameplay, the system detects the player's current emotional state through emotion recognition and generates music with corresponding emotional characteristics. This enables dynamic music adaptation based on the player's emotional state.

The system consists of the following four core modules:

- Game Environment Module
- Emotion Recognition Module
- Music Generation Module
- Music Playback Module

The system adopts a distributed architecture in which Unreal Engine and Python run independently. The Game Environment Module and Music Playback Module operate within Unreal Engine, while the Emotion Recognition Module and Music Generation Module run in a Python environment. Data transmission between modules is achieved through socket communication and audio file exchange.

During system operation, the camera captures the player's facial images, which are analyzed by the Emotion Recognition Module to determine the player's emotional state. The recognition results are transmitted to the Music Generation Module, which generates music files corresponding to the detected emotional state. The generated music is played when the player re-enters the scene, enabling music adaptation based on the player's emotional state.

This system establishes a complete emotion-driven interaction loop, allowing music to dynamically adapt to the player's emotional condition.

3.2 Game Environment Module

3.2.1 Unreal Engine Environment

The game environment was developed using Unreal Engine 5.2. Unreal Engine is a widely used real-time 3D engine that provides high-quality graphics rendering and audio processing capabilities.

In this system, Unreal Engine is responsible for:

- Constructing game scenes
- Playing background music
- Controlling scene transitions
- Communicating with Python modules

Audio playback is implemented using the AudioComponent. AudioComponent supports loading, playing, and switching audio files, enabling the system to play both pre-composed music and dynamically generated music.

The game runs on a desktop computer. The player experiences the game through a monitor, while a camera captures the player's facial images for emotion recognition.

3.2.2 Scene Design

Three experimental scenes with different emotional characteristics were designed:

These scenes were used to induce different emotional states in players and evaluate the effectiveness of dynamically generated music.

The Happy scene depicts family members dining together in a warm environment. The scene features bright lighting and a warm atmosphere intended to induce positive emotions. The default background music is warm and uplifting (Fig 1).

The Sad scene depicts a family argument. The environment is designed with a depressed atmosphere intended to induce sadness. The default background music reflects a sad emotional tone (Fig 2).

The Angry scene depicts two children arguing. The tense atmosphere is intended to induce anger or tension. The default background music reflects a tense emotional style.

Through these scenes, the system detects player emotional states under different emotional conditions and generates corresponding music (Fig 3).

Figure 1. Happy scene

Figure 2. Sad scene

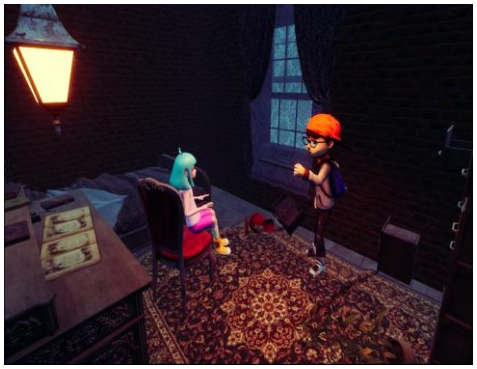

Figure 3. Angry scene

3.2.3 Scene Validation

To ensure that the designed scenes accurately conveyed the intended emotional categories, two pilot studies were conducted prior to the main experiment.

In the first pilot study, 10 university students participated in a questionnaire-based evaluation. Participants were shown screenshots and short animation clips of each scene and were asked to select the emotion that best matched their first impression. Based on the results and participant feedback, several revisions were made to improve emotional clarity. These revisions included adjustments to

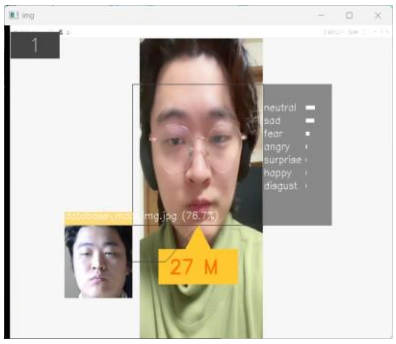

Figure 4. Emotion recognition module during operation.

scene lighting, character animations, and background music in order to strengthen the intended emotional expression of each scene.

Following these revisions, a second pilot study was conducted with 12 university students to validate the effectiveness of the modified scenes. Participants watched video recordings of each scene and selected the perceived emotional category. The results showed that 10 out of 12 participants correctly identified the happy scene, 12 out of 12 participants correctly identified the sad scene, and 9 out of 12 participants correctly identified the angry scene.

These results indicate that the majority of participants perceived the scenes as intended emotional categories. Therefore, the scenes were considered sufficiently valid for use in the main experiment.

3.3 Emotion Recognition Module

3.3.1 DeepFace Emotion Recognition

The system uses DeepFace for facial expression analysis. DeepFace is a convolutional neural network–based facial analysis framework capable of recognizing multiple emotion categories, including happy, sad, angry, neutral, fear, surprise, and disgust (Fig 4).

In this study, the system focuses on three emotional categories: happy, sad, and angry, which correspond to the emotional design of the experimental scenes.

For each detection, DeepFace outputs probability values for all emotion categories. The system uses the dominant emotion category (i.e., the category with the highest probability) as the primary input for music generation. This detected emotion category is used together with participant-specific music preference information collected in the pre-experiment questionnaire to generate music prompts for the music generation module.

In addition, after each scene experience, participants were asked to report their perceived emotional state. These self-reported emotions were used as reference information for interpreting the results and understanding the relationship between detected emotion categories and subjective experience, but were not directly used as inputs for music generation.

3.3.2 Emotion Detection Process

The Emotion Recognition Module runs in the Python environment and captures the player's facial images in real time using a camera.

The detection process consists of the following steps:

- Capturing camera images
- Detecting facial regions
- Analyzing facial expressions
- Outputting emotion classification results and probability values

Emotion detection is performed at fixed time intervals, enabling continuous monitoring of the player's emotional state.

3.3.3 Emotion Data Communication

Socket communication is used to enable data exchange between the Python modules and Unreal Engine.

The Emotion Recognition Module operates in the Python environment as a socket client and sends detected emotional data to Unreal Engine.

After receiving the emotion data, Unreal Engine uses the detected dominant emotion category as the primary input for generating music prompts and transmitting them to the Music Generation Module. The generated music is combined with participant-specific preference parameters obtained from the pre-experiment questionnaire.

Self-reported emotional responses collected after scene experiences were not used as inputs for music generation but were used as reference information for experimental analysis and interpretation of participant experience. The Music Generation Module also runs in the Python environment and generates music files based on the received emotional data.

This communication mechanism allows Unreal Engine to obtain the player's emotional state in real time while enabling independent execution of the Music Generation Module, thereby supporting a modular system architecture.

3.4 Music Generation Module

3.4.1 AI Music Generation Models

The system uses the following AI music generation models:

- Stable Audio
- Suno
- MusicGen

Stable Audio serves as the primary model due to its ability to run locally and generate music within approximately 20 seconds. Suno and MusicGen are used to validate the feasibility of different music generation approaches.

These models generate music with specific emotional characteristics based on input prompts.

3.4.2 Prompt Generation Based on Emotion and User Preference

Music prompts are generated based on two types of information:

- Player emotional state
- Player music preference questionnaire results

Before the experiment, players complete a music preference questionnaire, which includes information about preferred music genre, instruments, tempo, and mood.

For example, if a player's preferences are:

- Genre: Classical
- Instruments: Strings
- Tempo: Slow
- Mood: Warm

The system generates corresponding prompts for different emotional states.

For sad emotional state (goal: emotional relief):

Format: Orchestra | Genre: Classical | Instruments: Warm Strings, Solo Cello, Soft Violin | Moods: Comforting, Gentle, Hopeful, Sad | Tempo: Slow | BPM: 70

For angry emotional state (goal: tension reduction):

Format: Orchestra | Genre: Classical | Instruments: Soothing Strings, Viola, Cello, Violin | Moods: Calm, Peaceful, Warm | Tempo: Slow | BPM: 65

For happy emotional state (goal: maintaining positive emotion):

Format: Orchestra | Genre: Classical | Instruments: Bright Strings, Violin, Cello, Viola | Moods: Joyful, Warm | Tempo: Slow | BPM: 75

These prompts are provided to the AI music generation models to generate music that matches the player's emotional state and preferences.

This approach enables personalized music generation.

3.4.3 Music Generation Process

The music generation process consists of the following steps:

1. The Emotion Recognition Module detects the player's emotional state
2. The system generates music prompts
3. The AI model generates music
4. The generated music is saved as an audio file
5. The generated music is played when the player re-enters the scene

3.5 Music Playback and System Integration

The system uses Unreal Engine's AudioComponent to play background music. AudioComponent supports loading and playing audio files, allowing both fixed music and dynamically generated music to be played.
In the experiment, each scene is experienced twice:

- First experience: pre-composed fixed music is played
- Second experience: dynamically generated music based on the player's emotional state is played

The generated music is created and saved in the Python environment as audio files, which are then loaded and played by Unreal Engine.
The system receives emotion recognition results through socket communication and controls music generation and playback according to the experimental workflow.
This mechanism enables emotion-driven adaptive music playback.

3.6 Implementation Environment

The proposed system was implemented using Unreal Engine 5.2 on a desktop computer running Windows. Facial expression analysis was performed using the DeepFace library in Python, which estimates emotion probability distributions and dominant emotion categories from webcam input.
Music generation was performed locally using an AI-based music generation model (Stable Audio). Music generation requests were sent from the Unreal Engine game client to a local Python server via socket communication. The Python server generated approximately 20-second music clips based on the estimated emotion category and returned the generated audio file to the game system.
The generated audio files were saved locally and played in the game environment using Unreal Engine's audio playback system. This architecture enabled integration of emotion estimation, music generation, and playback within the interactive game environment.

3.7 System Summary

The proposed system integrates emotion estimation, AI-based music generation, and interactive gameplay within a unified framework. The system estimates the player's emotion category using facial expression analysis and generates corresponding music using an AI-based music generation model. The generated music is then played during gameplay to provide dynamically generated background audio.
The modular architecture allows independent processing of emotion estimation, music generation, and playback while enabling communication between components through socket-based communication. This design enables investigation of how music generated based on estimated emotion categories is perceived within interactive game environments.

4. Experiment

The purpose of this experiment is to evaluate the impact of a dynamically generated music system based on the player's emotional state on player experience. Specifically, this experiment compares fixed background music and dynamically generated music based on the player's emotional state in terms of player immersion, emotional experience, and the perceived congruence between music and game scenes.
Traditional games typically use pre-composed fixed background music, which cannot dynamically adapt to individual players' emotional states. The system proposed in this study generates music based on detected player emotion categories, allowing the music to reflect the player's emotional condition and potentially influence the gameplay experience.
The objective of this experiment is to explore how music generated based on estimated emotion categories is perceived by players within game scenes and to compare subjective responses between generated music and fixed background music conditions.
Specifically, this experiment examines perceived music–scene matching, immersion, emotional experience, and scene interpretation through subjective questionnaire responses and qualitative feedback.By analyzing questionnaire data collected from participants, this experiment evaluates the effectiveness of the dynamic music generation system.

4.2 Participants

A total of 10 participants took part in this experiment. All participants were university students and participated voluntarily.
Participants were between 20 and 30 years old and had normal or corrected-to-normal vision. No restrictions were placed on participants' musical background or gaming experience.
Before the experiment, participants were informed about the experimental procedure and objectives. All participants completed the full experimental protocol.

4.3 Experimental Conditions

The experiment included two music conditions:

- Fixed Music Condition
- Generated Music Condition

In the Fixed Music Condition, the system played pre-composed background music that did not change based on the player's emotional state.
In the Generated Music Condition, the system generated music based on the player's emotional state and music preferences and played the generated music in the scene.
Three experimental scenes with different emotional characteristics were used:

- Happy scene
- Sad scene
- Angry scene

Each scene was experienced under both music conditions.
The independent variable of the experiment was music type (fixed music or generated music).
The dependent variables were questionnaire ratings, including:

- Immersion
- Music–scene congruence
- Emotional impact
- Scene understanding

4.4 Experimental Design

This experiment used a within-subject design, in which each participant experienced all experimental conditions.
Each participant experienced three experimental scenes (Happy, Sad, Angry), and each scene was experienced twice:

- First experience: Fixed Music Condition
- Second experience: Generated Music Condition

Thus, each participant completed a total of six scene experiences.
To reduce order effects, three different scene presentation orders were used:

- ABC
- BCA
- CAB

Participants were randomly assigned to one of these order groups.
This design enabled comparison of the effects of different music conditions within the same participant.

4.5 Experimental Procedure

The experimental procedure was as follows:
First, participants completed a music preference questionnaire, including preferences for music genre, instruments, tempo, and emotional atmosphere.
Next, participants entered the experimental scenes.
In each scene, participants experienced the scene twice.
During the first experience, the system played fixed background music. At the same time, the Emotion Recognition Module detected the participant's facial expressions using a camera and determined the participant's emotional state. Based on the detected emotional state and the participant's music preferences, the system generated music.
After music generation was completed, participants entered the same scene again. During the second experience, the system played the generated music.
After completing both experiences, participants filled out a questionnaire to evaluate their experience.
This procedure was repeated for all three scenes.

4.6 Questionnaire Design

Two types of questionnaires were used in this experiment:

- Music Preference Questionnaire
- Post-experiment Experience Questionnaire

4.6.1 Music Preference Questionnaire

The Music Preference Questionnaire was completed before the experiment to obtain participants' music preference information.

4.6.2 Post-experiment Questionnaire

After each scene experience, participants completed an experience questionnaire.
The questionnaire used a 7-point Likert scale (1–7) to measure participants' experiences.

Table 1 pre-experiment questionnaire items

Item ID	Question	Response Options	Purpose
P1	Which music style do you prefer the most?	Ambient, Lofi, Classical, Electronic, Jazz, Rock, Pop	Music generation parameter reference
P2	Which instrument do you prefer the most?	Piano, Strings, Synth/Pad, Guitar, Bass, Drums	Music generation parameter reference
P3	Which music tempo do you prefer?	Slow, Medium, Fast	Music generation parameter reference
P4	In games overall, what kind of musical atmosphere do you expect?	Calm, Warm, Tense, Sad	Music generation parameter reference

Table 2 Post-Experiment Likert-Scale Questionnaire Items

Item ID	Question	Scale	Construct
Q1	I could clearly understand what was happening in this scene.	1–7 Likert scale	Scene Clarity
Q2	The emotional tone of this scene was clear to me.	1–7 Likert scale	Emotional Clarity
Q3	I felt immersed while staying in this scene.	1–7 Likert scale	Immersion
Q4	The second background music matched the scene well.	1–7 Likert scale	Music–Scene Matching
Q5	The changes in the second music affected my emotional experience.	1–7 Likert scale	Emotional Impact
Q6	During the second experience, the scene felt different from the previous time.	1–7 Likert scale	Scene Perception Change

Table 3 Open-Ended Questionnaire Items

Item ID	Question	Purpose
Q7	In your own words, please briefly describe what you felt in this scene.	Qualitative emotional response
Q8	Please describe how the changes in the second background music affected your feelings during this scene.	Qualitative evaluation of music influence

Table 1 shows the items of the pre-experiment music preference questionnaire. Table 2 presents the post-experiment Likert-scale questionnaire items used to measure scene clarity, emotional clarity, immersion, music–scene matching, emotional impact, and scene perception change. Table 3 lists the open-ended questionnaire items, which allowed participants to describe their experiences and the perceived influence of music in their own words. These questionnaire results were used to evaluate the effectiveness of the dynamic music generation system.

5. Results

This chapter evaluates the impact of dynamically generated music on player experience by combining quantitative questionnaire analysis and qualitative analysis of open-ended responses. In addition, this chapter validates the practical performance of the emotion recognition module during system operation.

5.1 Scene Understanding

Fig 5 presents participants' ratings of scene understanding for each emotional scene.

Overall, participants reported high levels of scene understanding across all three scenes. The mean ratings were 5.83 (SD = 1.11) for Scene A (Happy), 5.92 (SD = 1.16) for Scene B (Sad), and 5.33 (SD = 1.50) for Scene C (Angry). The median ratings were 6 for all three scenes. Most ratings were distributed between 5 and 7, indicating that participants were able to clearly understand the intended content and emotional context of each scene.

Scene B and Scene C showed slightly greater variability, with a small number of lower ratings. However, the majority of ratings remained in the higher range. Scene A showed relatively consistent ratings with fewer extreme

values.
These results indicate that the scene designs successfully conveyed their intended content and emotional context to participants.

5.2Emotional Clarity

Fig 6 shows participants' ratings of emotional clarity for each scene.
Participants generally reported high emotional clarity. The mean ratings were 5.83 (SD = 1.03) for Scene A, 5.75 (SD = 1.29) for Scene B, and 4.50 (SD = 1.73) for Scene C. The median ratings were 6 for Scene A and Scene B, and 4.5 for Scene C.
Scene A and Scene B showed consistently high ratings, indicating that participants were able to clearly perceive the intended emotional tone. Scene C showed greater variability and slightly lower ratings, suggesting that the emotional tone of the angry scene was less consistently interpreted.
Nevertheless, the overall results indicate that the emotional intent of the scenes was generally clear to participants.

5.3 Immersion

Fig 7 presents participants' ratings of immersion for each scene. Participants reported moderate to high immersion levels. The mean ratings were 5.42 (SD = 1.24) for Scene A, 5.08 (SD = 1.56) for Scene B, and 4.58 (SD = 1.78) for Scene C. The median ratings were 5.5 for Scene A, 6 for Scene B, and 5 for Scene C.
Scene A and Scene B showed relatively higher immersion ratings, while Scene C showed greater variability and slightly lower ratings overall. Some participants reported lower immersion in Scene C, indicating that immersion varied depending on individual perception.
Overall, the results suggest that participants reported moderate to high levels of immersion during the scene experiences, particularly in the happy and sad scenes.

5.4 Music–Scene Matching

Fig 8 shows participants' ratings of how well the second background music matched each scene. The mean ratings were 5.17 (SD = 1.47) for Scene A, 4.92 (SD = 1.73) for Scene B, and 4.83 (SD = 1.70) for Scene C. The median ratings were 5.5, 5.5, and 5, respectively.
Participants generally rated the second music as matching the scenes moderately well. Scene A showed slightly higher ratings and more consistent responses, while Scene B and Scene C showed greater variability.
These results suggest that the generated music was generally appropriate for the emotional context of the scenes, although individual responses varied.

5.5 Emotional Impact

Fig 9 presents participants' ratings of the emotional impact of the second background music. The mean ratings were 5.42 (SD = 1.24) for Scene A, 5.42 (SD = 1.16) for Scene B, and 4.75 (SD = 1.66) for Scene C. The median ratings were 5.5, 6, and 5, respectively.
These results indicate that participants reported that the second music influenced their emotional experience, particularly in Scene A and Scene B. Scene C showed greater variability and slightly lower ratings, suggesting that emotional influence varied depending on the participant.
Overall, the results suggest that the adaptive music system was able to affect participants' emotional experience.

5.6 Scene Perception Change

Fig 10 shows participants' ratings of whether their perception of the scene changed during the second experience. The mean ratings were 5.33 (SD = 1.11) for Scene A, 6.00 (SD = 1.22) for Scene B, and 5.92 (SD = 1.24) for Scene C. The median ratings were 5, 6, and 6, respectively.
Participants reported noticeable perception changes, particularly in Scene B and Scene C. These higher ratings suggest that participants reported that the change in background music influenced how they interpreted the scene.
These results indicate that adaptive music has the potential to alter participants' perception and interpretation of emotional scenes.

5.7 Qualitative Analysis

To better understand how dynamically generated music influenced player experience, we conducted a qualitative analysis of participants' open-ended responses. By organizing and summarizing all responses, several consistent experience patterns were identified.
Overall, the impact of dynamically generated music was reflected in the following aspects.

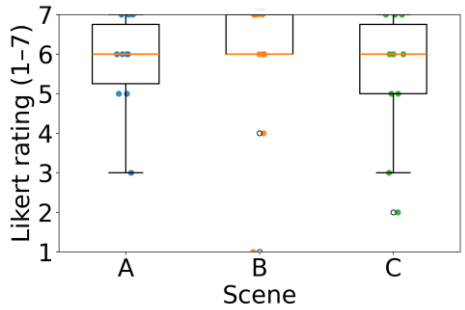

Figure 5. Scene Understanding Ratings for Each Scene (Q1).
Boxplots show Likert-scale ratings (1–7) of participants' understanding of the scene content for Happy (Scene A), Sad (Scene B), and Angry (Scene C) scenes. Individual participant ratings are overlaid.

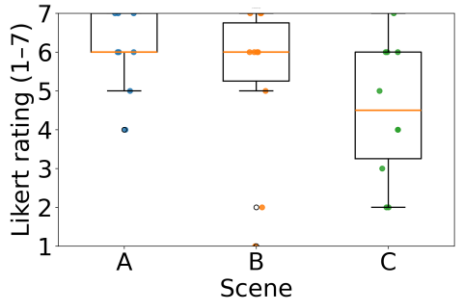

Figure 6. Emotional Clarity Ratings for Each Scene (Q2).
Boxplots show ratings of how clearly participants could perceive the intended emotion of each scene.

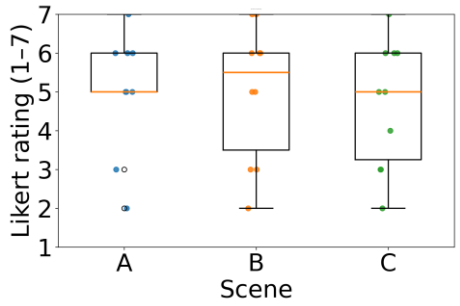

Figure 7. Immersion Ratings for Each Scene (Q3).
Boxplots show participant immersion ratings during scene experience.

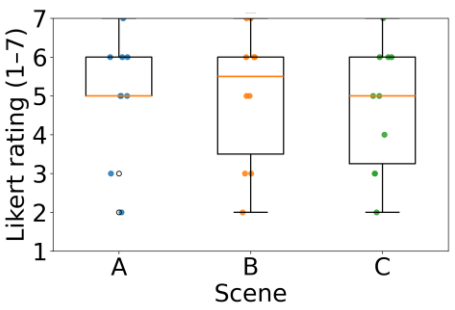

Figure 8. Music–Scene Matching Ratings for Second Music (Q4).
Boxplots show ratings of how well the generated second music matched each scene.

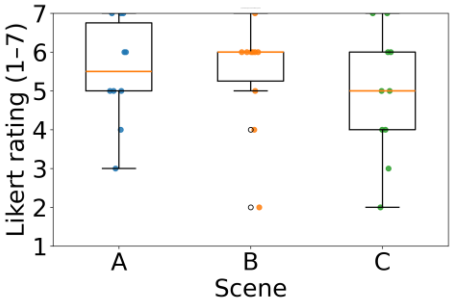

Figure 9. Emotional Impact Ratings of Second Music (Q5).
Boxplots show participant ratings of emotional influence of the generated music.

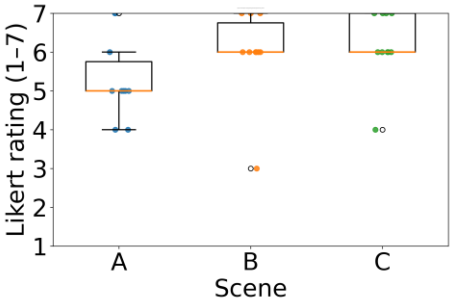

Figure 10. Scene Perception Change Ratings During Second Experience (Q6).
Boxplots show ratings of perceived changes in scene perception during the second experience.

5.7.1 Music Enhanced the Emotional Expressiveness of the Scene

Many participants reported that the music strengthened the emotional expression of the scene, making the emotional tone clearer and easier to interpret. Examples include:

"The music is also trying to describe how sad the whole scene is" "Warm turn to happy"

"A more lingering and simpler sorrow"

These responses suggest that music functioned not merely as a background element, but as an important component of emotional expression. While the visual scene provided emotional cues, the music further amplified these cues, enabling participants to perceive the intended emotions more clearly. These responses suggest that participants perceived the music as enhancing the emotional expressiveness of the scene.

5.7.2 Music Actively Changed Participants' Emotional Experience

Some participants explicitly noted that the music changed how they felt, rather than simply reflecting their existing emotion. Examples include:

"Turn nervous to humor"

"The background music became less sad"

"I sensed a shift in mood"

These responses indicate that music can actively influence the player's emotional state, suggesting that music can serve as an emotional regulation tool within interactive systems. This pattern is consistent with the quantitative results and suggests that dynamically generated music was perceived as influencing participants' experience.

5.7.3 Music Changed Players' Interpretation of the Scene and Narrative Meaning

Beyond emotional effects, some participants reported that music changed how they interpreted the scene's meaning. Examples include:

"It feels like revealing a sad story"

"It seems like the ending of the story"

"There seems to be more to the story"

These responses suggest that music influenced participants' understanding of the scene's narrative structure. Even with unchanged visual content, differences in music led participants to interpret the scene differently. This suggests that participants perceived the music as influencing their interpretation of the scene and narrative context.

5.7.4 Music–Scene Congruence Directly Influenced Immersion

Some participants reported that when the music matched the scene, the experience felt more natural and immersive. For example:

"The second background music is much more suitable"

"It feels more fitting than the first BGM"

Conversely, when music did not match the scene, immersion decreased:

"The first BGM doesn't match the scene and feels strange"

These responses suggest that participants perceived better music–scene congruence as contributing to a more immersive experience.

5.7.5 Individual Differences in the Effect of Music

Although most participants reported that music affected their experience, a small number noted that the impact was limited:

"music cannot change my emotion so much"

This suggests individual differences in sensitivity to music and its effect on experience. Possible reasons include differences in musical sensitivity, immersive tendency, and emotional expressiveness. This suggests that individual differences may be an important consideration for future adaptive music systems.

5.7.6 Summary of Qualitative Findings

The qualitative results suggest that participants perceived dynamically generated music as influencing both emotional and cognitive aspects of their experience.

Specifically, dynamically generated music can:

- Enhance emotional expression of the scene
- Change the player's emotional experience
- Alter players' interpretation and understanding of the scene
- Improve immersion
- Affect players to varying degrees depending on individual differences

These findings suggest that music generation based on estimated emotion categories was perceived as influencing player experience.

However, given the small sample size (N = 10), these qualitative findings should be interpreted cautiously and may not be generalizable to a broader population.

5.7.7 Qualitative Coding Analysis

To further analyze participants' open-ended responses systematically, we applied a thematic analysis approach to categorize and code all responses. Responses were grouped into major categories, and the frequency of each

Table 4. Qualitative Coding Results of Participants' Open-Ended Responses (N = 10).

Category	Description	Count (N = 10)	Percentage
Emotional Enhancement	Music strengthened the scene's emotional expression or clarified the emotion	8	80%
Emotional Change	Music changed the participant's own emotional experience	7	70%
Perception Change	Music changed the participant's interpretation of the story or meaning of the scene	7	70%
Immersion Improvement	Music made the experience feel more natural or immersive	6	60%
Music Matching Improvement	Music was perceived as more fitting for the scene	7	70%
Mismatch	Music did not match the scene and reduced the experience	3	30%
Limited Effect	Music had limited influence on the participant	2	20%

Table 4 summarizes the qualitative coding results, and the frequency of each category was calculated.

The coding results show that most participants reported a noticeable influence of dynamically generated music on either emotional change or a change in their interpretation of the scene. This suggests that participants perceived dynamically generated music as influencing their experience, including emotional perception and interpretation of game scenes.

Furthermore, 60% of participants reported increased immersion, indicating that music plays an important role in enhancing immersive gameplay experiences. Although a small number of participants (20%) reported limited emotional impact, the overall results suggest that dynamically generated music was positively perceived by most participants.

5.8 Emotion Detection System Operation And Integration

The emotion recognition module was operated during gameplay to capture and analyze participants' facial expressions using DeepFace.

The system successfully detected participants' faces and output probability values for multiple emotion categories. The dominant emotion category was transmitted to Unreal Engine via socket communication and used as the primary input for generating music prompts in the music generation module. This enabled integration between facial expression analysis, music generation, and gameplay.

The emotion recognition module operated continuously and stably during the experiment, allowing the system to generate music based on detected emotion categories.

In addition, participants provided self-reported emotional responses after each scene experience. These responses were used as reference information for interpreting participant experience and understanding the relationship between detected emotion categories and subjective perception.

Although formal quantitative accuracy evaluation was not conducted in this study, qualitative observations suggested that the detected emotion categories often aligned with participants' subjective impressions, particularly in happy and sad scenes. Greater variability was observed in angry scenes.

These observations indicate that the emotion recognition module was able to provide continuous emotion category outputs that could be used for music generation within the system. This section describes system operation rather than formal evaluation of recognition accuracy.

This section describes system operation and observed system behavior rather than formal evaluation of emotion recognition accuracy.

5.8 Qualitative Comparison Between Fixed and Generated Music

Although quantitative questionnaire ratings were collected primarily for the generated music condition, participants also provided verbal feedback regarding both the fixed music and generated music during the experiment. These verbal responses were reviewed to provide a qualitative comparison between the two conditions.

Overall, many participants reported that the fixed music was perceived as appropriate and pleasant for the scenes. Participants frequently indicated that the fixed music matched the scene well and provided a consistent emotional atmosphere.

At the same time, participants reported that the generated music was perceived as influencing their emotional experience. Some participants noted that the generated music contributed to changes in emotional perception or interpretation of the scene.

A smaller number of participants reported that the generated music was perceived as having a stronger influence on their emotional experience compared to the fixed music. These participants indicated that the generated music appeared to enhance emotional engagement or provided a more noticeable emotional effect.

These observations suggest that both fixed and generated music were generally perceived as appropriate, while generated music was perceived as influencing emotional experience for some participants. However, subjective responses varied across individuals.

5.10 Summary of Results

The results indicate that participants generally reported positive perceptions of the generated music across multiple evaluation dimensions, including scene understanding, immersion, music–scene matching, emotional impact, and scene perception change.

Qualitative responses further suggest that participants perceived the generated music as influencing their emotional experience and interpretation of the scenes. Comparison with fixed music indicated that both fixed and generated music were perceived as appropriate, while some participants reported that the generated music influenced their experience.

However, because this study was conducted with a small sample size (N = 10), the findings should be interpreted cautiously and may not be generalizable to a broader population. In addition, all evaluation results were based on subjective questionnaire responses and self-reported perceptions rather than objective physiological measurements.

Overall, the results suggest that emotion-category-based music generation was positively perceived by participants within the context of this experimental system.

6 Discussion

The results suggest that music generated based on estimated emotion categories was perceived as influencing player experience. Participants reported that changes in music affected immersion, emotional response, and interpretation of scenes. These findings indicate that music in interactive environments may contribute to perceived coherence between audiovisual elements.

Beyond simple accompaniment, music may function as an affective modulation mechanism, shaping both emotional intensity and narrative interpretation. Even when visual content remained unchanged, alterations in music were reported to modify perceived meaning and emotional tone.

However, individual differences were observed, suggesting that the influence of music varies depending on personal characteristics such as musical sensitivity and emotional responsiveness. These findings highlight the importance of considering personalization in affect-oriented game design.

From a system perspective, the successful integration of emotion estimation and AI-based music generation demonstrates the feasibility of incorporating emotion-category-based music generation into interactive environments.

7 Limitations

This study has several limitations. First, the system does not achieve continuous real-time adaptation, as music is generated between scene experiences rather than updated dynamically during gameplay. Second, emotion estimation relied solely on facial expression analysis and was not quantitatively validated. Third, the experiment involved a small sample size (N = 10) and relied on subjective evaluations.

Future work should incorporate multimodal emotion estimation, larger participant samples, and statistical comparisons to further clarify the role of emotion-category-based music generation in interactive systems.

8. Future Work

Future work will focus on strengthening both technical robustness and empirical validation. First, system logs should be systematically collected to enable quantitative analysis of agreement between estimated emotion categories and participants' self-reported emotional states. Second, experiments with larger and more diverse participant samples are necessary to improve generalizability and statistical reliability. Third, future research should explore more continuous and responsive integration between emotion estimation and music generation.

Beyond technical improvements, further investigation is needed to examine how AI-generated music may function

as an affective modulation mechanism across different gameplay contexts and emotional dimensions. Expanding scene diversity and incorporating objective measures such as physiological signals may provide deeper insight into the relationship between music generation and emotional experience.

9 Conclusion

This study designed and implemented a game system incorporating emotion-category-based music generation and examined its influence on player experience. A user study indicated that participants perceived the generated music as matching the scene and affecting their emotional experience and interpretation.

These findings suggest that music in interactive environments may function as an affective modulation mechanism rather than merely serving as background accompaniment. Although further validation with larger samples and objective measures is necessary, this study provides preliminary insights into the role of AI-generated music in affect-oriented game design.

Acknowledgements

The authors would like to thank all participants who took part in the experiment for their time and valuable feedback.

Reference

[1] M. T. Akbar, M. N. Ilmi, I. V. Rumayar, et al., "Enhancing game experience with facial expression recognition as dynamic balancing," Procedia Computer Science, vol. 157, pp. 388–395, 2019.

[2] L. A. Gerdner, "Effects of individualized versus classical 'relaxation' music on the frequency of agitation in elderly persons with Alzheimer's disease and related disorders,"International Psychogeriatrics, vol. 12, no. 1, pp. 49–65, 2000.

[3] E. G. Schellenberg, T. Nakata, P. G. Hunter, and S. Tamoto,"Exposure to music and cognitive performance: Tests of children and adults,"Psychology of Music, vol. 35, no. 1, pp. 5–19, 2007.

[4] G. Amaral, A. Baffa, J. P. Briot, et al.,"An adaptive music generation architecture for games based on the deep learning Transformer model,"in Proc. 21st Brazilian Symposium on Computer Games and Digital Entertainment (SBGames), IEEE, 2022, pp. 1–6.

[5] Z. Evans, C. J. Carr, J. Taylor, et al.,"Fast timing-conditioned latent audio diffusion,"in Proc. 41st International Conference on Machine Learning (ICML), 2024.

[6] A. Agostinelli, T. I. Denk, Z. Borsos, et al.,"MusicLM: Generating music from text,"arXiv preprint arXiv:2301.11325, 2023.

[7] J. Copet, F. Kreuk, I. Gat, et al.,"Simple and controllable music generation,"Advances in Neural Information Processing Systems (NeurIPS), vol. 36, pp. 47704–47720, 2023.

[8] A. Awana, S. V. Singh, A. Mishra, et al.,"Live emotion detection using DeepFace,"in Proc. 6th International Conference on Contemporary Computing and Informatics (IC3I), IEEE, 2023, pp. 581–584.

[9] P. E. Hutchings and J. McCormack,"Adaptive music composition for games,"IEEE Transactions on Games, vol. 12, no. 3, pp. 270–280, 2019.

[10] C. Plut and P. Pasquier,"Music matters: An empirical study on the effects of adaptive music on experienced and perceived player affect,"in Proc. IEEE Conference on Games (CoG), 2019, pp. 1–8.

[11] P. G. Hunter and E. G. Schellenberg,"Music and emotion,"in Music Perception, New York, NY, USA: Springer, 2010, pp. 129–164.

[12] C. L. Krumhansl,"Music: A link between cognition and emotion,"Current Directions in Psychological Science, vol. 11, no. 2, pp. 45–50, 2002.

[13] A. E. Lopez Duarte,"A progressive-adaptive music generator (PAMG): An approach to interactive procedural music for videogames,"in Proc. ACM SIGPLAN International Workshop on Functional Art, Music, Modelling, and Design, 2024, pp. 65–72.

[14] T. B. Alakus, M. Gonen, and I. Turkoglu,"Database for an emotion recognition system based on EEG signals and various computer games–GAMEEMO,"Biomedical Signal Processing and Control, vol. 60, 101951, 2020.

[15] S. J. Kirsh and J. R. W. Mounts,"Violent video game play impacts facial emotion recognition,"Aggressive Behavior, vol. 33, no. 4, pp. 353–358, 2007.