

Publication Date: 20 August 2024

Archs Sci. (2024) Volume 74, Issue S2 Pages 71-80, Paper ID 2024s29.
<https://doi.org/10.62227/as/74s29>

Exploration of Multimodal Teaching Methods in Emotionally Empowered Music Education

Leqi Tang^{1,*}

¹Xiamen Institute of Technology Liberal Arts Education & Art Media, Xiamen, Fujian, 361021, China.

Corresponding authors: (e-mail: 18659223323@163.com).

Abstract Multimodal as a new teaching concept and mode, applied to musicology professional education is still in the primary stage of development, still lacks a more perfect teaching system. This paper categorizes students' music learning types based on the modal types of sensory channels under multimodal teaching. The intelligent whiteboard technology is proposed to integrate the multimodal teaching technology. In the research of audio emotion recognition technology, the sample audio signal is preprocessed and emotional features are extracted in order to get the emotional features of the audio, and the construction of the WLDNN-GAN music emotion recognition model is completed. The emotional feature analysis of 500 music songs of the subjects shows that the emotional color with the largest standard deviation is "music", with a value of 11.6004, indicating a large degree of dispersion. The multimodal music teaching mode was designed and compared with the traditional music teaching mode to verify the application effect of the multimodal teaching mode. The multimodal teaching mode is basically better than the traditional teaching mode in terms of teaching matching degree, teaching interaction, and teaching effect, and the students' music classroom performance score is 8.2639 under the multimodal teaching mode, which is much higher than the traditional teaching mode.

Index Terms multimodal teaching, audio affective features, WLDNN-GAN model, music education

I. Introduction

Music serves as a pivotal vehicle of culture throughout the vast expanse of human history, evolving in tandem with the progression of human civilization and embodying the essence of human spiritual achievements. The renowned Orff music education system underscores this holistic perspective, stating, "Music is a multifaceted concept: it transcends singular forms, integrating movement, dance, and language into a cohesive artistic whole that transcends mere auditory experience or vocal expression." This notion underscores not only music's inherent nature as a unified entity comprising diverse modalities but also highlights the necessity for multi-sensory engagement in musical learning, as it constitutes the most direct pathway to acquiring skills and fostering emotional attitudes [1-4]. As the avenues for delivering aesthetic education broaden and teaching content diversifies, it becomes increasingly challenging to adhere to traditional, inflexible classroom methodologies. Consequently, the pedagogy of aesthetic education necessitates continuous innovation and efficiency enhancement. To elevate the efficacy of music classrooms, it is imperative to embed practicality within routine lessons, attending meticulously to the modalities employed by educators and being vigilant to students' feedback, thereby promptly identifying and addressing issues [5-8]. Furthermore,

educators should harness modern information technologies to augment their instructional strategies, fostering heightened student engagement. By integrating innovative teaching methodologies, we can strive to achieve the objectives of strengthening artistic expression, deepening aesthetic experiences, sparking creative endeavors, and enriching cultural comprehension—all essential components of reforming aesthetic education [9-12].

Music education underscores the cultivation of students' aesthetic perception, artistic expression, and cultural comprehension. From the perspective of nurturing the "holistic individual," the aesthetic environment and music classroom ought to be vibrant, multifaceted, and adaptable [13-15]. This necessitates addressing various requirements pertaining to three dimensions of the music classroom: the teacher's dimension, the dimension of educational technology, and the students' dimension. Numerous music educators have embarked on this path, consistently integrating innovative pedagogies and emerging information technologies into the classroom to facilitate effective classroom interactivity [16-18]. However, several teaching issues have surfaced, transitioning from the conventional "teacher-centered lecture format" to the "sole reliance on simplistic courseware demonstrations." Instances include teachers' PowerPoint presentations that are detached

from reality, often recycled without modification, and instances where teachers' efforts on the podium fail to ignite students' interest. This scenario may erroneously suggest a lack of enthusiasm on the teacher's side, yet the underlying issue transcends mere enthusiasm; it is about the ineffective engagement of students. Apparent "liveliness" and sensory enrichment can sometimes mask the blind copying of content, neglecting the fundamental nature of classroom teaching as a collaborative process where teaching information is transmitted and received, thereby creating meaning [19-21]. Hence, comprehending students' preferred sensory modalities and interests is paramount for enhancing the efficiency of information reception. This approach aligns with the "student-centered" pedagogy, aiming to elevate the effectiveness of music classrooms. Multimodal teaching underscores the promotion of multi-sensory learning via diverse channels. It is only through fostering multi-sensory experiences that we can offer students tailored and enriching educational encounters, thereby optimizing their learning outcomes.

This paper starts from the multimodal teaching model of music, categorizes the modal types of sensory channels under multimodal teaching, classifies students into visual, auditory, kinesthetic, and tactile learners based on the types of sensory channels, and develops corresponding teaching strategies. Integrate multimodal teaching techniques through intelligent whiteboards. A computational model of music emotion features is constructed for emotion processing in auditory modality. Adopt first-order digital filter to realize music pre-emphasis processing, and ensure the stability of audio signal by shortening the sub-frame length. A non-rectangular window is added to improve the frequency leakage problem in the music pre-processing process. Sequentially using Mel frequency cepstrum, perceptual linear pre, Schenkel analysis, and chord information extraction, the music emotion feature processing is completed. Based on Wacnrt and GAN algorithms, the WLDNN-GAN music emotion recognition model is constructed, and the model is applied to actual music emotion feature recognition. A multimodal music teaching model is designed with this core technology, and the effectiveness of multimodal music teaching is obtained through comparative experiments.

II. Integration of Multimodal Music Teaching Models and Pedagogical Technologies

A. Application of multimodal teaching model in music classroom

1) Types of sensory channel modalities under multimodal instruction

Figure 1 for music teaching multimodal categorization, multimodal teaching mode focuses on students' spiritual realm of aesthetic inculcation, and music education is aesthetic education, music education, the beauty of the ubiquitous, the beauty of this must be "relaxed", "pleasant to the ear," but also This beauty must be "heartwarming" and "pleasant to the ear", and it must be reached through the senses in terms of emotion, imagination and will. In the perspective of multimodal

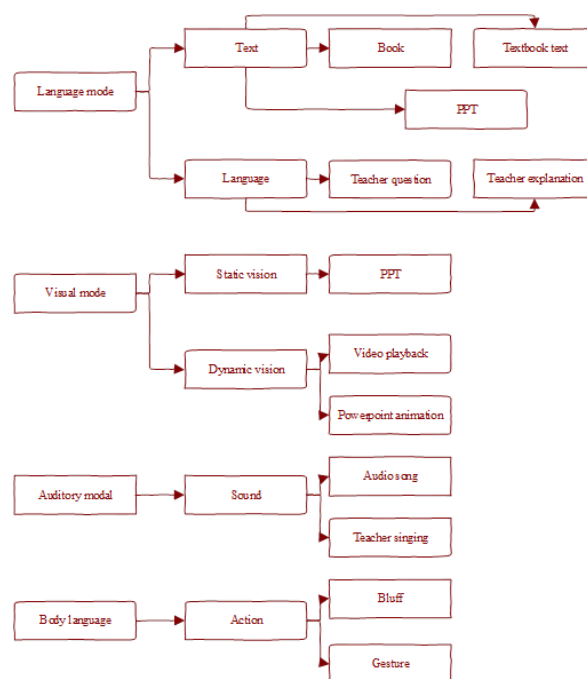


Figure 1: Music teaching multimodal classification

teaching, there are three elements in the classroom: teachers, students and teaching materials, and then there will be text, language, body language and other teaching modalities with meaningful representations, which are conveyed to the students through the sensory channels, and then the sensory modalities are created. It can be said that multimodal teaching exists in fact, even though it is currently receiving little attention from music educators. Therefore, the instructional design of multimodal teaching in the music classroom is a combination of modality and modality, of multiple channels and multiple senses, and it is one of the most common and straightforward concepts.

Multimodal teaching incorporates a myriad of symbols, including text, language, attitude, auditory stimuli, visual elements, and more, into the information transmission process. From this novel perspective, the music classroom transcends being merely a language or communication channel; it becomes a dynamic space where students engage with diverse modalities, encompassing listening, moving, and tactile experiences. The role of the teacher evolves beyond that of a mere presenter or manipulator of lectures; they now serve as the curator of various modal channels, a collaborator in devising teaching methodologies, an active facilitator and demonstrator, and a guide who nurtures students' sensory exploration and preferences. By harnessing multimodal teaching strategies, teachers can enhance students' understanding of their individual sensory orientations and preferences, effectively integrating various instructional approaches to foster innovation and creativity in every music class.

B. Types of student learning in multimodal instruction

Each student possesses unique characteristics in terms of their learning styles and preferred sensory modalities for information reception. Some students may exhibit a keener comprehension of linguistic elements and patterns, while others may have a more profound appreciation for tactile experiences, and yet, there are those who possess a heightened sense of smell. Multimodal teaching acknowledges this diversity and categorizes learners broadly based on their dominant sensory modalities: visual learners, auditory learners, kinesthetic learners, and tactile learners.

1) Visual Learners

They prefer to learn through visual media such as images, graphics, and animations. Visual demonstrations and image-based learning are particularly appealing to this type of learner, enabling them to comprehend various aspects of information effectively from visual sources. However, they may struggle with lengthy textual information, become easily frustrated with pure audio learning, and find themselves easily distracted by movements or vibrant colors. Consequently, when employing multimodal teaching methods, teachers must avoid using tedious audio and inserting irrelevant images that might detract from the focal content. Conversely, judicious use of pertinent images can significantly enhance these students' interest and attention. Furthermore, teachers should encourage students to maintain organized notes or create mind maps as a supplementary learning tool.

2) Auditory learners

Auditory learners prefer to learn primarily through listening, such as by attending to the teacher's explanations, audio playbacks, and appreciate it when the teacher integrates classroom content into a cohesive storyline for teaching purposes. In the context of a music classroom, auditory learners are particularly receptive to this approach.

3) Kinesthetic Learners

Kinesthetic learners learn predominantly through physical movement, preferring to directly experience and comprehend concepts through physical engagement. They excel at expressing themselves through the use of their bodies, dance, and movement. Nevertheless, these students often encounter challenges in maintaining concentration when confronted with books and other passive learning environments, as they require a greater proportion of practical, hands-on experiences compared to theoretical ones to grasp knowledge effectively. Consequently, incorporating activities, dance movements, or role-playing into the music classroom setting is crucial for enriching the learning environment and fostering a deeper understanding among kinesthetic learners.

4) Tactile learners

Tactile learners comprehend concepts more effectively when they engage their sense of touch, preferring to employ fine

motor skills to explore textures. Distinct from kinesthetic learners, tactile learners do not primarily focus on bodily movement but rather on tactile exploration and manipulation of objects. Nevertheless, they too face challenges in passive learning settings, often struggling to maintain concentration during monotonous lectures, and may resort to fidgeting with small objects on their desks as a result. In the context of music lessons, it is imperative to engage tactile learners through play-based learning methodologies and to capitalize on the unique qualities of music by introducing a diverse array of instruments and tactile props into the classroom. This approach enables them to experience music in a tangible and interactive manner, thereby fostering a deeper understanding and appreciation of the subject.

C. Multimodal music teaching based on intelligent whiteboards

The electronic whiteboard represents a high-tech innovation that integrates cutting-edge electronic technology, software engineering, and other advanced research and development methodologies. By leveraging the principle of electromagnetic induction in conjunction with computers and projectors, it facilitates the realization of a paperless environment for both office work and educational instruction. The incorporation of whiteboard technology into music teaching stems from the myriad advantages it offers within this context.

1) Enabling a dynamic teaching process

The utilization of whiteboards enables the integration of dynamic visuals, animations, and other multimedia content, thereby transforming the entire classroom experience into a dynamic and interactive process. This approach facilitates students' acquisition of a broader range of knowledge and skills. For instance, in the context of music teaching during the celebration of the New Year, the whiteboard can be employed to display videos and textual materials pertaining to the New Year festivities during the introductory phase of the course. This not only familiarizes students with the upcoming topic but also encourages their active participation in subsequent teaching activities, immersing them in the festive ambiance and enabling them to personally experience the joy and spirit of the New Year. Such an approach is significantly more engaging and captivating than solely relying on verbal explanations from the teacher.

2) Promote student interaction and cooperation

The use of whiteboard technology can simplify the teaching burden of teachers to a large extent, and transform more content that needs to be told directly into a dynamic process. Through the guidance of video, pictures, animation and other content, students' interactive and cooperative activities will have objects that can be imitated, and the effect of cooperative exchanges will be better. In the study of "I love China", I can use whiteboard technology to create a realistic situation for students. Students are then asked to work interactively in groups to sing these songs. This gives students a basis for

interactive cooperation and allows practical activities to lead to the improvement of students' practical skills.

3) Diversification of teaching methods

By harnessing the advancements of modern technology, the electronic whiteboard serves as a versatile tool that significantly enriches the teaching methodologies employed by educators. Beyond merely presenting teaching content, instructors can leverage the capabilities of the whiteboard to engage in interactive board writing sessions, display engaging teaching activities, and much more.

III. Multimodal Music Emotion Recognition Technology and Analysis

In this paper, we address the challenges pertaining to auditory emotion processing by focusing on the cognitive laws governing auditory emotion in the brain. We integrate methods for emotion analysis and recognition of both sound signals and electroencephalography (EEG) signals, thereby embarking on an investigation into the cognitive mechanisms underlying auditory emotion and developing a computational approach for audio emotion analysis.

A. Music pre-processing

1) Pre-emphasis

Because the quality of the read music data varies, to avoid the music data in the transmission process of signal weakening, so take the means of pre-emphasis to improve the high-frequency part. For a sound signal such as music, the energy is mainly distributed in the low-frequency band, and the power spectral density of the sound decreases as the frequency increases, which will have a great impact on the signal quality. Pre-emphasis is realized by a first-order digital filter with the formula:

$$H(z) = 1 - \mu z^{-1}. \quad (1)$$

In Eq. (1), the pre-emphasis factor μ is a constant decimal close to 1, and z represents the input signal, which is passed through this digital filter, and the output at moment n is:

$$\overline{Y(n)} = Y(N) - \mu Y(n-1). \quad (2)$$

2) Framing

In the realm of music processing, the duration of sub-frames is typically set within the range of 10-40 milliseconds (ms). From a macroscopic perspective, the frame length must be sufficiently short to ensure that the signal within the frame remains stable and approximately time-invariant. This is due to the fact that the duration of instrument articulation loudness generally spans from 40 to 200 ms. Consequently, the frame length is commonly chosen to be less than 40 ms. On the other hand, from a microscopic viewpoint, the sub-frame length must encompass at least one complete vibration cycle, implying that it must capture the fundamental frequency of the instrument, which often approximates 100 Hz. Given that the fundamental frequency of a piano is approximately 100 Hz,

the sub-frame duration is typically set to be at least 10 ms to accommodate this requirement.

The selection of speech frame lengths can significantly influence the performance of a given task, emphasizing the importance of choosing an appropriate frame length. To ensure a seamless transition between consecutive voice frames, it is crucial to incorporate overlap between frames. This overlap, measured by the temporal shift in the starting position between two adjacent frames, is commonly referred to as the frame shift. Typically, the frame shift is set to be either half (1/2) or three-quarters (3/4) of the frame length.

3) Adding windows

In order to make an improvement in the frequency leakage, the method of adding a non-rectangular window will be used, and the usual solution is to add a Hamming window, which is very suitable for this experiment because it has the property of a large sidelobe attenuation, which is able to attenuate the peak of the main and the peak of the first sidelobe by up to 40db. The formula is as follows, where a_0 is taken as 0.53836:

$$\omega(n) = a_0 - (1 - a_0) \bullet \cos\left(\frac{2\pi n}{N-1}\right), 0 \leq n \leq N-1. \quad (3)$$

B. Music emotion feature extraction

1) Mel frequency cepstrum coefficients

A musical piece can be segmented into numerous frames, with each frame undergoing Fast Fourier Transform (FFT) computation to yield a corresponding spectrum. This spectrum captures the relationship between frequency and energy, specifically, the variation in amplitude across different frequencies. The sound spectrogram, which is a visual representation of the spectrum over time, should comprehensively reflect the relationships between all frequencies and their corresponding energies. As a result, the output is often a substantial map. To extract sound features that are pertinent for the ultimate goal of music emotion recognition, this map is typically processed through a Mel-scale filter bank, transforming it into a Mel spectrum. The transformation from the conventional frequency scale to the Mel frequency scale is carried out as outlined below, and the mapping relationship is detailed as follows:

$$Mel(f) = 2595 \bullet \log_{10}(1 + f/700), \quad (4)$$

where f denotes the frequency (in Hz).

2) Perceived linear prediction

Music is not a simple splicing of information from frame to frame, so it is necessary to add some features to indicate that the cepstrum coefficients are changing with time, i.e., when making the inherently static waveform graph move. We call Δ^* , Δ^{2*} as Δ features, that is, dynamic features to time derivation, cepstrum features Δ features of the formula is shown below:

$$d(t) = \frac{c(t+1) - c(t-1)}{2}, \quad (5)$$

$c(t)$ is the dynamic features of music at 2 moments.

Figure 2 shows the perceptual linear prediction computation process, through the derivation of the way can reflect the dynamic characteristics of the music information to a certain extent, can be more optimized for the extraction of features.

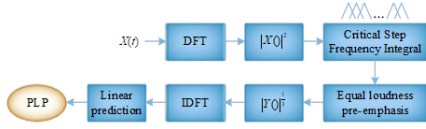


Figure 2: Perceptual linear prediction calculation process

3) Schenkel analysis

The central idea of Schenker's theory is melodic reduction, i.e., retaining the backbone of the musical weight and removing some of the other passing tones and ornamental tones, which play little role in the expression of the main musical feeling.

This is computed as a string of notes N , where all possible Schenkel analyses of N can be enumerated as $A_1 \dots A_m$, and for some integer m we wish to obtain the most probable analysis as $\arg \max_i P(A_i|N)$, with $T_{i1}, \dots, T_{ip}$ denoting the set of triangles that make up the analysis A , where L_{ij}, M_{ij}, R_{ij} is the three vertices of the triangle T_{ij} , with the formula shown in (6) below:

$$\begin{aligned} \arg \max_i P(A_i|N) &= \arg \max_i P(T_{i1}, \dots, T_{ip}) \\ &= \arg \max_i \prod_{j=1}^p P(T_{i,j}) \\ &= \arg \max_i \prod_{j=1}^p P(M_{i,j}|L_{i,j}, R_{i,j}). \end{aligned} \quad (6)$$

The problem of computing the probability $P(M_{ij}|L_{ij}, R_{ij})$ in which the musical information embedded in M_{ij}, L_{ij}, R_{ij} is represented by 18 feature values that represent the melodic, harmonic, beat and timing information of the music. Each type of triangle is determined by a combination of 18 features with too large a dimension, so a random forest approach is used to learn the conditional probability $P(M_{ij}|L_{ij}, R_{ij})$, where the set of 18 features contains a subset of six features that depend on the middle tone M . A single random forest cannot adequately learn to predict all six features, so the conditional probability is decomposed into six different products:

$$\begin{aligned} P(M|L, R) &= P(M_1, M_2, \dots, M_6|L, R) \\ &= P(M_1|L, R) * P(M_2, M_3, \dots, M_6|M_1, L, R) \\ &= P(M_1|L, R) * P(M_2|M_1, L, R) \\ &\quad * P(M_3, M_4, \dots, M_6|M_1, M_2, L, R) \\ &= P(M_1|L, R) * P(M_2|M_1, L, R) \\ &\quad \dots P(M_6|M_1, \dots, M_5, L, R). \end{aligned} \quad (7)$$

4) Chord information extraction

Using the backpropagation function, the loss E_1 is computed from the inputs and updated inversely for W . Assuming that the output value is u , i.e., $hW = u$, the predicted value is:

$$y_i = \text{Softmax}(u_j) = \frac{e^{u_j}}{\sum_{k=1}^V e^{u_k}}. \quad (8)$$

For an input of length N , the objective function is as follows:

$$\begin{aligned} E &= \arg \max_{W, W'} \prod_{i=1}^N \prod_{j=i-w}^{i+w} P(y_j|y_i) \\ &= \arg \max_{W, W'} \prod_{i=1}^N \prod_{j=i-w}^{i+w} P(y_j|y_i) \\ &= \arg \max_{W, W'} \sum_{i=1}^N \sum_{j=i-w}^{i+w} \log P(y_j|y_i) \\ &= \arg \max_{W, W'} \sum_{i=1}^N \sum_{j=i-w}^{i+w} \log \frac{e^{w_j}}{\sum_{k=1}^V e^{u_k}}. \end{aligned} \quad (9)$$

C. Music emotion recognition model construction

WaveNet is an autoregressive probabilistic model that models the joint probability distribution of musical information $x = \{x_1, \dots, x_T\}$ as $p(x) = \prod_{t=1}^T p(x_t|x_1, \dots, x_{t-1})$. This modeling approach is similar to DeepAR.

WaveNet model where part 1 is: the input layer of WaveNet, which is used to receive the input of information, part 2 is the causal convolutional layer, part 3 is the expanded convolutional network of WaveNet, which is the most central part of the whole network, and part 4 is the output layer, and the output information is convolved by two ReLU and two $1*1$, and finally outputted by Softmax function.

The residual block is internally using the gate activation function with the formula:

$$\text{output} = \tanh(W_{f,k} * \text{input}) \odot \sigma(W_{g,k} * \text{input}). \quad (10)$$

This formula outperforms ReLU in activating audio signals. The WaveNet model provides a good modeling idea for the processing aspect of audio information, which can achieve a large sensory field even with a small number of layers, and can be better used for learning emotional features of music and automatic labeling of basic information of music.

1) GAN

The objective function to be optimized by the network is:

$$\begin{aligned} \min_G \max_D V(D, G) \\ = E_{x \sim p_{data}(x)} [\log D(x)] + E_{z \sim p(z)} [\log(1 - D(G(z)))]. \end{aligned} \quad (11)$$

Specifically for G network optimization the objective function is:

$$\begin{aligned} L &= \log(1 - \text{sigmoid}(D(G(z)))) \\ &\approx -\log(\text{sigmoid}(D(G(z)))). \end{aligned} \quad (12)$$

The objective function of optimization for a D network is:

$$L = -\log(\text{sigmoid}(D(x))) + \log(1 - \text{sigmoid}(D(G(z)))) \\ \approx -\log(\text{sigmoid}(D(G(z)))) \quad (13)$$

D. WLDNN-GAN modeling

Firstly, the preprocessed data is input into the Wavenet layer to extract trainable music emotion features that correlate across different data points. By employing null causal convolution, the learning of music emotion features becomes more efficient, achieving better results with reduced dimensionality. Subsequently, these features are integrated with the LSTM layer, which effectively summarizes the preceding musical information. This allows the subsequent portions of the model to more readily classify emotions from the input, while the final DNN layer deepens the previous hidden layers and fuses the processed features from Wavenet, transforming the enriched information into a discriminative feature space.

Initially, the Generative Adversarial Network (GAN) was proposed for data generation. However, the emotion recognition problem addressed in this paper necessitates a classification model, requiring modifications to the original GAN framework. To this end, we devise a feature separation framework grounded in GANs specifically for emotion recognition. This framework utilizes constraints such as partial feature exchange, content loss, center alignment, classification loss, and adversarial loss to facilitate a direct and highly purified separation of music emotion-related features from irrelevant ones. During classification, the irrelevant features are subsequently disregarded, mitigating their influence on the classification task and thereby enhancing classification accuracy.

E. Results of music sentiment analysis

1) Overall musical sentiment analysis

Employing the music emotion analysis model developed in this study, the sample music was subjected to an emotional analysis, resulting in the classification of music emotions into seven distinct emotional categories: joy, happiness, anger, sadness, fear, eeriness, and surprise.

Figure 3 presents a descriptive analysis of the overall music emotion dataset. Through the application of descriptive statistics to the emotion analysis data, we observe the relationships and correlations among the relevant values. Given the significant variation in the overall emotional value across individual songs, the scores for the seven emotional categories (good, joy, anger, sadness, fear, eeriness, and surprise) among the 500 songs exhibit substantial differences. A direct statistical analysis of these emotional scores would reveal a high degree of dispersion, thereby limiting the reference significance. Consequently, we focus on describing the proportion of each of the seven emotional categories within the total emotional value of the 500 songs. The standard deviation serves as a metric to quantify the dispersion of data around its mean. A larger standard deviation indicates that the data points are more dispersed from the mean, whereas a smaller standard deviation implies that the data values are closer to the

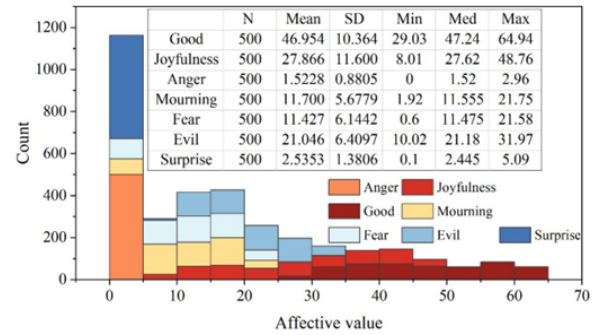


Figure 3: Descriptive analysis of overall music emotional data

mean. In this context, the emotional category "joy" exhibits the largest standard deviation of 11.6004, suggesting a high degree of dispersion in the data. Conversely, the emotional category "anger" displays the smallest standard deviation of 0.8805, with a minimal range between the maximum and minimum values (1.52), indicating stability and a low degree of dispersion. The substantial difference between the largest and smallest standard deviations (10.7199) underscores the varying degrees of dispersion across the data.

2) Song-specific emotional validation

Three classical music pieces were selected, and their affective values were quantified through music modeling. The outcomes of the subjects' music affective analysis are presented in Figure 4.

In Song 1, the emotional hues of "good" and "happy" constitute 66.37% of the total, indicating a dominant presence and resulting in a relaxed and optimistic emotional analysis. This high proportion of positive hues suggests a positive emotional tendency. Conversely, the emotional hues of "sadness," "shock," and "fear" account for 13.22%, 3.04%, and 1.52% respectively, which are relatively lower in comparison. Notably, the hue "evil" comprises 15.48% of the emotional breakdown, albeit less significant than the positive hues, suggesting a minority presence. Collectively, these findings reveal that Song 1 exhibits a predominantly positive and optimistic emotional profile, with a relatively small proportion of negative hues such as "sadness," "fear," and "shock." Given the overall balance and prevalence of positive hues, this song can be classified within the emotion category of "smoothness and gentleness," which encapsulates its characteristic emotional tone. Consequently, it is a fitting example within this broader emotion category.

Based on the values presented in the chart, the emotional profiles of Song 2 and Song 3 are analyzed sequentially. For Song 2, the cumulative proportion of positive emotional hues amounts to 73.51%, exceeding half of the total and surpassing the average benchmark, thereby aligning the emotional analysis of the song with the category of positive emotions. Conversely, in Song 3, the emotional hues of "good" and "joy" constitute 35.78% and 15.63% respectively, which are

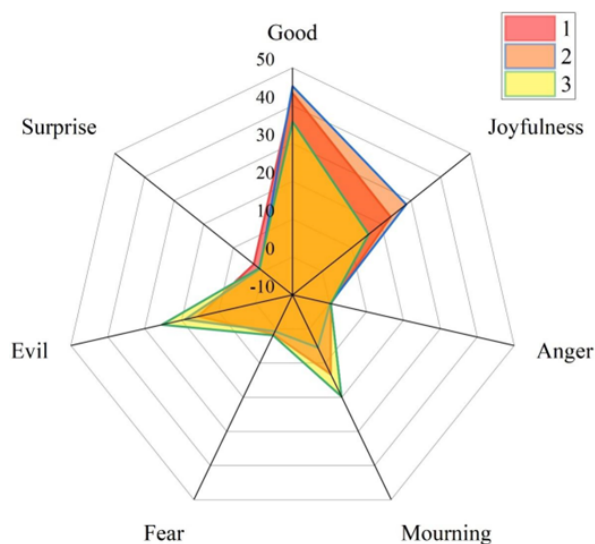


Figure 4: The results of the analysis of the emotional analysis

notably below the average levels for these two hues. As a result, the proportion of positive emotions, when compared to other emotional categories, is relatively low. Additionally, the emotional hues of "sorrow" and "evil" occupy 19.76% and 25.36% respectively, both exceeding their respective averages, suggesting that Song 3 falls within the category of sadness and pessimism.

IV. Research Design for the Application of Multimodal Teaching in the Music Classroom

A. Multimodal music instructional design strategies

This chapter delves into students' emotional experiences with songs in the context of music learning, drawing upon the aforementioned music emotion recognition techniques. Subsequently, the chapter examines the formulation of pertinent music learning strategies in light of these emotional experiences.

1) Establishment of teaching programs

Teaching objectives are established in accordance with the needs of students for developing musical literacy, emphasizing the four pivotal aspects of aesthetic perception, artistic expression, creative practice, and cultural understanding. These objectives are further tailored to the requirements outlined in the music curriculum standards. When setting unit-specific teaching objectives, adherence to the materials prescribed in the national curriculum is crucial, ensuring that they neither exceed the syllabus nor anticipate future learning outcomes. Additionally, the setting of music teaching objectives must be grounded in the actual circumstances of the students. Teachers must be cognizant of the students' current learning abilities and, based on their existing capabilities, devise appropriate learning unit content. Consequently, the implementation of a thorough pre-course student learning assessment is paramount to achieving this end.

2) Study music curriculum standards and materials

The most salient feature of the music curriculum standard lies in its prioritization of musical literacy as the cornerstone of student training, thereby further elucidating the central role of musical literacy in shaping the positioning of music education. By placing the promotion of students' holistic development at the forefront, it ensures ample preparation for students to navigate societal advancements and lead healthy lives, simultaneously laying a robust foundation for their lifelong learning and development.

3) Integration of pedagogical content

To facilitate students' mastery of musical concepts, teachers must promptly evaluate their learning progress. Adhering to the music curriculum standards, educators should meticulously observe whether students have achieved the objectives of teaching and learning activities, fulfilled their assigned tasks, and met the criteria for developing musical literacy. This assessment serves as the cornerstone for classroom, annual, and ultimately, academic quality evaluations. The content of instructional materials is systematically integrated and interconnected, adhering to this evaluation approach, ensuring that the materials are seamlessly incorporated into music teaching with an emphasis on nurturing students' holistic development. Furthermore, students' capacity to tackle practical issues in society is seamlessly integrated into classroom instruction through subject-specific knowledge, ultimately fostering their musical literacy within the music discipline as the ultimate goal.

B. Analysis of the effectiveness of multimodal music instruction

1) Analysis of the effect of multimodal music teaching application

In this research, the questionnaire data collected both prior to and subsequent to the implementation of multimodal teaching were analyzed using a paired-samples t-test in SPSS 21.0. The survey instrument was structured into four distinct sections: affective attitudes and values, process and methodology, knowledge and skills acquisition, and overall satisfaction. A five-point Likert scale was employed, where a score of 1 represented complete disagreement, 2 signified disagreement, 3 denoted neutrality or fairness, 4 conveyed agreement, and 5 stood for complete agreement.

Figure 5 illustrates the outcomes of the paired samples t-test, revealing that the overall mean post-multimodal teaching was 1.0185 units higher than that prior to the intervention, with a standard error of the mean (SEM) of 1.265. Statistically significant differences ($p < 0.01$) were observed across all four dimensions of the questionnaires administered before and after the implementation of multimodal teaching. Consequently, the efficacy of music teaching subsequent to multimodal teaching was notably enhanced compared to pre-intervention levels, indicating that the multimodal teaching classroom model contributes to the improvement of music

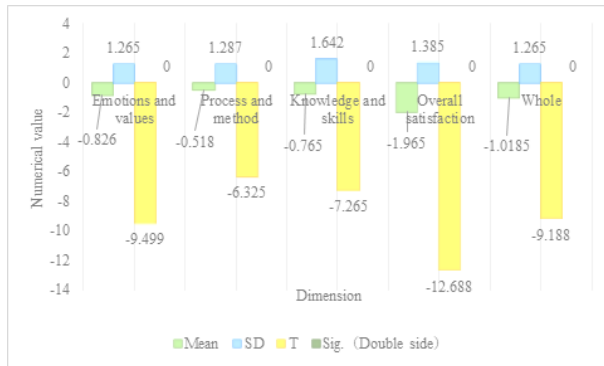


Figure 5: Matching sample t test results

teaching and learning outcomes. Furthermore, this teaching modality facilitates the enhancement of music classroom instruction's effectiveness. Analyzing the changes in the mean values across dimensions, the most pronounced increase was observed in overall satisfaction, with a difference of 1.965 between pre- and post-multimodal teaching. It is imperative to recognize that music teaching should not merely be regarded as imparting specialized knowledge to students but rather as fostering an appreciation of music as a source of joy and pleasure. Therefore, in the teaching process, emphasis should be placed on the methodology and approach adopted by students, emphasizing their experiential journey with music.

2) Analysis of the effectiveness of music teaching

Table 1 presents an analysis of the effectiveness of music teaching, specifically examining the impact of teaching methodology on student comfort levels. The results indicate that students who were exposed to the traditional teaching model of the music course reported significantly less comfort with the teaching approach, in contrast to those who underwent the multimodal teaching model, who expressed a high degree of comfort with the manner in which the course was conducted. Notably, the difference in adaptation to the classroom teaching style between the multimodal and traditional instructional approaches amounted to 1.56 units on the relevant scale. In terms of teaching objectives, students instructed via the traditional teaching mode reported a notably clear understanding of the objectives and tasks of the music course. Conversely, students exposed to the multimodal teaching approach indicated an average level of comprehension regarding the teaching objectives and tasks, which seemingly correlated with their subpar performance. Consequently, the degree of alignment between the teaching objectives and student understanding, as measured by the value of the match, was slightly inferior to that observed in the traditional teaching mode.

In terms of teaching interactivity, the mean values pertaining to teacher-student interaction, student-student interaction, and the variety of interaction modes were all notably higher in the multimodal teaching mode compared to the traditional teaching mode, exhibiting differences of 1.1704,

0.513, and 0.0791 units, respectively. Analogously, in assessing teaching effectiveness, the scores achieved under the multimodal teaching mode surpassed those of the traditional mode across various dimensions, including self-evaluations of learning efficacy, classroom performance, perceived learning pressure, and evaluations of the teaching approach itself. This superiority is particularly evident in the realm of classroom performance, where the average score of students' classroom performance in the multimodal teaching mode, at 8.2639, significantly outpaces that of the traditional music teaching mode.

V. Conclusion

In this paper, we commence by categorizing the modal types of sensory channels employed within the framework of multimodal teaching, with the aim of identifying the corresponding types of students' music learning experiences. To integrate the information derived from these multimodal teaching techniques, we utilize intelligent whiteboard technology. Subsequently, we constructed the WLDNN-GAN music emotion computation model, specifically tailored to address the challenge of auditory emotion processing. The practical efficacy of this model is rigorously validated through comparative case studies. The subsequent sections present the outcomes of both the music emotion analysis and the analysis of the music teaching's effectiveness.

- 1) In our analysis of seven types of music-evoked emotions across 500 musical subjects, the emotion category exhibiting the greatest variability, as indicated by the largest standard deviation, is "music" itself, with a value of 11.6004, suggesting a wide dispersion of responses. Conversely, the emotion with the smallest standard deviation, indicative of more uniform responses, is "anger", having a deviation of just 0.8805. It appears that there was an oversight in stating the maximum and minimum standard deviation values, as they were erroneously repeated as the same (0.8805). The actual difference between the highest (11.6004) and lowest (0.8805) standard deviations amounts to 10.72, highlighting a substantial disparity in the dispersion of responses across different emotion categories.
- 2) For the purpose of specific analysis, three songs were carefully selected from a pool of 500 songs. In song 1, the combined percentage of the emotional colors "good" and "happy" amounted to 66.37%, indicating a substantial proportion of positive sentiments and thus suggesting a positive emotional inclination. Consequently, the emotional classification of this song can be deemed as relaxed and optimistic.
- 3) A paired-sample t-test was conducted to compare the outcomes before and after the implementation of the multimodal teaching model. Subsequent to the application of the multimodal music teaching approach, the overall level witnessed a statistically significant increase of 1.0185 units, accompanied by a standard error of 1.265. This finding was confirmed by a p-value less

Teaching matching						
Index	Teaching mode	N	Max	Min	Mean	SD
Course teaching	Tradition	55	1	5	2.5965	1.5955
	Multimodal state	55	1	5	4.1565	0.6244
Teaching target	Tradition	55	1	5	4.6855	0.8611
	Multimodal state	55	1	5	4.4255	1.1455
Teaching content	Tradition	55	1	5	3.5522	1.2658
	Multimodal state	55	1	5	3.7952	1.2684
Learning demand	Tradition	55	1	5	3.8564	1.2965
	Multimodal state	55	1	5	3.9755	1.2648
Learning interactivity						
Student interaction	Tradition	55	1	5	3.4855	1.2684
	Multimodal state	55	1	5	4.6559	1.0325
Biointeraction	Tradition	55	1	5	3.7554	1.1982
	Multimodal state	55	1	5	4.2684	1.2682
Interactive mode	Tradition	55	1	5	3.9154	1.1544
	Multimodal state	55	1	5	3.9945	2.0589
Teaching effect						
Learning effect self-evaluation	Tradition	55	1	5	3.7811	1.2654
	Multimodal state	55	1	5	4.0366	2.0644
Classroom performance	Tradition	55	1	10	6.5944	1.1658
	Multimodal state	55	1	10	8.2639	1.1765
Learning pressure	Tradition	55	1	5	3.4752	1.3264
	Multimodal state	55	1	5	3.6748	2.6748
Teaching method evaluation	Tradition	55	1	10	7.1679	1.2647
	Multimodal state	55	1	10	7.9958	1.5956

Table 1: Analysis of the effect of music teaching

than 0.01, indicating a high level of statistical significance. Notably, errors were observed across all four dimensions under consideration. Among these, the most pronounced effect was observed in the improvement of overall satisfaction, which exhibited an amplitude of enhancement amounting to 1.965 units.

- 4) In the comparative analysis of traditional music teaching and the multimodal teaching mode, students reported a higher level of adaptability towards the latter, exhibiting a notable difference in adaptability scores of 1.56 points. Regarding pedagogical compatibility, the multimodal teaching mode surpassed the traditional teaching mode in all other assessed dimensions, albeit marginally falling short in terms of meeting specific teaching objectives as compared to the traditional approach.

In conclusion, the multimodal music teaching mode proposed in this paper has a good application effect.

References

- [1] Laes, T., & Westerlund, H. (2018). Performing disability in music teacher education: Moving beyond inclusion through expanded professionalism. *International journal of music education*, 36(1), 34-46.
- [2] Sirek, D. (2018). Our culture is who we are!“Rescuing” Grenadian identity through musicking and music education. *International Journal of Music Education*, 36(1), 47-57.
- [3] Hayes, L. (2017). Sound, electronics, and music: A radical and hopeful experiment in early music education. *Computer Music Journal*, 41(3), 36-49.
- [4] Tahirbegi, D. (2022). Higher music education students’ experiences and management of performance anxiety: A qualitative study. *Psychology of Music*, 50(4), 1184-1196.
- [5] Cayari, C. (2019). Demystifying trans*+ voice education: The transgender singing voice conference. *International Journal of Music Education*, 37(1), 118-131.
- [6] Schneider Grings, A. F., & Hentschke, L. (2017). Attributional Theory in investigating public music performance in higher music education. *International Journal of Music Education*, 35(1), 31-46.
- [7] Jorritsma, M. (2022). Towards an eco-literate tertiary music education: Notes from a South African context. *International Journal of Music Education*, 40(1), 14-25.
- [8] Shaw, R. D., & Auletto, A. (2022). Is music education in tune with the pursuit of equity? An examination of access to music education in Michigan’s schools. *Journal of research in music education*, 69(4), 364-381.
- [9] Richerme, L. K. (2022). Naming moral-political discourses in music education: a philosophical investigation. *International Journal of Music Education*, 70(1), 48-65.
- [10] Fitzpatrick, K. R. (2022). “Finding the Other Half of Me”: Culture-Based Approaches to Music Education in Hawaii. *International Journal of Music Education*, 70(1), 22-47.
- [11] De Villiers, R. (2021). The Force Field Model applied to a Music Education teacher training framework in a South African context. *British Journal of Music Education*, 38(3), 219-233.
- [12] Machado, J., Ody, L. C., & Ilari, B. (2021). Popular education and music education: Lessons from a strings program in Erechim, Brazil. *International Journal of Music Education*, 39(1), 80-92.
- [13] Forrester, & Eros. (2020). Nontraditional students in traditional music education programs: a case study. *Bulletin of the Council for Research in Music Education*, (223), 42.
- [14] Yang, Y., Yin, A., & Guan, T. (2021). Mapping music education research in Mainland China (2007–2019): A metadata-based literature analysis. *International Journal of Music Education*, 39(2), 234-246.
- [15] Shevock, D. J. (2021). Music education for social change: constructing an activist music education: by Juliet Hess, New York, NY, Routledge, 2019, 211 pp.
- [16] May, B. N., Broomhead, P., & Tsugawa, S. (2020). A music literacy-based rationale for popular music ensembles and experiences in music education. *International Journal of Music Education*, 38(3), 470-491.
- [17] Prendergast, J. S. (2021). Music education and educators in Missouri, Iowa, and Illinois. *International Journal of Music Education*, 69(2), 228-243.
- [18] de Villiers, A. C. (2021). (Re) organizing the music curriculum as multi-cultural music education. *International Journal of Music Education*, 39(4), 383-393.
- [19] Conway, C., Vaughan-Marra, J., & Marra, C. (2021). Balancing the varied needs of Master of Music Education students: Perceptions from program directors and alumni. *Journal of Research in Music Education*, 69(1), 105-120.

- [20] Woody, R. H. (2021). Music education students' intrinsic and extrinsic motivation: A quantitative analysis of personal narratives. *Psychology of Music*, 49(5), 1321-1343.
- [21] Westerlund, H. M. (2020). Stories and narratives as agencies of change in music education: narrative mania or a resource for developing transformative music education professionalism?. *Bulletin of the Council for Research in Music Education*, (223), 7-25.
- [22] Jorgensen, E. R. (2020). Life, liberty, and the pursuit of happiness: Values for music education. *Bulletin of the Council for Research in Music Education*, (226), 66-79.
- [23] Niknafs, N. (2020). Music education as the herald of a cosmopolitan collective imperative: On being human. *International Journal of Music Education*, 38(1), 3-17.
- [24] Crawford, R. (2019). Using interpretative phenomenological analysis in music education research: An authentic analysis system for investigating authentic learning and teaching practice. *International Journal of Music Education*, 37(3), 454-475.

...