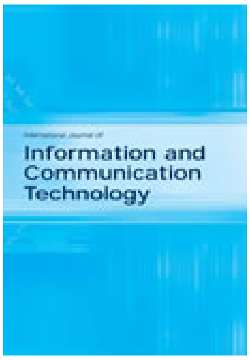


International Journal of Information and Communication Technology

ISSN online: 1741-8070 - ISSN print: 1466-6642
https://www.inderscience.com/ijict

Generative music composition teaching system based on mobile interaction

Yuanyuan Xie, Xiaole Zhu, Xing Hu

DOI: 10.1504/IJICT.2026.10075963

Article History:

Received:	01 December 2025
Last revised:	28 December 2025
Accepted:	30 December 2025
Published online:	11 February 2026

Generative music composition teaching system based on mobile interaction

Yuanyuan Xie, Xiaole Zhu and Xing Hu*

Xiamen Huaxia University,
Xiamen, 361021, China
Email: xieyy@hxxy.edu.cn
Email: zhuxl@hxxy.edu.cn
Email: hxingmx@sina.com
*Corresponding author

Abstract: Current music composition teaching systems face challenges such as limited resource quality and insufficient real-time interactivity. To address these issues, this paper proposes a generative music composition teaching system based on mobile interaction technology. The system first designs a music teaching resource generation module utilising multi-scale feature filtering, combined with a multi-discriminator structure to enhance the discriminative capability of generated samples. Building upon the generation of rich music teaching resources, this paper introduces an expandable network communication module and an interactive collaboration module supporting bidirectional collaborative control and user state management. Experimental results demonstrate that the designed system achieves a CPU utilisation rate of 43% and a single interaction response time of only 11.8ms. It not only generates high-quality music creation resources but also exhibits outstanding real-time interactive performance, holding significant value for advancing the widespread application of interactive mobile teaching.

Keywords: music composition teaching; music generation; mobile interaction; feature selection; collaborative learning.

Reference to this paper should be made as follows: Xie, Y., Zhu, X. and Hu, X. (2026) 'Generative music composition teaching system based on mobile interaction', *Int. J. Information and Communication Technology*, Vol. 27, No. 9, pp.89–112.

Biographical notes: Yuanyuan Xie is a Lecturer at Xiamen Huaxia University, China. She obtained a Bachelor's degree (2014) and a Master's degree (2017) from Fuzhou University. She has published two academic papers. Her research interests include music and technology, music education, and arts management.

Xiaole Zhu is a Professor at Xiamen Huaxia University, China. He earned a Bachelor's degree from Xi'an Conservatory of Music (2005) in China, followed by a Master's degree (2008) and a PhD (2021) from the Odessa State Conservatory of Music in Ukraine. He has published one SSCI-indexed paper and one CPCI-SSH-indexed paper. His research interests include music education, piano performance, and music and artificial intelligence.

Xing Hu is an Associate Professor at Xiamen Huaxia University, China. She obtained a Bachelor's degree in Advertising from Guizhou Minzu University (2006) and a Master's degree in Advertising from China Media University (2010). She has published 1 CPCI-indexed paper and 2 EI-indexed papers. Her research interests include new media operations and art communication.

1 Introduction

As the digital music industry rapidly growing, generative music creation has emerged as a new frontier in music education due to its innovative and efficient nature. However, current generative music instruction heavily relies on specialised equipment and fixed settings, presenting challenges such as high operational barriers, limited interactivity, and restricted teaching scenarios. These limitations hinder learners' ability to engage in creative practice anytime and anywhere, thereby constraining the widespread adoption of generative music creation in public education (Zhang, 2023). The rapid advancement of mobile interaction technologies offers new possibilities for addressing these challenges. The widespread adoption of mobile devices such as smartphones and tablets, coupled with the maturation of interaction technologies like touchscreens, gesture recognition, and sensors, can overcome the temporal and spatial constraints of traditional teaching. This enables the creation of more immersive and convenient creative environments for learners (Sun and Zang, 2025; Hernandez-Olivan and Beltran, 2022). Integrating mobile interactive technology with generative music composition instruction not only lowers the barrier to entry for creators but also enhances teaching efficiency and learner engagement through real-time feedback and personalised guidance. This approach aligns with current trends toward digitalisation and personalisation in education (Cheng, 2025; Larsson and Georgii-Hemming, 2019).

The core of mobile interactive learning lies in empowering education with technology and reconstructing the teaching-learning-assessment relationship through personalised, interactive and data-driven design. Its advantages are not only reflected in the improvement of efficiency, but also in cultivating learners' autonomous learning ability and digital literacy, laying the foundation for a lifelong learning society. In the future, with the development of technologies such as 5G, AI, and the metaverse, mobile interactive learning will further integrate virtual and real scenarios, promoting the evolution of education towards a smarter and more inclusive direction. Mobile interactive learning emphasises students' self-directed learning, which is a dynamic communicative and collaborative learning approach formed between teachers, students, and peers through the use of mobile devices and technology to promote music composition teaching effectiveness (Dai et al., 2025). Tong (2016) points out that in a mobile interactive teaching environment, teachers can adopt multiple teaching methods to facilitate interactions among peer students, teacher-student interactions, and with music composition teaching resources. Zhao et al. (2024), through feasibility analysis of a mobile learning support system for smartphones, proposed a multi-level music composition teaching model based on mobile internet. Gong and Wang (2023) point out that building a synchronous interactive music composition teaching system that meets students' growth and individual needs can promote learners' acquisition of musical knowledge and deepen their understanding. Duarte-García et al. (2020), under the support of mobile environments, used relevant software to build a music composition teaching model based on mobile whiteboards, thus enhancing classroom interactivity. Ye and Zhang (2023) point out that in interactive learning supported by mobile devices, the quality of music composition resources provided through mobile devices is relatively low, and the format of resources is too monotonous, making it difficult to attract continuous engagement from learners during interactions. Uludag and Satir (2025) found in music composition teaching case studies that when teachers use mobile technology for interactive learning in class, they often remain at the operational interaction level, which makes it difficult to achieve deep learning outcomes.

Although some achievements have been made in the research on music composition teaching systems based on mobile interactive technology, and there are more and more mobile learning products on the market, the quality of musical teaching resources varies greatly. Most students choose to learn in a casual manner without clear goals or systematic plans, which significantly affects their learning effectiveness (Shi, 2023; Zheng, 2024). To address these issues, researchers have studied music composition resource generation models based on artificial intelligence (AI) algorithms, with main foundational model architectures including long short-term memory (LSTM), transformer, and generative adversarial network (GAN) (Meng et al., 2025). In the generation of AI music creation teaching resources, the Transformer achieves the best overall performance, particularly excelling in multi-part, complex structures, and long-sequence compositions. LSTM is suitable for lightweight, rapid validation and small-scale data scenarios. GAN demonstrates advantages in timbre and stylistic diversity, but suffers from weaker training stability and structural controllability. Model performance improvements stem from the synergistic effects of architectural mechanisms, training strategies, data optimisation, and engineering refinements. Hewahi et al. (2019) trained an LSTM network with a large number of high-level musical vocabulary items to generate a series of music composition resources. Lei (2023) proposed a music composition resource generation model in light of deep convolutional neural networks (CNN). This model can learn composers' musical styles and dynamic features from training sets, achieving good generation results. Ferreira et al. (2023) applied the Transformer to music generation, leveraging its attention mechanism to achieve excellent results in generating long musical sequences with coherence. Min et al. (2022) proposed a model applying GAN to the field of music generation, where both the generator and discriminator use LSTM networks. The generator is trained to transform random noise into a melody. Cheng and Qu (2025) optimised the traditional GAN, first generating the main melody and chords of music, then extracting features from the generated parts, and finally generating multi-track music composition resources based on the extracted features. Although GANs have many advantages in music composition resource generation tasks, they still face issues such as unstable training, difficult model convergence, excessively random generation results, and poor auditory experiences (Jiang and Mou, 2024).

Currently, music composition teaching systems based on mobile interaction technology face several core challenges. First, teaching resources are unevenly distributed. High-quality music composition instructors and hardware equipment are often concentrated in specialised institutions or developed regions, making it difficult to provide widespread access to the general public. Second, delayed feedback and incentives. Learners struggle to receive immediate, effective positive feedback during the creative process. The solitary practice environment lacks interactivity, making it difficult to sustain motivation. Although numerous music creation apps (such as GarageBand and FL Studio Mobile) are available on the market, their core remains rooted in traditional digital audio workstation models, lacking creative interaction. Therefore, how to integrate generative AI's music generation capabilities with mobile interaction technology to form a closed-loop creation-learning-feedback system has become a key issue worthy of in-depth research.

To cope with these challenges, this article suggests a generative music composition teaching system based on mobile interactive technology. First, regarding the issue of excessive randomness and poor quality of generated music in resource generation, this study designs a music generation model based on a multi-scale screening module and multiple discriminators. The generator features a multi-scale feature screening module, with the objective of identifying and extracting the most critical features of authentic music. At the same time, this model introduces a multi-discriminator structure to enhance the discriminator's ability to identify music samples, thereby optimising the training process of the generator. Secondly, in terms of interaction model construction, an interactive model supporting mobile learning is built based on students, teachers, and music composition resources. Based on this, a music composition teaching system is designed. The system adopts the concept of separating control from transmission, decoupling network communication and interaction mechanisms in mobile interactions. It designs a scalable network communication module and an interactive collaboration mechanism supporting bidirectional collaborative control and user state management, reducing network pressure and improving interactivity among users.

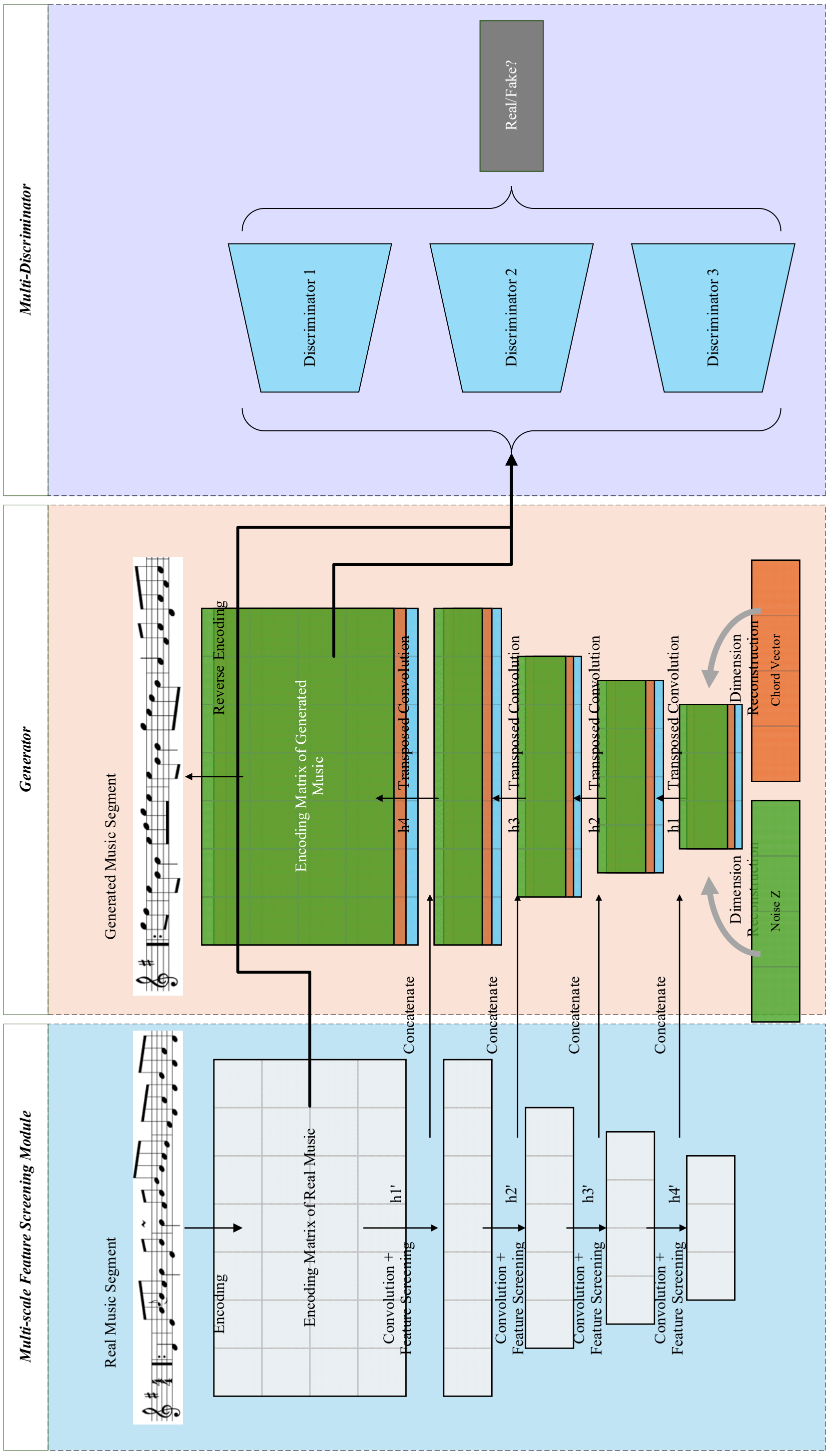
2 Generation of music composition teaching resources based on multi-scale feature selection

2.1 Overall framework of music composition teaching resource generation

Music composition teaching resource generation and generative music composition teaching systems are closely dynamically related, with the former providing core materials and sources of inspiration for the latter, while the latter enhances the application efficiency of the former through technical integration and optimisation of the teaching framework. Both jointly promote the transition of music education from knowledge transmission to creativity cultivation. In current tasks of generating music composition teaching resources, the initial goal is to generate music that better aligns with contemporary auditory aesthetics, thereby assisting musicians in their music composition teaching (Schmidt-Jones, 2018). However, due to traditional GAN's relatively random generation results, the generated music often has poor quality and lacks good listening experience.

To enable the model to find a reasonable balance during the generation process, ensuring that the generated music has sufficient richness without becoming chaotic. This chapter proposes a music generation model based on a multi-scale characteristic screening module and multiple discriminators (MSFC-MD). The structure of the MSFC-MD model is shown in Figure 1. The multi-scale characteristic screening module filters characteristics from real music samples at different scales, thereby selecting part of significant guiding melodic features to guide the generator in music generation. The structure of multiple discriminators structure introduces three distinct discriminators to extract symbolic music data features at different scales for discrimination, thus improving the discriminators' feature extraction capability. The ultimate goal is for the generator to output music that is perceptually indistinguishable from genuine human-composed pieces.

Figure 1 The structure of the music composition teaching resource generation model (see online version for colours)



2.2 Music feature screening network

Since the music feature matrix contains its inherent musical logic, to enable the model to extract certain effective information from real music samples to guide the generator in producing higher-quality music samples, this chapter introduces a multi-scale feature screening module. In this module, each scale's feature matrix uses a feature screening network to extract effective musical information. The extraction process is as follows: First, the feature matrix *X* passes through the feature screening module to obtain the feature matrix *X*', which is then combined with the original feature matrix *X* using weighted fusion, ultimately obtaining the screened feature matrix *Y*. In the feature screening module, this paper primarily draws on the feature recalibration strategy from squeeze-and-excitation network (SENet) (Pereira et al., 2019). Taking the feature matrix in $C \times H \times W$ as an example, for a feature matrix *X* obtained through a convolutional layer, it is first compressed via an average pooling operation. Each two-dimensional feature channel is converted into a real number. The output dimension equals the input feature channel count, resulting in an output vector *z* that represents the global receptive field. Its equation is as follows.

$$z_c = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} X_c(i, j) \tag{1}$$

where *c* stands for the amount of channels, and *H* and *W* stand for the feature dimensions for each channel.

The output vector *z* is input into a two-level fully connected layer to produce weights for each feature channel s_c, ultimately obtaining the weight distribution vector *s*, as shown below, where δ is the ReLU function, σ is the sigmoid function, W_1 is the first fully connected layer that compresses c-dimensional features to capture relationships between channels, and W_2 is the second fully linked level used for dimensionality restoration, resulting in the weight distribution vector *s*.

$$s = \sigma\left(W_2 \delta\left(W_1 z\right)\right) \tag{2}$$

Finally, *s* is applied channel-wise to the feature matrix *X* to obtain the weighted feature matrix *X*′, as shown below, where *c* represents the feature dimensions.

$$X_c' = s_c \cdot X_c \tag{3}$$

2.3 Multiple discriminator and loss function

Since musical information in the music feature matrix is concentrated near the time axis and distant parts from it are mostly blank, while complex intrinsic musical logic exists between densely packed features on the time axis, a single discriminator cannot extract sufficient effective information during feature extraction. This may lead to certain musically logical but monotonous samples being identified by the discriminator, which will affect the training of the generator and reduce the diversity of generated melodies. Therefore, this chapter introduces multiple discriminators into the music generation model to enhance its feature extraction capability. In this part, three discriminators are designed for feature extraction at different scales, and their discrimination results are weighted and fused. The network structure of each discriminator is similar, consisting of two convolutional layers and two fully connected layers. However, to achieve feature extraction at different scales, this chapter uses different parameter settings for the size of the convolution kernels and the amount of feature channels in the convolutional layers across different discriminators. In addition, to determine whether the melody of a musical sample is harmonious with the given chord, this chapter also concatenates the given chord vector with the feature matrix obtained from the discriminator multiple times along the channel dimension, which is denoted as *concat*.

For the loss function of the generator, its goal is to generate musical segments that can be identified by the discriminators as real music samples. Therefore, its loss function and optimisation objective are given in the following formula, in which θ_G stands for the network parameters of the generator, z is a random noise vector input into the generator, $p_z(z)$ is defined as the distribution of data that arises from the latent vector, $G(z)$ is the encoded matrix for music composition resources generated by the generator, and $D(G(z))$ is the output result of the discriminator based on the generator's created musical resource.

$$\min_{\theta_G} L_G = E_{z\sim p_z(z)}\left[\log\left(1-D\left(G(z)\right)\right)\right] \tag{4}$$

For the discriminators, their goal is to correctly distinguish real musical samples from music compositions generated by the generator. Therefore, during training, both real music composition resources and those produced by the generator are used simultaneously for training. Its loss function and optimisation objective are as follows: θ_D stands for the network parameters of the discriminator, x denotes the encoded matrix derived from a real musical sample, $p_{data}(x)$ refers to the distribution of the actual data, and $D(x)$ is the output result of the discriminator based on a real musical sample.

$$\max_{\theta_D} L_D = E_{x\sim p_{data}(x)}[\log(D(x))] + E_{z\sim p_z(z)}\left[\log\left(1-D\left(G(z)\right)\right)\right] \tag{5}$$

Finally, through the integration of the generator and discriminator losses, we can obtain the overall model's loss function. Its total loss function and optimisation objective are as bellow.

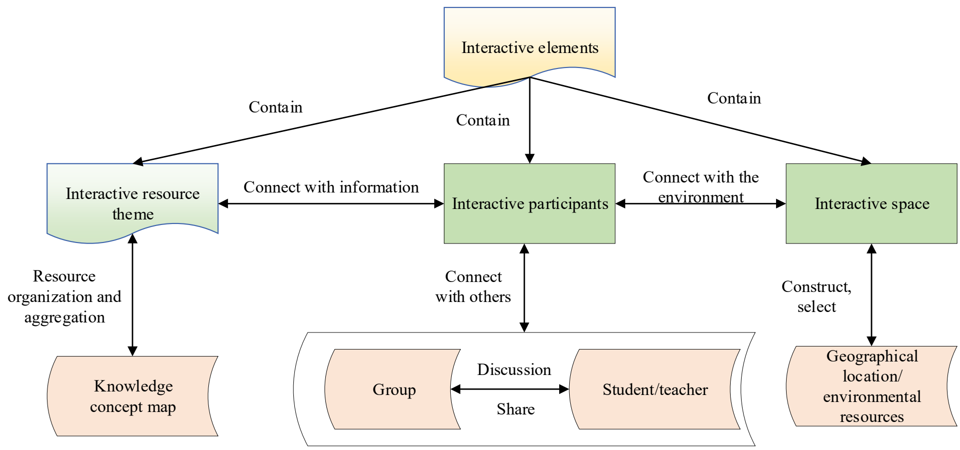
$$\min_{\theta_G} \max_{\theta_D} L(D,G) = E_{x\sim p_{dan}(x)}\left[\log\left(D(x)\right)\right] + E_{z\sim p_z(z)}\left[\log\left(1-D\left(G(z)\right)\right)\right] \tag{6}$$

3 Construction of a mobile interactive model integrating music composition teaching resources

3.1 Elements and relationships of the mobile interaction model

In the current mobile collaborative learning environment, learners have more interaction opportunities, richer and more flexible interaction methods, and broader interaction spaces compared to traditional teaching environments. This new teaching model should support more complex agent interaction relationships, meeting learners' needs for effective interactions. In connectivist (Alam, 2023) learning, learners' learning goals are the establishment of connections and the creation, continuous development, and optimisation of (cognitive, social, conceptual) networks. Therefore, combining with the music composition teaching resources generated in the previous section, we present a mobile learning interaction element model in Figure 2.

Figure 2 Mobile learning interaction elements model diagram (see online version for colours)



A complete interaction consists of three elements: the interactive resource topic, interactive participants, and the interactive space. Among these, the interactive participants are at the core, engaging in three types of connective interactions, including connectivity with teaching resources, connectivity with people, and connectivity with mobile devices.

1 Connectivity with music composition teaching resources. In connecting with teaching resources, learners organise and aggregate fragmented knowledge concepts, based on which they consolidate the content into conceptual knowledge maps. The ultimate goal of this process is to allow key nodes in the knowledge concept map to be aggregated as interactive topics, thereby connecting more content and learners. At the same time, learners find interaction topics through searching, looking up, or adding, and features such as labelling, recommendation, and association can also help learners quickly locate relevant interactive themes.

2 Connectivity with people. Interaction as a social learning process includes information sharing and discussion among interacting participants. In the interaction process, if learners have stronger group presence, both students and teachers show a greater willingness to participate actively and take initiative in collaborative group work.

3 Connectivity with mobile devices. In mobile learning, learners are no longer confined to computer screens but interact freely through mobile terminals within real-world spatial environments. Moreover, as the context-aware technologies and mobile techniques rapidly growing, learners take geographic information system services as a basis for location data and can independently construct or choose more personalised environmental resources via context-aware devices to form interactive spaces. In this interactive space, interaction participants are able to spark ideas through face-to-face communication and control the process of interaction in real-time, accurately, and dynamically. At the same time, mobile devices continuously receive and push expressions of physical environment information, automatically searching for interaction participants that meet specific conditions

through various information sources, ultimately making the form of interaction richer and more flexible.

3.2 Heterogeneous interaction model supporting mobile learning

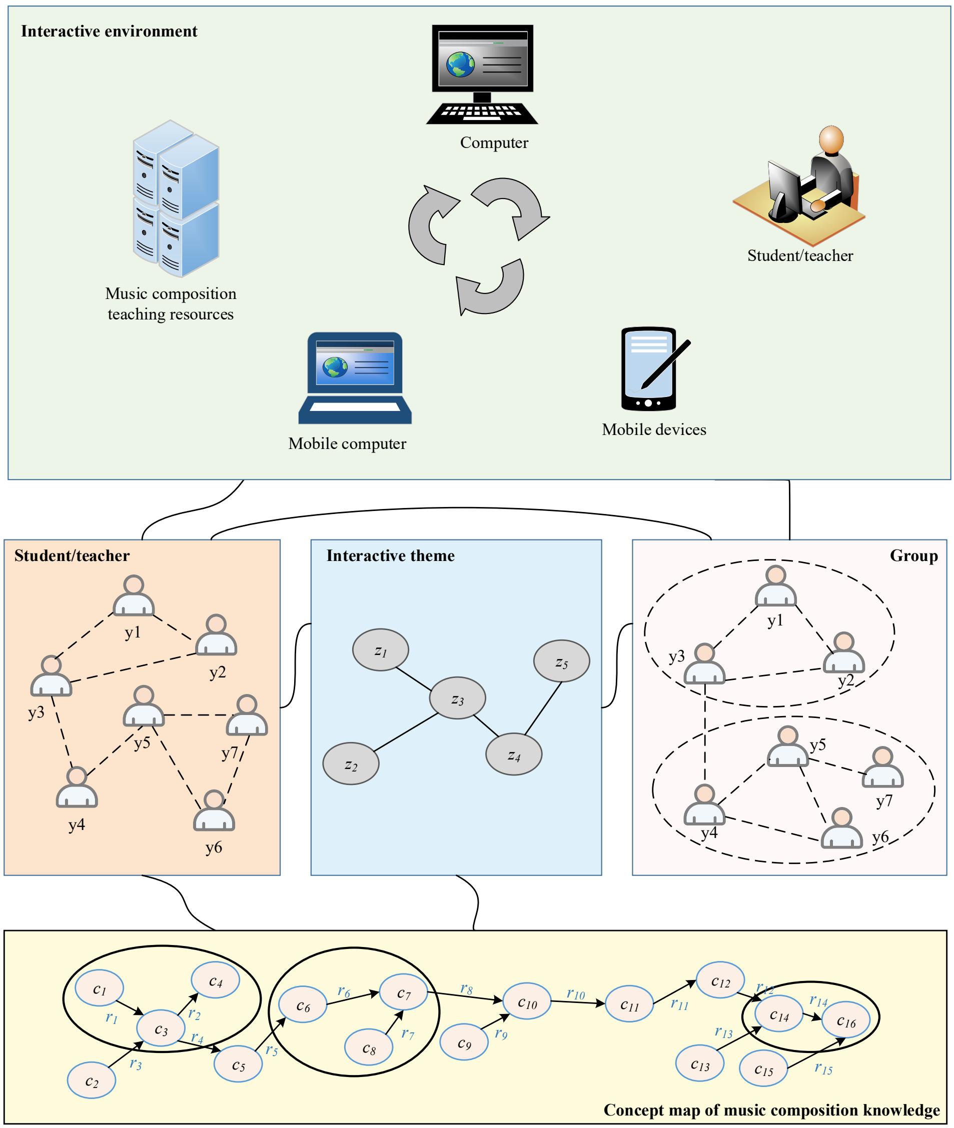
Based on the interactive structural characteristics of mobile learning networks, a heterogeneous undirected graph is used to establish an interaction model. The interactive elements included in the model are learners/teachers, groups, interaction topics, mobile devices, and knowledge concepts as graph nodes. Node connections are established according to the relationships between these interactive elements. Figure 3 displays the schematic diagram of the model. Let a heterogeneous graph be represented by $G = \{V, E, R\}$, where V is the node set of graph G, containing all objects involved in the interaction activities and expressed as: the learner/teacher node set being $Y = \{y_1, y_2,\ldots,y_{|Y|}\}$, the musical composition resource topic node set for interactions being $Z = \{z_1, z_2,\ldots,z_{|Z|}\}$, the group node set being $U = \{u_1, u_2,\ldots,u_{|U|}\}$, the environment node set being $S = \{s_1, s_2,\ldots,s_{|S|}\}$, and the knowledge concept node set being $C = \{c_1, c_2,\ldots,c_{|C|}\}$. E is the edge set of graph G, containing all interaction relationships between these objects. Then R represents a relation type set among the objects. There are six interaction relationship types in total according to the interactive activity's objects, represented as $R = \{(Y, U), (Y, Z), (U, Z), (Y, S), (U, S), (Z, C)\}$. The process for constructing interaction nodes and edges in this heterogeneous graph is as follows.

1 Construction of interaction nodes. Learner/teacher objects Y, group objects U, and knowledge concept graph C are directly added to G. For the interaction environment S in mobile learning, both the positional distance expressing the physical proximity between environment nodes and attribute features expressing environmental resource similarity need to be considered. Therefore, all locations in the interaction model should undergo clustering processing; during this clustering process, road network distances of locations and resource tags are used separately as spatial and non-spatial distances for weighted calculation to obtain a weighted environment node distance. Then, combined with k-means method for clustering, the results of this clustering will be added as spatial node set S into G. In addition, although online interaction does not have an actual physical interaction environment, it is also included in the interaction model as a form of interaction; therefore, the online environment type marked as s^* is added to S.

2 Interaction edge construction. Learner/teacher and group (Y, U), representing the affiliation relationship between learners/teachers and groups. Set edge weight $\omega(y_i, u_j)$ to measure the affiliation degrees of these two types of interaction nodes. For m users and n groups, their user set Y and U can form a matrix A_{Y_U}, where any element $a_{(Y\_U)ij}$ indicates that learner/teacher y_i is affiliated with group u_j. If y_i has not joined u_j, then $a_{(Y\_U)ij} = 0$. Generally, the more times learners/teachers interact with other members within group activities, the higher their membership degree to that interaction group node. Therefore, the edge weight between these two types of nodes is represented by the ratio of the amount of interactions from the learner/teacher node to the total member interactions, as shown in equation (7), where $Mes(y_i, u_j)$ represents the total number of interactions made by member y_i within group u_j during all interaction activities. A single teaching resource publication, participation in a discussion, or question reply can be regarded as one interaction.

$$\omega\left(y_i,u_i\right)=\frac{Mes\left(y_i,u_j\right)}{\sum_{a_{(Y\_U)kj}\neq 0}Mes\left(y_k,u_j\right)} \tag{7}$$

Figure 3 The proposed heterogeneous interaction model supporting mobile learning (see online version for colours)



Learner/teacher and interaction topic (Y, Z), representing the interest relationship between learners/teachers and interaction topics. If y_i participated in an interaction activity with an interaction topic of z_j, for m users and n topics, their user set Y and topic set Z can form an interest-topic matrix A_{Y_Z}. Any element $a_{(Y\_Z)ij}$ within this matrix indicates the number of times y_i participated in interactions with a topic of z_j. If y_i did not participate, then

$a_{(Y\_Z)ij} = 0$. Set edge weight $\omega(y_i, z_j)$ to measure the interest degrees between these two types of interaction nodes. The more times learners/teachers participate in an interaction on a certain topic, the higher their interest degree towards that topic node. For each user y_i, its corresponding interest-topic vector is as follows.

$$a^T_{(Y\_Z)i} = \left(a_{(Y\_Z)i1}, a_{(Y\_Z)i1}, \ldots, a_{(Y\_Z)\text{in}}\right) \tag{8}$$

Therefore, define the interest correlation degree as the ratio between the number of interactions for a specific topic by the learner/teacher node and the total number of interactions they participated in. This is used to represent the edge weight between these two types of nodes, as shown below.

$$\omega\left(y_i, z_j\right) = \frac{a_{(Y\_Z)ij}}{\left\| aa^T_{(Y\_Z)i} \right\|_1} \tag{9}$$

Group and interaction topic (U, Z), representing the interest relationship between groups and interaction topics. If group u_i initiated or participated in an interaction activity with an interaction topic of z_j, for group u_i and topic set Z, the group's interest-topic vector is as follows.

$$a^T_{(U\_Z)i} = \left(a_{(U\_Z)i1}, a_{(U\_Z)i1}, \ldots, a_{(U\_Z)in}\right) \tag{10}$$

The elements $a(U\_Z)_{ij}$ in this vector indicate the number of members within u_i who participated in interactions with a topic of z_j. If u_i did not initiate or participate in an interaction activity with a topic of z_j, then the interest-topic vector is a zero vector. The more members in a group participate in interactions on a certain topic, the higher their interest degree towards that topic node. Therefore, these two types of nodes' edge weights are defined as follows, where $SME(u_i)$ represents the total number of members within u_i.

$$\omega\left(u_i, z_j\right) = \frac{a_{(U\_Z)ij}}{SME\left(u_i\right)} \tag{11}$$

Learner/teacher and interaction space (Y, S), representing the positional relationship between learners/teachers and interaction spaces. For offline interactions, if y_i's check-in location on a mobile device is within s_j, then edge weight $\omega(y_i, s_j) = 1$. Otherwise, $\omega(y_i, s_j) = 0$. For online interactions, use $\omega(y_i, s^*)$ to represent the preference of y_i for online interaction. It is defined as the ratio among the amount of times y_i participates in online interactions and their total number of interactions. If this user has never participated in any online interaction, then $\omega(y_i, s^*) = 0$.

Group and interaction space (U, S), representing the positional relationship between groups and interaction spaces. The edge weight is defined by calculating the ratio of the group's total check-ins at location to its total membership size, as shown in equation (12). For online interactions, the calculation method for weights $\omega(u_i, s^*)$ remains the same and will not be repeated here.

$$\omega\left(u_i, s_j\right) = \frac{\sum_{\forall\, yk \in u_i} \omega\left(y_k, s_j\right)}{SME\left(u_i\right)} \tag{12}$$

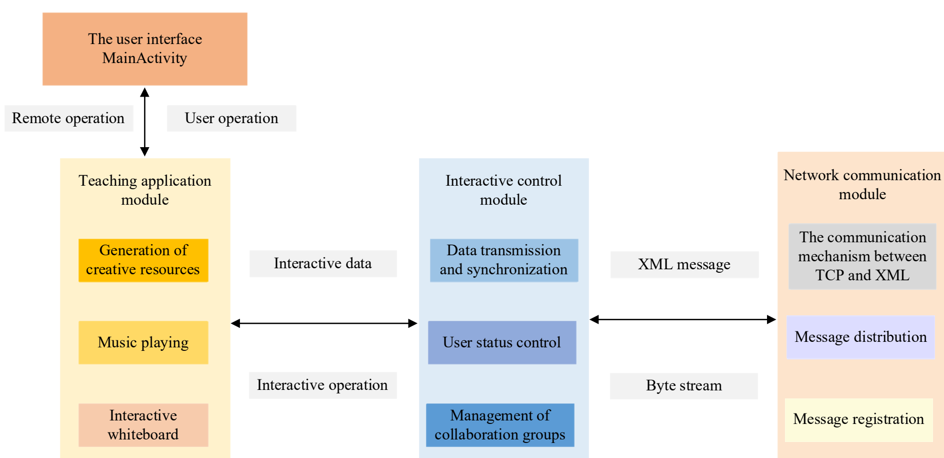
Interaction topics and knowledge concept graphs (Z, C), representing the inclusion relationship between interaction topics and knowledge concept graphs. This paper adopts the Novak concept map model, where learners and groups assist in maintaining the completeness attribute of knowledge within the concept map. If an interaction topic z_a includes a branch node c_b from the concept knowledge graph, then the edge weight is $w(z_a, c_b) = 1$. Otherwise, it is $w(z_a, c_b) = 0$.

4 Design of the music composition teaching system based on the mobile interaction model

4.1 System function description

Based on constructing a mobile interaction model integrating music composition teaching resources, this paper designs a music composition teaching system. The system adopts the idea of separating control and transmission, decoupling network communication from the interaction mechanism during interactions, and designing a network communication module supporting extensible eXtensible Markup Language (XML) messages as well as an interaction mechanism for two-way control and user state management. The system mainly solves the problem of bidirectional multiplexed synchronisation in interactive data within music composition teaching. The mobile end of the music composition teaching system studied in this paper adopts an operation-based interaction mechanism, avoiding the use of video recordings for instruction, thus reducing network pressure and improving interactivity among users.

Figure 4 Primary application scenarios of the system (see online version for colours)



The primary role of the interactive online teaching system is to simulate a real-world classroom, creating a virtual learning environment for teachers and students. The focus of this study is to investigate an interactive online teaching system mobile end that supports interaction, enabling students to access classes anytime and anywhere. The primary application scenarios of the system are shown in Figure 4, which includes a music composition teaching resources centre, teachers, and various teaching terminals.

Integrated within this framework, the mobile component of the interactive teaching system enables students to access virtual classrooms via Android devices, thus meeting the needs of learning anytime and anywhere. To effectively access virtual classrooms, the mobile system must offer students two core capabilities: a wealth of learning materials and robust tools for interaction, collaboration, and communication with their peers (Li et al., 2019).

1 Generation of music composition teaching resources. In the music generation module, this chapter divides the module into two functions, where the first function is weight selection. Through training, this paper will provide the MSFC-MD model with more than two trained models and weights that generate better results. Users can select the corresponding weights according to their preferences for music generation. The second function is music generation. In terms of music generation, the system provides a model channel representing the music generation model MSFC-MD. After selecting the model weights, users only need to click the generate musical instrument digital interface (MIDI) button; the system will pop up a window and let users choose the save path for the music file. Once the save path is confirmed, the system can generate the corresponding MIDI music. However, if users forget to select the weight path, the system will display a window requesting that the user select the weights first.

2 Playback of music composition teaching resources. The system needs to provide participants with instructional tools for accessing music composition teaching resources; playback tools are used to meet this need. The playback tool can synchronise the playing of teaching resources between teachers and students and supports bidirectional control by both sides, enhancing the interactivity of course materials.

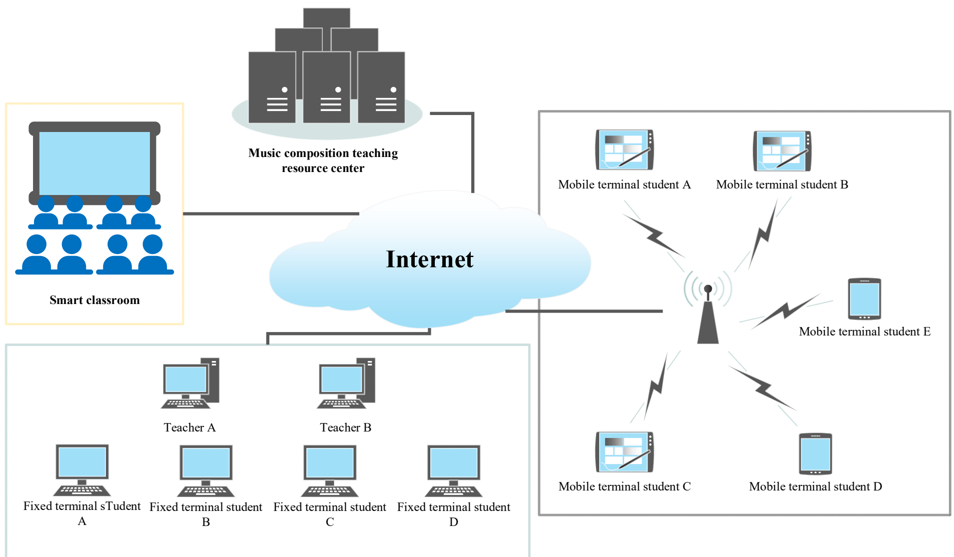
3 Interactive electronic whiteboard sharing. Provide an interactive shared electronic whiteboard for both parties in teaching as a replacement for traditional blackboards. The interactive whiteboard tool can synchronise content created on the whiteboard in real-time between teachers and students, supports bidirectional synchronisation by both sides of instruction, provides control capabilities to students, and enhances interactivity during teaching.

4 Communication connection and login: It has the function of establishing network communication, enabling the mobile student terminal to connect with the teacher's terminal and providing communication support for other teaching interaction-related modules of the system.

5 Interactive and collaborative music composition teaching: In order to make music composition teaching interactive, it is necessary to provide operational synchronisation and data processing capabilities. This function allows participants in music composition teaching to synchronise their operations on teaching resources with those of other participants according to teaching strategies, and enables the processing of signalling data for interaction to ensure real-time teaching interaction. During class, teachers are responsible for managing teaching resources and have control over them. In addition, students can apply to the teacher for control rights of the teaching tools, allowing their operations on these tools to be synchronised with those on other participants' systems so that they can participate in the teaching.

The interactive instruction system is architected with a modular design and built upon the Android operating system. The system is principally partitioned into three functional modules: namely, the teaching application module, the interaction control module, and the network communication module. The interaction control and network communication modules provide services to the teaching application module, enabling user interaction and collaboration during instruction. The teaching application module is presented as custom View controls overlaying the MainActivity and engages in direct interaction with users. Figure 5 presents the system architecture.

Figure 5 Overall structure of a generative music composition teaching system based on mobile interaction technology (see online version for colours)



4.2 Design of music composition teaching application module

The module's main objective is to serve as an interactive platform for the delivery of instruction, supporting both educators and learners. To this end, the teaching application module must be equipped with a suite of diverse tools to facilitate multidimensional user interaction. To allow each teaching application to function autonomously and be rendered in various activity regions, it is essential to decouple the views of the teaching applications from their particular functionalities.

The Android application framework provides the foundational view controls used to design the teaching application tools. Within the Android framework, view controls represent the fundamental, native building blocks for user interfaces in Android, they define a rectangular screen region responsible for both drawing its content and handling user events within its bounds. View controls provide an onDraw () method for redrawing regions and an invalidate () method for updating them. The event response process is categorised into local events and remote events. Local events primarily refer to user operations; for example, a user performing a page-turn operation on a tool used to play music composition teaching resources. In this case, after the event listener captures the user's touch interaction with the View control, it notifies the controller of the teaching resource tool, and the controller completes the page turn for the teaching resource playback tool, updating the view. At the same time, the event listener can also notify the interaction control module about this operation to complete the interactive function. The interaction control module may also inform the controller of remote operations and request synchronisation locally.

4.3 Network communication module design

Access to the online virtual classroom requires users to employ mobile terminals as computational nodes. These devices must connect via the system's designated network model to facilitate internet access and user communication, receive guidance from remote teachers, and collaborate with students in different locations for learning.

The network communication module employs transmission control protocol (TCP) for connections and utilises XML for message formatting. To optimise messaging efficiency, the module employs a single-threaded Selector pattern instead of spawning a dedicated thread per connection, managing all socket I/O through a central Select mechanism. Android supports non-blocking network communication through its ServerSocketChannel and Selector classes.

The network module begins by registering a channel with a Selector. After initiating the listening socket, it enters a loop to continuously call the select () method. Whenever a new SelectionKey emerges, it could originate from either an existing connection that now has newly readable data or from the establishment of a brand-new connection. In such cases, judgments and processing should be based on events of the SelectionKey. The Select call blocks indefinitely until either a new SelectionKey is ready or an explicit wake-up signal is received.

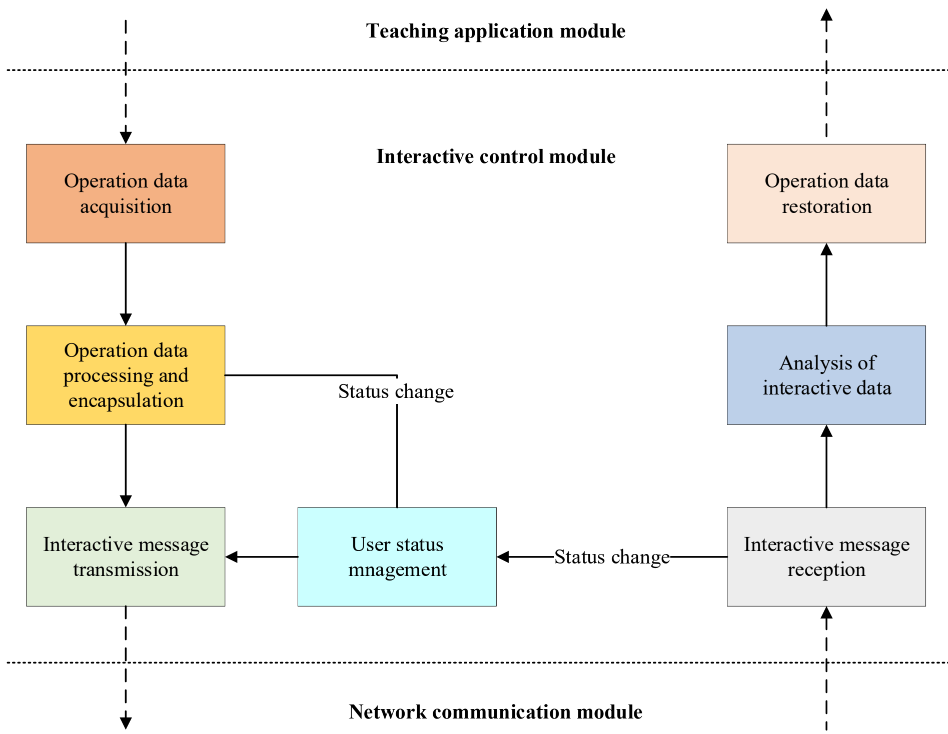
After the interactive teaching system is launched, when a student wishes to join a teacher's class, they must request a connection from the teacher server. If the teacher server is already running and functioning normally, it will accept the student's connection request and add that student to the current student list. If the initial connection attempt fails, the system will retry several times. Should all attempts remain unsuccessful, the system will notify the user.

During the teaching process, either the teacher or the student may terminate the communication. For the mobile version of the interactive teaching system, if the teacher properly ends the class, the mobile app will disconnect normally upon receiving the teacher's disconnection message. If a teacher disconnects unexpectedly, the interactive teaching system will terminate the current communication session upon detecting the disconnection and attempt to reconnect. Should multiple reconnection attempts fail, the system will generate a report for the user. When the mobile device of the interactive teaching system initiates a connection termination, it sends a disconnection notification to the teacher server. Upon receiving the teacher server's acknowledgment, the mobile device actively disconnects. The teacher server then removes the student from the current student list.

4.4 Interactive control module design

Interactivity is the biggest difference between the music composition teaching system and traditional asynchronous video teaching, as it enables real-time interaction between teachers and students, thereby improving teaching quality. These interactive functions are built upon a reasonable network transmission and synchronisation control strategy of the system. Network transmission serves as the carrier for interactive data, responsible for transmitting user interactions and states; meanwhile, the synchronisation control strategy defines the specifications for interactive data, describing the included user operations in the interactive data, as well as establishing rules for users' interactive behaviour.

Figure 6 Data flow direction of the interactive control module (see online version for colours)



The interactive control module is responsible on one hand for receiving operation data from the teaching application module, processing and encapsulating it, then transmitting through the network communication module. On the other hand, it needs to receive and parse data from the network communication module, reconstructing remote operations locally in the teaching application, as shown in Figure 6.

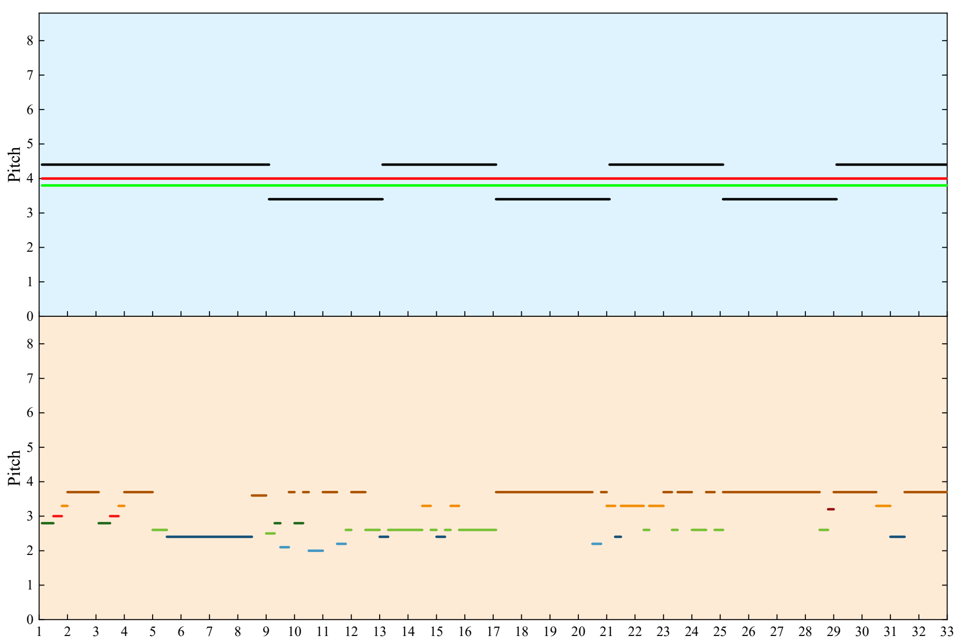
5 Experimental results and performance analysis

5.1 Analysis of music creation resource generation effects

The system architecture employs the Android OS for mobile client devices, while the instructor's interface operates on the Windows platform. Evaluation was performed using a plurality of Android mobile endpoints. The experimental dataset comes from MIDI files of 1022 popular songs in TheoryTab (Huo, 2024). Each song provides two MIDI files: one file contains the melody of the song, and the other contains the chords in the song, mainly used to assist the model in generating music based on chord characteristics. An NVIDIA GeForce GTX 1060 (6GB) graphics card was used for model training. In the training process, both the generator and discriminator use Adam optimiser to optimise network parameters, with a learning rate set at 0.0002. The weighted coefficient in the multi-scale feature screening module is 0.01, and the weighted coefficient of the multi-discriminator structure is 0.1. In addition, the batch size was set to 72, with a total of 20 training rounds conducted.

From Figure 7, it can be seen that the music generated by the model proposed in this chapter demonstrates good continuity, with smooth transitions between notes and certain variations in time values between each note, all within reasonable ranges. From the melodic perspective, its generated melody aligns with the chords on the same natural scale, resulting in a harmonious combination of chords and melody in terms of auditory perception. All are in the C major natural scale, indicating that the model proposed in this chapter can learn the scale patterns in the dataset, and its generated samples comply with music theory standards.

Figure 7 Sample generation results for music composition instruction (see online version for colours)



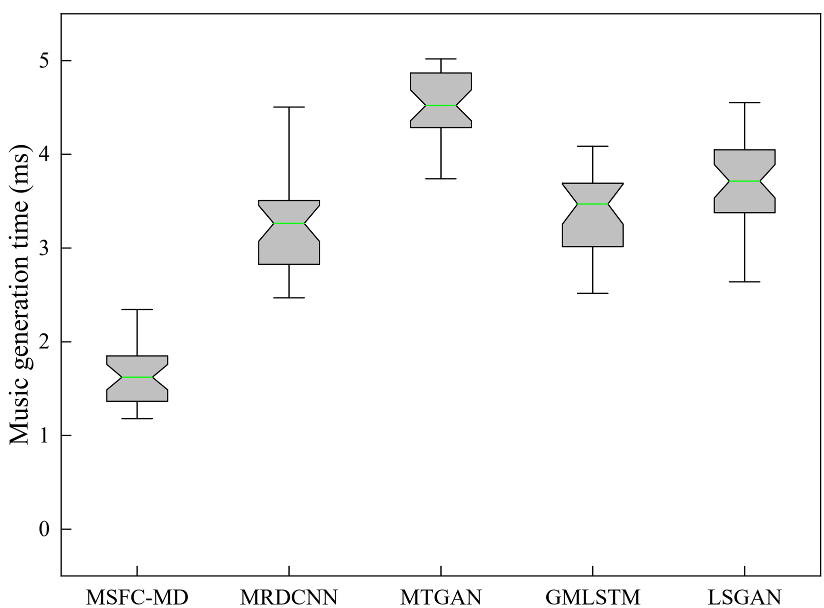
To demonstrate the superiority of the MSFC-MD model in generating music creation teaching resources, this chapter conducts comparative experiments using objective evaluation indicators such as note diversity (ND), average note duration (AND), centricity (CE), and maximum pitch difference (MPD). The models involved in the comparison are GMLSTM (Hewahi et al., 2019), MRDCNN (Lei, 2023), LSGAN (Min et al., 2022), and MTGAN (Cheng and Qu, 2025). To ensure fairness, this chapter still uses randomly generated 80 bars of music by each model to evaluate the objective indicators. The final experimental results are shown in Table 1. To reflect the difference among the music samples generated by each model and actual music samples, the table also provides the absolute differences (in blue font) between the generated music samples and real music samples for each indicator.

Table 1 Objective quality metrics for music composition teaching resources

Real music	*ND*	*AND*	*CE*	*MPD*
	11.1	*0.990*	*23.20%*	*17.4*
MSFC-MD	10.4	1.095	23.22%	17.6
	0.7	0.019	0.02%	0.2
GMLSTM	9.7	1.009	25.75%	20.2
	1.2	0.105	2.55%	2.8
MRDCNN	9.3	0.837	27.47%	21.5
	1.5	0.153	4.27%	4.1
LSGAN	9.0	0.905	26.25%	18.9
	2.1	0.085	3.05%	1.5
MTGAN	15.4	1.000	21.98%	42.7
	4.3	0.137	1.22%	25.3

From the experimental results, it can be seen that MSFC-MD achieves the level closest to real music samples in all indicators, with absolute differences of 0.7, 0.019, 0.02%, and 0.2, respectively. This also shows that based on objective evaluation indicators, after introducing the multi-scale feature screening module and multi-discriminator structure, the model proposed in this chapter has improved its ability to extract music features to some extent, generating music samples closer to real music and demonstrating certain advantages in music generation quality.

Figure 8 Music composition teaching sample generation time consumption of each model (see online version for colours)



To further validate the real-time distribution of music generation resources across different models, this paper statistically analysed the music composition teaching sample generation time consumption of each model, as shown in Figure 8. The generation time consumption for MSFC-MD was 1.8 ms, while the evaluation time consumption for GMLSTM, MRDCNN, LSGAN, and MTGAN was 4.5 ms, 3.8 ms, 3.6 ms, and 3.3 ms, respectively. Compared to the baseline model, MSFC-MD exhibits smaller median, upper quartile, and lower quartile values, fewer outliers, and a more concentrated overall distribution. Furthermore, the gap in the MSFC-MD box plot does not overlap with the corresponding intervals of the other four models, indicating a significantly smaller median, shorter generation time, and markedly superior real-time performance compared to the other models.

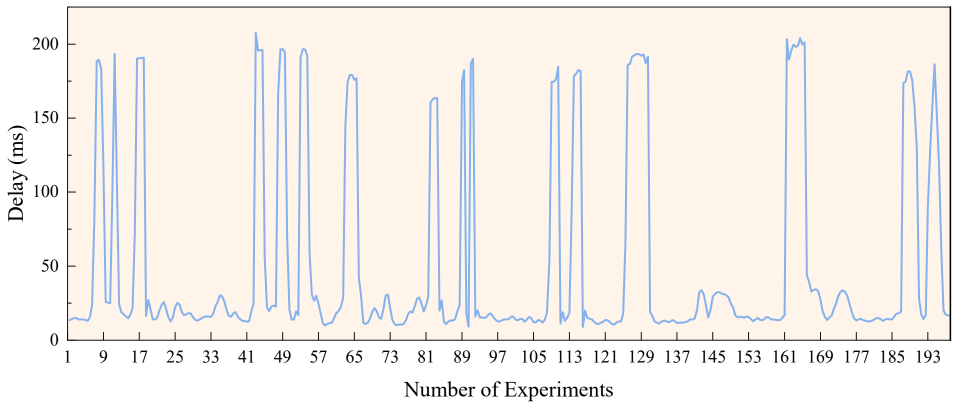
5.2 Performance analysis of the music composition teaching system

The scope of the performance testing is to verify the operational effectiveness of the mobile application under the load of real-time interaction with the teacher's station. Low-latency data response is essential for a seamless interactive experience; therefore, it demands special consideration. Testing is conducted in two scenarios: before and after optimisation, primarily evaluating the round-trip delay of interaction data between teacher and student mobile devices. Given the high frequency of interactions, the testing focuses primarily on the interactive whiteboard tool. Compared to the synchronisation of music composition teaching resources playback, the interactive whiteboard involves a larger amount of synchronised data, which better reflects real-world situations.

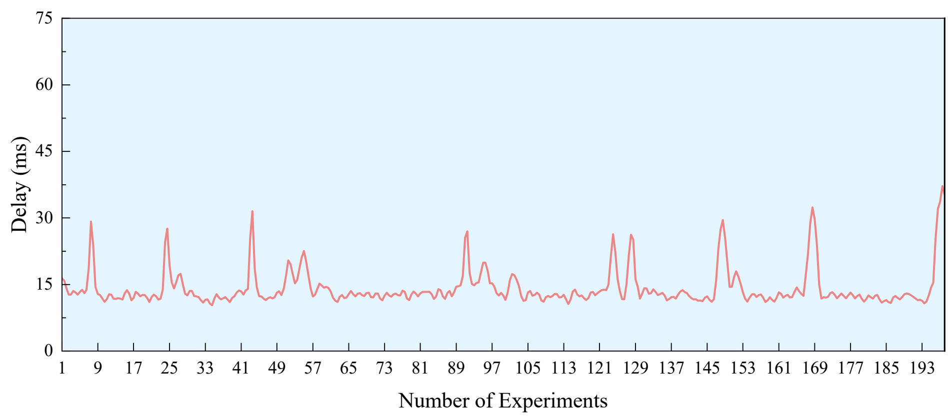
As shown in Figure 9, the initial evaluation compares the interaction data round-trip time between teacher and student ends with the Nagle algorithm enabled and disabled. Wireshark is used to capture and analyse network traffic originating from the mobile end of the teaching system. After Student A obtains permission, Student A's synchronisation data is first sent by A to the teacher's end and then transmitted from the teacher to other students. At the same time, when Student C joins the class midway, he or she first confirms the current classroom status with the teacher's end before performing state synchronisation, which aligns with the design. By comparing Figure 9(a) and Figure 9(b), before transmission optimisation, the average round-trip time interval for students to receive acknowledgments after sending interaction messages is 48.6275ms, and the round-trip time interval is unstable, with occasional longer intervals of around 200ms. After optimisation by disabling the Nagle algorithm, the average round-trip time interval is significantly reduced to 14.2195ms, with a more stable interval mostly below 30ms. In actual interactive operations, especially during operation processes, users also felt a reduction in interaction delay, indicating significant optimisation effects.

For the use of the music composition teaching system, frequent user access and excessively high node hardware resource loads can increase CPU and memory usage. If hardware load is too high, it may affect data storage and scheduling performance at nodes. Under identical experimental settings, the CPU and RAM utilisation for ILEML (Gong and Wang, 2023), IMSME (Duarte-García et al., 2020), SRIMR (Ye and Zhang, 2023) and OURS systems are observed, as shown in Figures 10 and 11 respectively. The CPU and memory usage of OURS are 43% and 37%, respectively, which decreased by 41% and 38% compared to ILEML, by 13% and 8% compared to IMSME, and by 22% and 24% compared to SRIMR. Whether in terms of CPU usage or RAM usage, OURS performs well, can better utilise system performance, and is capable of handling computational tasks at some high-load nodes, improving the efficiency of system interaction.

Figure 9 Round-trip time for mobile interaction data transmission with and without Nagle algorithm enabled, (a) round-trip time for unoptimised interactive data transmission (b) optimised round-trip time for interactive data transmission (see online version for colours)



(a)



(b)

Figure 10 CPU usage across different systems (see online version for colours)

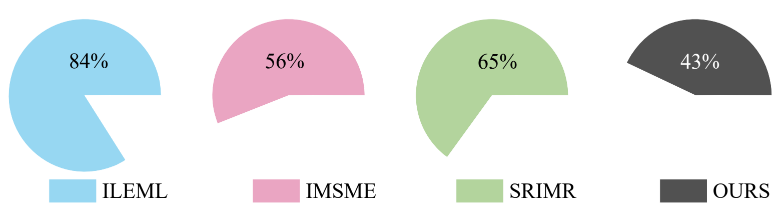
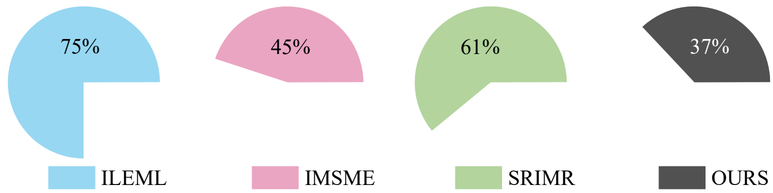


Figure 11 RAM usage across different systems (see online version for colours)



This system is not intended for individual users, so it must account for scenarios involving multiple concurrent users accessing the system. Stress testing is required to evaluate system performance under high concurrency conditions. Since the system has not yet been deployed for general use, software was employed to simulate the stress test. The system underwent stress testing with approximately 500 users. The results of the stress test are shown in Table 2. The system can generally handle concurrent access from approximately 500 users. Furthermore, when the number of concurrent users reached 550, the server response time was around 5 ms. This level of latency is essentially imperceptible to students and teachers.

Table 2 Music composition teaching system stress testing

Number of concurrent sessions	*User login response time*	*Interactive creation time*
550	5 ms	5 ms
500	4 ms	4 ms
450	3 ms	2.5 ms
400	1.8 ms	2 ms
300	1.5 ms	1 ms
200	1 ms	0.8 ms
100	0.5 ms	0.5 ms

Table 3 System interaction performance comparison

System	*ST/ms*	*SI/%*	*CS/%*
ILEML	23.2	86.4	89.6
IMSME	13.5	93.4	97.1
SRIMR	17.9	88.2	92.0
OURS	11.8	96.9	99.3

To further verify the efficiency of the proposed system, this paper selects quantitative evaluation metrics such as single interaction time (ST), interaction smoothness (SI), and user satisfaction (CS) to evaluate the performance of different music composition teaching systems. The results are shown in Table 3. OURS has a single interaction time of only 11.8ms, which is reduced by 11.4 ms, 1.7 ms, and 6.1 ms compared to ILEML, IMSME, and SRIMR respectively. The SI and CS of OURS are 96.9% and 99.3%, representing at least an increase of 3.5% and 2.2% over the other three models, indicating that the OURS system not only meets the requirements for real-time interaction but also significantly improves the smoothness and satisfaction of interaction, providing technical support to enhance the quality of music composition teaching.

6 Conclusions

As the mobile interaction technology and generative AI rapidly growing, music composition education is undergoing unprecedented changes. To address common issues such as low-quality creative resources and poor interactive real-time performance in music composition teaching systems, this paper proposes a generative music composition teaching system based on mobile interaction technology. Firstly, to solve the problem of low-quality music teaching resource generation, the system designs a multi-scale feature screening-based music generation method. A multi-scale characteristic screening module is introduced into the generator to guide it in learning key feature information from real music, while a multi-discriminator structure is also incorporated to enhance the discriminator's ability to distinguish musical samples, thereby improving generation quality. Secondly, this paper combines heterogeneous graph theory with interaction element analysis to build a new heterogeneous interaction model for mobile learning environments. Based on this, a mobile interaction-supported music teaching system architecture was further designed, which separates network communication from interactive collaboration mechanisms, achieving an extensible network communication module and an interactive cooperation mechanism with bidirectional collaboration control and user state management capabilities.

A large number of simulation experiments were conducted in real environments, and the outcome indicates that the proposed system can achieve high interactive real-time performance and improve the quality of music composition teaching resource generation. This research establishes a novel technical framework for music composition pedagogy while simultaneously providing a practical blueprint for the profound integration of mobile interactive technologies and generative AI within arts education.

Declarations

All authors declare that they have no conflicts of interest.

References

Alam, M.A. (2023) 'From teacher-centered to student-centered learning: the role of constructivism and connectivism in pedagogical transformation', *Journal of Education*, Vol. 11, No. 2, pp.154–167.

Cheng, D. and Qu, X. (2025) 'Intelligent generation and optimization of resources in music teaching reform based on artificial intelligence and deep learning', *Scientific Reports*, Vol. 15, No. 1, pp.30–42.

Cheng, L. (2025) 'The impact of generative AI on school music education: challenges and recommendations', *Arts Education Policy Review*, Vol. 13, No. 2, pp.1–8.

Dai, B., Wu, Y. and Niu, L. (2025) 'Enhancing student engagement and classroom interaction through mobile interactive technologies', *International Journal of Interactive Mobile Technologies*, Vol. 19, No. 15, pp.142–156.

Duarte-García, M.A., Wilde, E., Cortez, R. and Sigal, J.R. (2020) 'The use of an interactive music system as an aid for exploring sound in music education in a rural area', *Revista Música*, Vol. 20, No. 1, pp.357–380.

Ferreira, P., Limongi, R. and Fávero, L.P. (2023) 'Generating music with data: application of deep learning models for symbolic music composition', *Applied Sciences*, Vol. 13, No. 7, pp.45–53.

Gong, L. and Wang, J. (2023) 'Interactive learning environment for effective music learning in Chinese primary and secondary schools', *Sustainability*, Vol. 15, No. 3, pp.21–43.

Hernandez-Olivan, C. and Beltran, J.R. (2022) 'Music composition with deep learning: a review', *Advances in Speech and Music Technology: Computational Aspects and Applications*, Vol. 10, pp.25–50.

Hewahi, N., AlSaigal, S. and AlJanahi, S. (2019) 'Generation of music pieces using machine learning: long short-term memory neural networks approach', *Arab Journal of Basic and Applied Sciences*, Vol. 26, No. 1, pp.397–413.

Huo, X. (2024) 'Design of automatic generation model for teaching music based on user differences and transfer learning', *International Journal of Embedded Systems*, Vol. 17, Nos. 1–2, pp.62–72.

Jiang, R. and Mou, X. (2024) 'The analysis of multi-track music generation with deep learning models in music production process', *IEEE Access*, Vol. 12, pp.110322–110330.

Larsson, C. and Georgii-Hemming, E. (2019) 'Improvisation in general music education – a literature review', *British Journal of Music Education*, Vol. 36, No. 1, pp.49–67.

Lei, H. (2023) 'Reset and integration of music instructional resources using deep convolutional neural networks', *Journal of Environmental and Public Health*, Vol. 4, No. 2, pp.25–38.

Li, Y., Lin, Q., Qian, C. and Zhao, S. (2019) 'Research on evaluation model for interactive classroom enabled with mobile terminals', *Interactive Learning Environments*, Vol. 27, No. 2, pp.163–180.

Meng, Z., Zhu, L. and Chen, G. (2025) 'Innovative design of ceramic products based on extension reasoning and generative adversarial networks', *International Journal of Information and Communication Technology*, Vol. 26, No. 38, pp.75–90.

Min, J., Liu, Z., Wang, L., Li, D., Zhang, M. and Huang, Y. (2022) 'Music generation system for adversarial training based on deep learning', *Processes*, Vol. 10, No. 12, pp.25–37.

Pereira, S., Pinto, A., Amorim, J., Ribeiro, A., Alves, V. and Silva, C.A. (2019) 'Adaptive feature recombination and recalibration for semantic segmentation with fully convolutional networks', *IEEE Transactions on Medical Imaging*, Vol. 38, No. 12, pp.2914–2925.

Schmidt-Jones, C. (2018) 'Open online resources and visual representations of music: new affordances for music education', *Journal of Music, Technology & Education*, Vol. 11, No. 2, pp.197–211.

Shi, Y. (2023) 'The use of mobile internet platforms and applications in vocal training: synergy of technological and pedagogical solutions', *Interactive Learning Environments*, Vol. 31, No. 6, pp.3780–3791.

Sun, N. and Zang, Y. (2025) 'Innovative applications and teaching effectiveness analysis of interactive mobile technology in music education', *International Journal of Interactive Mobile Technologies*, Vol. 19, No. 1, pp.74–82.

Tong, J. (2016) 'Design and implementation of music teaching platform in college based on android mobile technology', *International Journal of Emerging Technologies in Learning (Online)*, Vol. 11, No. 5, pp.15–31.

Uludag, A.K. and Satir, U.K. (2025) 'Seeking alternatives in music education: the effects of mobile technologies on students' achievement in basic music theory', *International Journal of Music Education*, Vol. 43, No. 2, pp.172–188.

Ye, Y. and Zhang, S. (2023) 'Development of music teaching system by using speech recognition and intelligent mobile remote device', *International Journal of System Assurance Engineering and Management*, Vol. 7, pp.1–9.

Zhang, S. (2023) 'Interactive environment for music education: developing certain thinking skills in a mobile or static interactive environment', *Interactive Learning Environments*, Vol. 31, No. 10, pp.6856–6868.

Zhao, H., Narikbayeva, L. and Wu, Y. (2024) 'Interactive systems for music composition: a new paradigm in music education', *Interactive Learning Environments*, Vol. 32, No. 4, pp.1194–1203.

Zheng, Y. (2024) 'E-learning and speech dynamic recognition based on network transmission in music interactive teaching experience', *Entertainment Computing*, Vol. 50, pp.71–84.