A Continuous Hidden Markov Algorithm-Based Multimedia Melody Retrieval System for Music Education

Yingjie Cheng

Conservatory of Music, Hubei Engineering University, Xiaogan, Hubei 432000, China
E-mail: chengyingjie1001@163.com

Abstract

Education professionals receive instruction in Music Education (ME) to prepare for prospective jobs like secondary or primary music teachers, schools ensembles executives, or ensembles directors at music institutions. In the discipline of music education, educators do original research on different approaches to teaching and studying music. The most accurate and effective method of extracting music from huge music databases has become one of the most frequently discussed participants in contemporary multimedia information retrieval development. The essence of multimedia material is presented within a range of techniques since it is not bound to a single side. These several categories could consist of the song's audio components and lyrics for musical information. Retrieving melodic information, subsequently, becomes the main focus of most recent studies. Aside from being an expensive deviate from academics, music programs are neither a viable profession neither a valid pastime. Therefore, in this study, we offer a Continuous Hidden Markov Algorithm (CHMA) related a novel method for recovering

melodies from musical multimedia recordings. CHMA is considered to be the most basic dynamic Bayesian network. Two various types of audio frame features and audio example features are extracted throughout the feature extraction procedure from the audio signal according to unit length. Every music clip receives a unique approach that we implement with concurrently using various CHMA. The initial music gets processed using a trained CHMA that monitors fundamental frequencies, maps states, and generates retrieval outcomes. The training time for Traditional opera reached 455.76 minutes, the testing time for Narration achieved 56.10 minutes, and the recognition accuracy for advertisement reached an impressive 98.02%. A subsequently experimental result validates the applicability of the proposed approach.

Keywords: Music education, multimedia melody retrieval system (MMRS), audio features, continuous hidden Markov algorithm (CHMA).

1 Introduction

The Internet is becoming more populated with digital music resources due to the fast development of multimedia and digital technology. Enormous library resources encourage users to create a wide range of sophisticated music retrieval requirements. Users could, for instance, be eager to hear a specific type of music or a song with a particular mood at an exact time [1]. The development of multimedia technology has revolutionized based on teach and learn, especially in music education. With the increasing demand for high-quality education and the advancements in multimedia technology, it has become necessary to integrate multimedia tools and techniques into the teaching and learning process. Multimedia technology provides a platform for synchronous, integrated, and interactive learning experiences, which can significantly enhance the effectiveness and efficiency of music education. By incorporating multimedia elements such as audio, video, graphics, and text, music educators can create engaging and interactive learning materials that cater to the diverse learning needs of students. Overall, integrating multimedia technology in ME can enhance the quality of teaching and learning and promote the development of a more dynamic and interactive ME system [2]. Overall, ME based on computer music education, music coordinating practice, digital stacks, digital aesthetics, and music information retrieval is a rapidly evolving field which is shown in the Figure 1. The integration of digital aesthetics and computer music creation in the current music education scene has resulted in different processes. As music coordinating
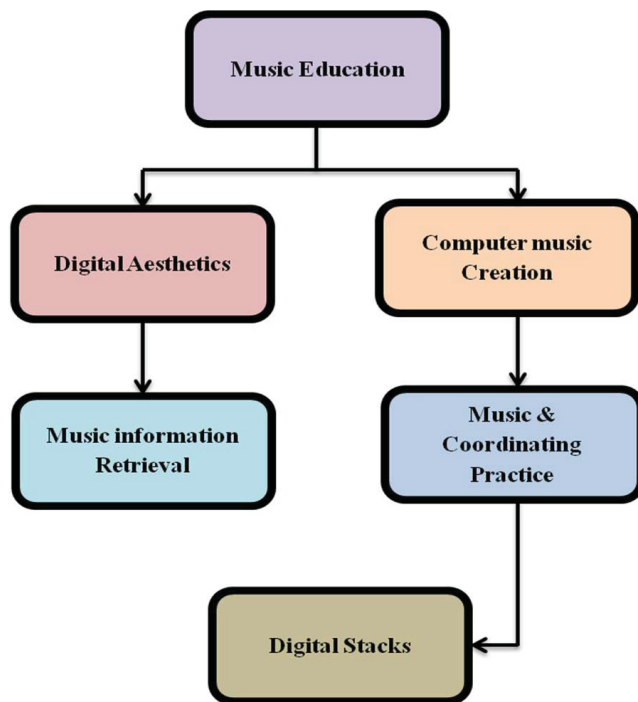
Figure 1   Structure of music education.

practice facilitates teamwork, music information retrieval gives musicians access to large musical knowledge repositories. The continuous interaction inside the digital stack improves music instruction and offers up new avenues for creative thought in the constantly changing field of music.

Early practical music through the contemporary development of music technology has been strongly linked with the development of music art. University-level ME mainly fosters the next generation of musicians and music professionals. The use of multimedia technology (MT) may improve ME by fostering an intuitive and engaging learning environment. Music educators can develop engaging and interactive learning materials that can assist students in developing sound scientific concepts [3]. The advantages of MT can assist students in creating their performing style, which is essential in the modern music business. The importance of MT and its potential benefits for music instruction should be better understood scientifically by college vocal music instructors. To develop creative and efficient teaching strategies that can boost the effectiveness and caliber of vocal music training, they should

research and use the advantages of MT. College vocal music instructors may assist students in reaching their full potential and prepare them for a lucrative future in the music business by figuring out the ideal tech-and-teaching blend [4]. Implementing MT into university-level ME can ultimately transform how the instructors teach and study music students and the music business. Finding distinct material sources from the original multimedia data of popular music is a considerable difficulty, the popular music multimedia signals are often organized chronologically. Popular music material must be clearly defined, making categorizing and retrieving detailed information challenging [5]. The current approaches for classifying popular music multimedia are relatively straightforward and cannot adequately reflect the characteristics of popular music. The music may make it challenging to employ intelligence teaching strategies for popular music multimedia. Researchers must investigate more complex and precise categorization techniques for popular music multimedia to address the difficulties. To do, sophisticated methods like machine learning, data mining, and deep learning may be used to pinpoint popular music's distinctive qualities and traits [6]. The accuracy and relevancy of popular music multimedia retrieval and categorization may also be improved by including user-generated information and feedback. Adopting many intelligence teaching modalities in ME depends on developing more exact and advanced categorization algorithms for popular music multimedia. By strengthening comprehension of the distinctive elements and traits of popular music, which may design more exciting and functional ME programs that meet the various educational requirements of students [7]. With content-based retrieval, a related piece that fits the mood and tone of the video is found by examining the visual and aural aspects. User-based recommendations look at the audience's or the video creator's musical tastes to provide appropriate background music. To offer the right background music, collaborative filtering examines the preferences of other users who have made or watched comparable movies [8]. A wide range of musical abilities and skills need to be formed and developed for music education. More specifically, these include an aesthetic sense for listening to and appreciating music, the capacity to choose a performance repertoire, acoustic qualities of tone, tempo, hearing, and voice, and the capacity to express and control emotions and feelings while playing music, musical instrument proficiency, among many others [24]. Music is a type of expression that uses structured sound and rhythm to communicate ideas, narratives, and emotions. It spans a variety of genres, styles, and cultural influences, making it a shared and engrained aspect of the human experience. On the other hand, the mathematical idea of Markov's

rule deals with probabilistic changes in states or occurrences. Exploring the patterns and dynamics that underpin this ageless and varied art form is made possible by combining the rich tapestry of music with the Markov's rule's predictive capacity [25].

Music programs are expensive, music is only a diversion from academics, and music is neither a legitimate profession than a legitimate pleasure. The motivation is to use the CHMA to retrieve melodies while assisting in the creation of a multi-media musical arrangement. The techniques offer a keyboard foundation it promotes the development of musical abilities.

As we proposed, Multimedia Music's layout, with the help of the Continuous Hidden Markov Algorithm, melody retrieval may be done. The methods provide a keyboard foundation that supports the growth of musical skills.

The remaining portion of the manuscript is organized as follows: Segment 2 explains the prior Study about the Research's goals or objectives and notes any restrictions or differences. Segment 3 provides suggestions for further Study based on the results and outlines the research technique and methods utilized to gather and analyze data. We go through the Discussion and Results in Segment 4 before presenting the research findings succinctly and methodically, evaluating and describing them in light of the study goals or objectives. The Study's primary characteristics, significance and contributions, possible effects on practice or policy, and prospective future research topics are all summarized in Segment 5.

2 Related Works

The Research [9] suggested a The Generative Theory of Tonal Music (GTTM) approach. Grouping Preference Rules (GPR) replace melodic phrases within a musical composition according to the resemblance of the term vectors. Via a user survey, the naturalness of the new melody was assessed, and the majority of people unfamiliar with the musical composition were unable to identify where the song had been substituted. The case study preferred an alternate strategy using deep neural networks to learn end-to-end similarity metrics. The findings for the Meer Tens Music Collections, which include many vocal and instrumental monophonic compositions from Dutch musical sources that span five centuries, are presented in the article to show how reliable the learned similarity measures are [10]. The article [11] examines the music melody which can be utilized to identify the musical technique genre and find musical compositions. The convolutional neural network (CNN), a deep learning method, was employed in the paper to

extract lyrical melody characteristics and identify genres. In the training procedure, three tuple samples were utilized as training samples. Every simulation tests, contrasting studies were carried out on the number of musical segments and the kind of activation function in the algorithm. The paper [12] suggested the Support vector machine (SVM) and the conventional CNN methods were contrasted with the CNN algorithm. The study [13] examines artistic Endeavour and, social engagement, group music making. By engaging musically with their patients and developing music, therapists may use music's social benefits to improve their patients' well being. As the activity calls for highly qualified specialists, the customers must pay a hefty price. The Research suggested a serious game that allows players to collaborate to create a beat using drum pads, allowing those without musical experience to engage. The gaming system enhances the aesthetic experience essential to musical engagement and its therapeutic properties by analyzing the player's performance. The article evaluated the approach using quantitative indicators demonstrating its ability to maintain a user-set pace. A questionnaire given to test participants revealed that they liked it and thought it exciting and that adding music facilitated their involvement [14]. The Research has proven that listening to music is a habit inseparable from everyday life and a tool for individuals to express their true feelings [15]. The essay preferred the primary technique for people to locate music is still text-based information retrieval. However, the approach has glaring flaws and needs to be more laborious and effective. The article suggested a technique based on the contour of the musical tune to address the issue [16]. The report investigated the retrieval technique depending on the rhythm, musical elements and music's melody. It includes several challenges; including the expression of the song, feature extraction from the melody, melody matching, user query building, and music database development [17]. The Research implemented the algorithms are being improved, and the system uses the improved algorithms to do audio preprocessing, feature extraction, feature post-processing, and matching retrieval [18]. The study [19] used a music library of 100 different types of MP3 files and randomly chose four songs each time to test the system with various system settings, recording lengths, and background noise. To increases the effectiveness of music retrieval and offers theoretical backing for creating music-retrieving system implementation systems. The paper [20] used the technology of sound signal parameter analysis and extraction to present a purpose technique for evaluating the quality of voice music based on comparisons of sound parameter characteristics. A basic frequency discrimination model is used to identify the primary melody in singing.

The study [21] suggested a feature extraction LAM technique depending on the melody's form. The most crucial extraction element for content-based music retrieval is melody. To compare and recognize the rhythm and melodic aspects of the initial tune recorded in the database, user could humming a song from memories followed by extract the melody, rhythm, and additional details of the hummed audio information. The paper [22] proposed a nonlinear feature detection-based technique for segmented identification of musical notes. The division vector is decided by establishing the constraint requirements. The vector space matrices are utilized to evaluate and process the musical notes to represent the correlation of musical notes and acquire the results of music retrieval. The nonlinear feature detection kernel function approach is presented to music note recognition. The study [23] used Back-Propagation approach to create the music classification model to precisely characterize music data and extract various sorts of music characteristics. The method enables musicians to compose music in real time, resolves the conflict between listeners' assessments and the rate that music is produced, and preserves the benefits of conventional genetic algorithms (GA).

Problem Statement

The lack of availability and diversity in ME currently is a concern. Access to high-quality ME becomes difficult for many students, particularly those from disadvantaged circumstances. There is also growing worry that traditional methods of teaching music can't properly engage students or prepare them for the changing nature of music in modern times. The major goal of this study is to use the CHMA to improve the performance of Multimedia Melody's retrieving abilities in the field of music. Through implementing the CHMA, the multimedia melody retrieval system can be improved and optimized to better satisfy user expectations and possibly open up new opportunities in the field of music-related material retrieval.

The statement implies that the Research has improved our understanding of better deploying multimedia education for the music platform. The following might be among the Study's particular passages:

1. This project aims to increase Multimedia Melody's efficiency for music retrieval by using Continuous Hidden Markov Algorithm and assessing its effects.
2. We evaluate several trials with a group of music melody learners to assess the efficacy of our technique. We contrasted their effectiveness and happiness with those who used a conventional multimedia platform.

3 Experimental Procedure

3.1 Basic Principle of Music Multimedia Retrieval

Music multimedia melody retrieval is searching for music multimedia based on its melody. The goal is to retrieve music multimedia with a similar theme to a given query melody. There are several approaches to music multimedia melody retrieval, including **Template-based rescue**: In this approach, a set of template melodies are pre-defined and stored in a database. When a query melody is submitted, it is compared to the templates in the database to find the closest match. **Feature-based retrieval**: In this approach, the theme of the music multimedia is represented as a set of features, such as pitch, rhythm, and tempo. When a query melody is submitted, its features are compared to the parts of the music multimedia in the database to find the closest match. **Content-based retrieval**: In this approach, the melody of the music multimedia is analyzed in terms of its spectral and temporal characteristics. This analysis generates a set of descriptors that represent the melody. When a query melody is submitted, its descriptors are compared to the descriptors of the music multimedia in the database to find the closest match. Figure 2 analyzing the music multimedia retrieval which using demolishing trackers and interrogate engines as a powerful tool for analyzing and identifying music. It can be used to search for specific elements of music, such as identical melodies, and can also be used to search for music that has similar characteristics to a target piece of music. Table 1 demonstrates the Template-based rescue, Feature-based retrieval, and Content-based retrieval are all methods of searching and retrieving information from a database or information system.

Figure 2 analyzes the music multimedia retrieval which using demolishing trackers and interrogate engines as a powerful tool for analyzing and identifying music. It can be used to search for specific elements of music,

Table 1   Principles of music multimedia retrieval

| Principles	Benefits
Template-based rescue	• It is quick and easy to use. • It is ideal for simple queries or searches.
Feature-based retrieval	• It can search for features across multiple documents. • It is ideal for complex queries or searches.
Content-based retrieval	• It is ideal for searching large collections of unstructured data. • It is useful for finding information that may not be explicitly.
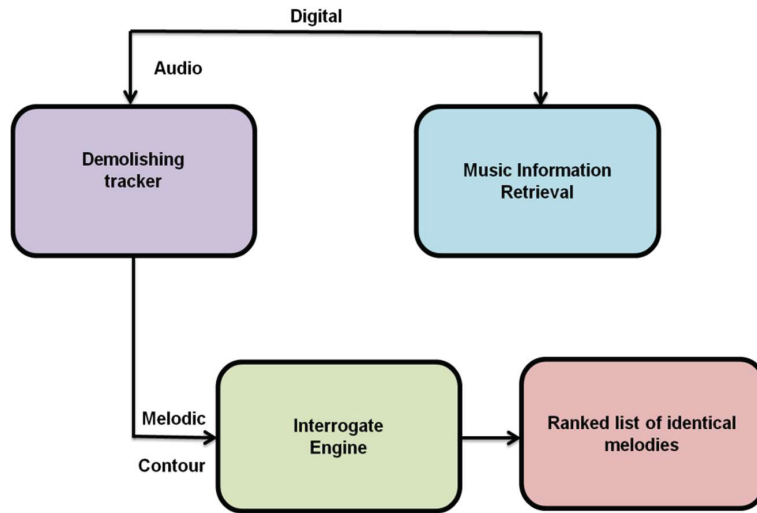
Figure 2   Analysis of music multimedia retrieval.

such as identical melodies, and can also be used to search for music that has similar characteristics to a target piece of music.

3.2 Music Multimedia Retrieval Using Continuous Hidden Markov Algorithm

3.2.1 CHMA architecture

The CHMA algorithm is a variant of the Hidden Markov Model (HMM) algorithm used to model continuous data streams. In contrast to the standard HMM, where the observed data is discrete and typically consists of categorical or binary variables, the CHMA deals with continuous-valued data that may be sampled from a continuous distribution. A collection of hidden states and a set of observable conditions make up our system's CHMA structure. Whereas the visible states indicate the acoustic characteristics of the song or subject, the hidden conditions reveal the underpinning structure of the piece. Whereas an emission matrix models the emission probabilities of the visible states, a transition matrix models the transitions between concealed states. Figure 3 shows the performance of CHMM-based music retrieval systems depends on several factors, such as the quality of the music features, the number of states in the HMM, and the amount of training data. The training and testing time for CHMM-based retrieval solutions can vary depending on the complexity of the HMM model and the size of the music.
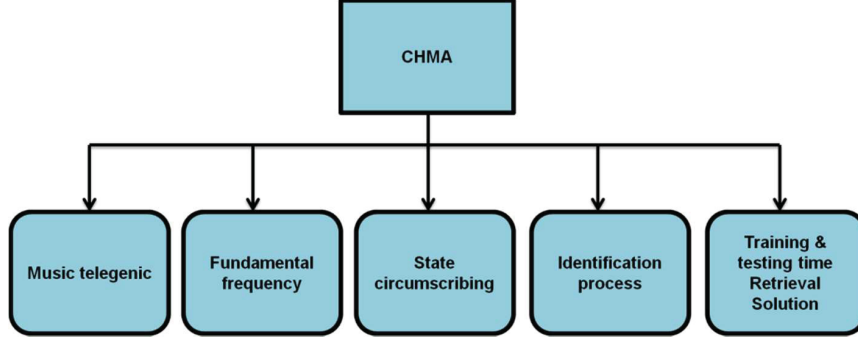
Figure 3   Implementation of the CHMA algorithm.

Using a left-right CHMA, where a state represents each note, it is possible to aspects in an accessible fashion. In a CHMA, notes consecutively pitched at the same level are characterized by a connected state. A t-dimensional Gaussian probability density function (PDF) generally describes each state in the CHMM model:

$$s(y, \mu, \Sigma) = \frac{1}{\sqrt{(2\pi)^t |\Sigma|}} \exp[-(y - \mu)^D \Sigma^{-1}(y - \mu)/2], \qquad (1)$$

In this situation, the PDF's average vector and covariant matrix are as follows,

$$s(y, \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{y - \mu}{\sigma}\right)^2\right], \qquad (2)$$

When y is a scalar observation, the state probability of observation y in the state i for a given word is the value of $s(y, i, i2)$ for the condition I indicated by $b\ g(y, i)$. Moreover, we construct a transition probability (abbreviated as b d (j, i)) between states j and. Simply condition j or condition $j + 1$ is available as potential future steps for observations in our left-right model at state j. As a result, the transition probabilities are subject to the following restrictions:

$$\begin{cases} 0 \leq b_d(j, i) \leq 1, \forall j, i \\ b_d(j, i) = 0, \text{ if } i < j \text{ or } i > j + 1 \\ b_d(m, m) = 1 \\ b_d(j, j) + b_d(j + 1, j) = 1, \forall j \end{cases} \qquad (3)$$

where m is the amount of state in the restrictions above, and $b\ d(m, m) = 1$ means that the sole transition from the final state is to itself.

3.2.2 CHMA assessment

We must assess the model by determining the probability of the data it produced given a CHMA and a series of observations (pitch vector). Dynamic programming is often used to achieve this. More precisely, we need to build a nm table Z, where each entry is calculated by the following recurring equation about a certain measured velocity vector $y = [y\,1, y\,2, \ldots y\,n]$ of extent in addition to a CHMM denoted through m condition:

$$Z(j, i) = \text{In } b_g(y_j, i) + \max \begin{cases} Z(j - 1, i) + uInb_d(i, i) \\ Z(j - 1, i - 1) + uInb_d(i - 1, i) \end{cases} \quad (4)$$

where the transition probability's weighting factor is u. $Z(j, i)$ Stands for the most excellent cumulated log probability of the scenario in which the Ist $j\ observations = [y\,1, y\,2, \ldots y\,n]$ subsist produced through every Ist I condition of the CHMM method. The aforementioned recurrent equation's beginning conditions are as follows:

$$\begin{cases} Z(1, 1) = \text{In } b_g(y_1, 1) \\ Z(1, i) = -\infty \text{ for } i > 1 \end{cases} \quad (5)$$

Also, this model's overall maximum probability of the information $y = [y\,1, y\,2, \ldots y\,n]$ is equal to $(max\ Z(n, i)@i = 1m)$. After identifying the highest probability, we may locate the best route for assigning each pitch point to a state. Hence, using the above method, we can determine the probability of every song's model within a specific observable pitch sequence, $y = [y\,1, y\,2, ...y\,n]$.

3.2.3 CHMA training

To evaluate a model, we must first determine the best variables for a CHMA. According to the accessibility of human input samples or music scores, that could be done. These two techniques will be referred to as retraining according to scores and training based on connections.

- **Training based on scores**

Every song in the database is a monophonic series about musical notes, with every note's pitch and length as its identifiers. The procedures listed below may be used to determine the matching CHMA for a given music score in this format:

1. Every state in the CHMA corresponds to a musical note in the song. Every condition is equivalent to the pitch level of the message in semitones.

2. Every condition 2 has a subjective value of 1
3. Let the pitch points in this condition equal r. The conversion prospect $b\,d(j, j+1)$ is similar into 1/ due to the fact that only the concluding pitch point of the condition will advance to the following stage (r).

Hence, $1 - b\,d(j, j+1)$ is set for $b\,d(j, j)$. We have $b\,d(m, m) = 1$ and $b\,d(m, m+1) = 0$ for the last note. For example, if the sampling rate is 8000 Hz, the frame size is 256, and the overlap is 0, then the appropriate value of r for a state representing a note with a length of 0.5 seconds is.

$$0.5 \times \frac{\text{sampling rate}}{\text{frame size} - \text{overlap}} = 0.5 \times \frac{8000}{256 - 1} \approx 16 \qquad (6)$$

- **Corpus-based Training**

There is at least one recording of the appropriate song to go on with corpus-based training of a model. For a particular model, corpus-based training entails the re-estimation processes listed below: Gather every pitch points that match each state I in the model using the best alignment route from the previous step. These pitch points may be combined into a vector, $[x\,1, x\,2, \ldots x\,o]$ (here, we omit the subscript i for convenience), from which the best parameters can be determined using the maximization of likelihood estimation.

$$\mu_i = \left(\sum_{r=1}^{o} x_r \right) \Big/ o, \qquad (7)$$

$$\sigma_i^2 = \left[\sum_{r=1}^{o} (x_r - \mu_i)(x_r - \mu_i)^D \right] \Big/ o, \qquad (8)$$

$$b_d(i, i+1) = \frac{k}{\sum_{r=1}^{k} t_r}, \qquad (9)$$

$$b_d(i, i) = 1 - b_g(i, i+1) \qquad (10)$$

while r is the entire amount of pitch points in the recording's r that are unique with respect to i, and k is the entire amount of measurements generated with this model. Table 2 compare the statistical models used in both approaches are typically based on probability theory, and the training process involves estimating the probabilities of different linguistic events given the available training.

Table 2   Methods of CHMA training

| Methods	Aspects
Corpus-based training	It can be effective when there is a large amount of training statistics available
	It does not require manual annotation of metrics.
Score-based training	It can be effective when there are limited training facts available.
	It is ideal for tasks that require high accuracy.

4 Analyzing the Factors Affecting the Effectiveness of CHMA

The CHMA is a statistical algorithm that is used to the dynamics of a system that is characterized by continuous observations. In this algorithm, the system is represented as a Markov chain, where the hidden states of the system are not directly observable, and the observations are the result of the underlying states. The CHMA is used to estimate the factors of the algorithm from the observations.

While the CHMA makes a priori assumptions about the distribution, these assumptions may not always be valid. As a result, the neural predictor (NP) is incorporated into HMMA in this experiment, enabling the proposed system to determine the likelihood of each state. It can deal with the issue of fixed input mode as well as the shortcomings of HMMA. The suggested approach may then be expanded to large-capacity audio retrieval. In other words, we mix the CHMA and NP models to recover audio data. The article's experimental audio samples were obtained from TV stations. In this experiment, we split them into 10 categories: (1) Advertisement, (2) poetry recitation, (3) weather forecast, (4) a basketball match, (5) song, and (6) traditional opera, (7) Live events, (8) Narration, (9) Sound effects and (10) Dialogue in Shows and Movies.

By using a diverse set of audio data, the experiment can provide a more robust evaluation of the CHMA performance across different types of audio signals. For example, some audio signals, such as weather forecasts, may have a more predictable structure, while others may have more complex patterns, such as songs with varying tempos and pitch. By evaluating the algorithm's performance across different types of audio signals, the experiment can provide insights into its strengths and limitations in other contexts.

The CHMA algorithm's training time is shown in Figure 4. It has been shown that the forecasts made about and generated by the recommended approach give more accurate results than the ones presently being used.
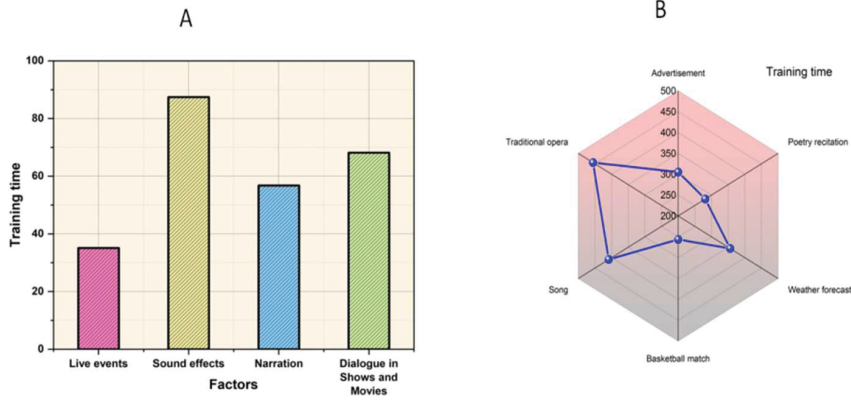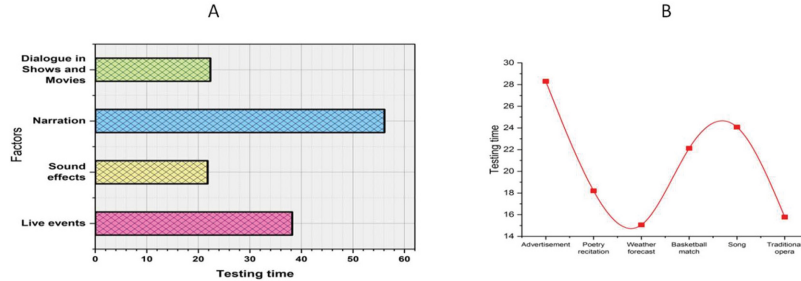
Figure 4   Training time of CHMA algorithm.

Table 3   Comparison of training time

Factors	Training Time
Advertisement	305.21
Poetry recitation	281.34
Weather forecast	356.08
Basketball match	256.41
Song	408.61
Traditional opera	455.76
Live events	35.12
Sound effects	87.43
Narration	56.80
Dialogue in Shows and Movies	68.14

A percentage of the total is often used to describe the amount of training time. The proposed algorithm CHMA has indicators that there might be erroneous predictions. The comparison of the training time for specified criteria, a poetry reading for an advertisement (305.21), Poetry recitation (281.34), Weather forecast (356.08), a basketball game (256.41), a song (408.61), a Traditional opera (455.76), Live events (35.12), Sound effects (87.43), Narration (56.80) and Dialogue in Shows and Movies (68.14). Table 3 displays a comparison of training time.

The CHMA algorithm's testing time is shown in Figure 5. It has been shown that the forecasts made about and generated by the recommended approach give more accurate results than the ones presently being used. A percentage of the total is often used to describe the amount of training time.

Figure 5   Testing time of the CHMA algorithm.

Table 4   Comparison of testing time

Factors	Testing Time
Advertisement	28.3
Poetry recitation	18.2
Weather forecast	15.06
Basketball match	22.12
Song	24.08
Traditional opera	15.78
Live events	38.2
Sound effects	21.8
Narration	56.10
Dialogue in Shows and Movies	22.32

The proposed algorithm CHMA has indicators that there might be erroneous predictions. The comparison of the testing time for specified criteria, a poetry reading for an advertisement (28.3), Poetry recitation (18.2), a basketball game (22.12), a song (24.08), and a classic opera, a Traditional opera (15.78), Live events (38.2), Sound effects (21.8), Weather forecast (15.06), Narration (56.10) and Dialogue in Shows and Movies (22.32). Table 4 displays a comparison of testing time.

The Recognition accuracy of CHMA algorithms is shown in Figure 6. It has been shown that the forecasts made about and generated by the recommended approach give more accurate results than the ones presently being used. A percentage of the total is often used to describe the amount of training time. The proposed algorithm CHMA has indicators that there might be erroneous predictions. The comparison of the recognition accuracy for specified criteria, a poetry reading for an advertisement (98.02), Poetry recitation (83.14), a basketball game (68.23), a song (55.23), and a classic opera, a Traditional opera (38.12), Live events (88.23), Sound effects (73.35),
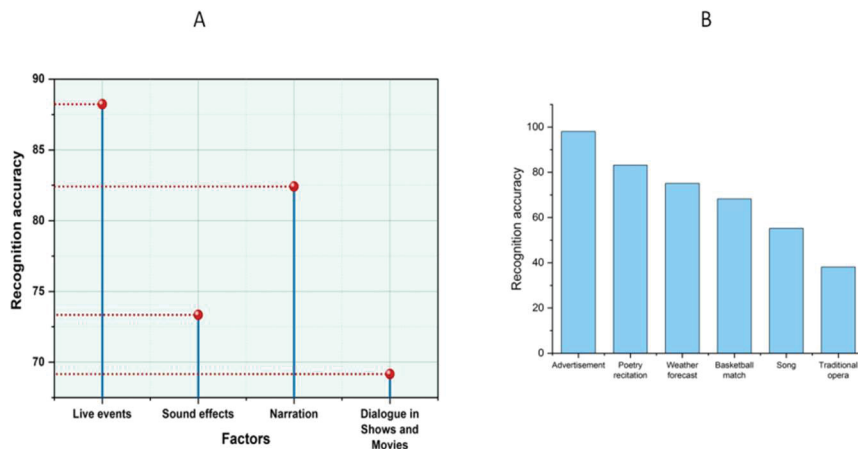
Figure 6   Recognition accuracy of the CHMA algorithm.

Table 5   Comparison of recognition accuracy

Factors	Recognition Accuracy
Advertisement	98.02
Poetry recitation	83.14
Weather forecast	75.08
Basketball match	68.23
Song	55.23
Traditional opera	38.12
Live events	88.23
Sound effects	73.35
Narration	82.42
Dialogue in Shows and Movies	69.17

Weather forecast (75.08), Narration (82.42) and Dialogue in Shows and Movies (69.17). Table 5 displays a comparison of Recognition accuracy.

5 Conclusion

The CHMAs have shown great potential in music multimedia melody retrieval due to their ability to model the temporal nature of music and use hidden states to capture the underlying structure of musical notes. We used an algorithm based on the HMM framework, but it is designed to model

continuous observations, making it suitable for analyzing multimedia music data. Using CHMA-based music multimedia melody retrieval, it is possible to accurately identify and retrieve melodies from an extensive database of multimedia music resources. This can be useful in various applications, such as music recommendation systems, search engines, and music education. However, the effectiveness of the CHMA-based music multimedia melody retrieval algorithm is the high quantity of training data and the selection of appropriate features and parameters. Training time for traditional opera took 455.76 minutes, while testing time for narration took 56.10 minutes. The recognition accuracy for advertisements significantly increased to peaking of 98.02%. In addition, the algorithm may face challenges when dealing with complex music genres or pieces with multiple melodies. Despite these challenges, CHMA-based music multimedia melody retrieval holds great promise for the future of music information retrieval and analysis. Further Research and development in this area can lead to more sophisticated and accurate music retrieval systems that meet the diverse needs and preferences of music listeners and creators.

Funding Statement

Acknowledgement

References

[1] Zhang, J., 2021. Music feature extraction and classification algorithm based on deep learning. Scientific Programming, 2021, pp. 1–9.

[2] Li, N. and Ismail, M.J.B., 2022. Application of artificial intelligence technology in the teaching of complex situations of folk music under the vision of new media art. Wireless Communications and Mobile Computing, 2022, pp. 1–10.

[3] Qin, T., Poovendran, P. and BalaMurugan, S., 2021. RETRACTED ARTICLE: Student-Centered Learning Environments Based on Multimedia Big Data Analytics. Arabian Journal for Science and Engineering, pp. 1–1.

[4] Wang, D., 2022. Analysis of multimedia teaching path of popular music based on multiple intelligence teaching mode. Advances in Multimedia, 2022.

[5] Dong, K., 2022. Multimedia pop music teaching model integrating semi finished teaching strategies. Advances in Multimedia, 2022, pp. 1–13.

[6] Rui, W., 2021, February. Application of the singing techniques in Oroqen folk songs teaching with the help of multimedia technology. In Journal of Physics: Conference Series (Vol. 1744, No. 3, p. 032244). IOP Publishing.

[7] Kratus, J., 2019. On the road to popular music education: The road goes on forever. The Bloomsbury Handbook of Popular Music Education: Perspectives and Practices, pp. 455–463.

[8] Wang, W., Li, Q., Xie, J., Hu, N., Wang, Z. and Zhang, N., 2023. Research on emotional semantic retrieval of attention mechanism oriented to audio-visual synesthesia. Neurocomputing, 519, pp. 194–204.

[9] Hirai, T. and Sawada, S., 2019. Melody2vec: Distributed representations of melodic phrases based on melody segmentation. Journal of Information Processing, 27, pp. 278–286.

[10] Karsdorp, F., van Kranenburg, P. and Manjavacas, E., 2019, October. Learning Similarity Metrics for Melody Retrieval. In ISMIR (pp. 478–485).

[11] Zhang, Z., 2023. Extraction and recognition of music melody features using a deep neural network. Journal of Vibroengineering, 25(4).

[12] Wu, J., Liu, X., Hu, X. and Zhu, J., 2020. PopMNet: Generating structured pop music melodies using neural networks. Artificial Intelligence, 286, p. 103303.

[13] Carnovalini, F., Roda, A. and Caneva, P., 2023. A rhythm-aware serious game for social interaction. Multimedia Tools and Applications, 82(3), pp. 4749–4771.

[14] Fedotchev, A.I. and Bondar, A.T., 2022. Adaptive Neurostimulation, Modulated by Subject's Own Rhythmic Processes, in the Correction of Functional Disorders. Human Physiology, 48(1), pp. 108–112.

[15] Lu, P., Tan, X., Yu, B., Qin, T., Zhao, S. and Liu, T.Y., 2022. MeloForm: Generating melody with musical form based on expert systems and neural networks. arXiv preprint arXiv:2208.14345.

[16] Zhang, J., 2022. Music Data Feature Analysis and Extraction Algorithm Based on Music Melody Contour. Mobile Information Systems, 2022.

[17] Chabin, T., Pazart, L. and Gabriel, D., 2022. Vocal melody and musical background are simultaneously processed by the brain for musical predictions. Annals of the New York Academy of Sciences, 1512(1), pp. 126–140.

[18] Li, Z., Yao, Q. and Ma, W., 2021. Matching Subsequence Music Retrieval in a Software Integration Environment. Complexity, 2021, pp. 1–12.

[19] Wang, T., 2022. Neural Network-Based Dynamic Segmentation and Weighted Integrated Matching of Cross-Media Piano Performance Audio Recognition and Retrieval Algorithm. Computational Intelligence and Neuroscience, 2022.

[20] Zheng, X., 2022. Research on the whole teaching of vocal music course in university music performance major based on multimedia technology. Scientific Programming, 2022.

[21] Zhang, J., 2022. Music Data Feature Analysis and Extraction Algorithm Based on Music Melody Contour. Mobile Information Systems, 2022.

[22] Shi, Y., 2022, April. Music Note Segmentation Recognition Algorithm Based on Nonlinear Feature Detection. In International Conference on Multi-modal Information Analytics (pp. 578–585). Cham: Springer International Publishing.

[23] Shi, N. and Wang, Y., 2020. Symmetry in computer-aided music composition system with social network analysis and artificial neural network methods. Journal of Ambient Intelligence and Humanized Computing, pp. 1–16.

[24] Ouyang, M., 2023. Employing mobile learning in music education. Education and Information Technologies, 28(5), pp. 5241–5257.

[25] Li, T., Choi, M., Fu, K. and Lin, L., 2019, December. Music sequence prediction with mixture hidden markov models. In 2019 IEEE International Conference on Big Data (Big Data) (pp. 6128–6132). IEEE.

Biography

Yingjie Cheng (1970.10–) male, Han Nationality, Hubei Qichun people, master of art, associate professor of Hubei Institute of Engineering, research direction: computer music creation, music education.