Optimal Design of Preschool Children’s Music Classroom Teaching System Based on Music Melody Retrieval System

Lijuan Tao
Yunnan Institute of humanities, Career Academy, Anning 650400, China

Abstract

With the continuous progress of social economy in China, people’s subsistence problem has been basically solved and the demand for spiritual civilization has further expanded. Under this background, the mass music culture has also been rapidly developed. Nowadays, music has become an integral part of China’s education system. As a crucial art form of spiritual civilization, music exerts a subtle influence on the thinking and behaviours of pre-school children. Consequently, the education community holds a higher awareness of music education for preschool children and conducts an in-depth study on the impact of music education for pre-school children. The current research results generally argue that music can enhance preschool children’s cognitive ability, improve their emotional intelligence, cultivate their cognitive movement system ability and reinforce their stress response system and memory system, and that music is of vital significance to promote the healthy development of preschool children’s learning and life. Influenced by a variety of subjective or objective factors, China’s traditional pre-school music education still holds some problems, leading to the poor quality of preschool music education and the difficulty to play its prominent educational role. In the meantime, the resources of pre-school children’s music education are limited and children often do not receive more diversified music categories, which becomes one of the most key factors restricting the development of preschool children’s music education. For this purpose, this paper constructs the music melody retrieval system, integrates it with the traditional preschool children’s music education, and provides a reference for the reform of preschool children’s music education.

Keywords: Music Melody Retrieval System, Music Education for Pre-school Children, Mass Music Culture.

1. RESEARCH OVERVIEW

1.1 Research background

Information retrieval technology organizes information according to certain rules and searches for the relevant information on basis of the basic requirements of users in the database or the Internet. At the earliest, this technology originated from the literature indexing and reference services in the library and was manually performed by human, with a relatively low efficiency and a long duration. With the continuous advancement of information technology, the traditional method of information retrieval has gradually changed from manual search to information retrieval. Especially in the context of the rapid progress of the Internet, the explosive rapid growth of information, knowledge and consultation exerts a tremendous impact on people’s lives and work. How to accurately and efficiently retrieve such information has also become one of the key issues facing the development of the Internet. The currently used main information retrieval tools, such as Baidu, Yahoo and Google, are mostly based on keyword matching retrieval and provide feedbacks to users in accordance with the relevance. However, this method has some defects as far as music is concerned, mainly due to the relatively high requirements for users to remember song names. Numerous users can not search for tracks that they can only remember the melody but not the song information. As time goes by, the amount of data in the music database will show a trend of rapid growth, and the workload of manual annotation will also be increasingly heavy.

1.2 Literature review

Music is an art form rich in spiritual and cultural connotations. Under the Internet age, music has become a undivided part of people’s lives and an integral component of the vast resources of the Internet. Most people search for and download the desired music through the Internet. The Internet also stores a substantial number of
songs, and the number of songs also demonstrates a rapid growth trend. In view of an enormous database of songs, how to accurately and quickly search out the desired songs of users has become a central challenge to the song search engine (Wu, 2012). The retrieval system based on searching music content in the music database by using the feature information of music files, including melody, tempo, pitch, etc. This retrieval method makes full use of users’ perception of the retrieved music and is more in line with users’ search habits than the traditional keyword retrieval method. Under this mode, users only need to hum some of the content by memory and then obtain accurate and fast retrieval results without having to memorize some of the song information, which is more convenient and quicker compared to the traditional search mode and brings a better user experience (Liu and Ma, 2012). The retrieval system based on music melody is applied in the following area. The first application is to retrieve in the music database through the Internet or mobile terminals. With the continuous progress of mobile phones and other mobile terminals, their technical level becomes higher and higher and their information processing capability is as good as computers in many ways. Users can search and download by humming a melody through a phone. The second application is the song requesting system. The traditional way of requesting a song is through remembering the name or the singer of a song. The song requesting system based on the melody detection system only requires humming to provide a search list and accurate results. The last one is music education. The melody retrieval system can extract users’ extended information, conduct a comprehensive analysis, identify users’ defects and deficiencies, put forward some targeted improvement strategies, and effectively improve their singing performance (Bai and Yan, 2012).

2. AN OVERVIEW OF MUSIC FEATURES

Before the study on the music melody retrieval system, it is necessary to first identify the parameters and the basic features of music. Prior to the existence of this system, most of the music searches were mainly based on text messages, and retrieval could only be conducted by searching the keywords of music singer, lyricist, composer and lyrics in the database (Wang and Li, 2012). The music melody retrieval system is an approach to retrieve information based on the characteristics of the music files itself, which requires little information about the music works. This retrieval mode can make people search in a more intuitive and natural way and it is more aligned with the basic needs of music retrieval (Huang, 2013).

2.1 Fundamental music theory

In physics, sound is generated by the object vibrating at a certain frequency and takes sound waves as the main mode of transmission. Sound can pass through solid, liquid, gas and other media, spread to human ears, and cause the vibration of tympanic membrane. Consequently, sound is transformed into brain waves through vibration so that humans can process and interpret sound, which ultimately forms the sound in people’s daily lives. The frequency of sound is extremely high. However, human ears have a limited capacity, and only the sound with a frequency between 20Hz and 20000Hz can cause aural responses (Yin and Fei, 2013). As a special form of sound, music has a certain height where voice, musical instruments and other sounding objects periodically vibrate with regular rules and generate a generally fixed frequency. Music mainly has the following features.

The first feature is note. Notes are used to record the core symbols of various music information, and different notes represent different lengths of sound and also have different manifestations in the music score.

The second feature is pitch. Pitch mainly represents the level of music notes. From the perspective of physics, pitch is mainly affected by sound frequency. The higher the frequency is, the higher its pitch is. In music, C, D, E, F, G, A and B respectively correspond to 1, 2, 3, 4, 5, 6 and 7 in music (Wang and Li, 2013). In general, because of the different physiological development of men and women, women’s vocal cords vibrate at a higher frequency than men, so their pitch is generally high, which is also the main reason that sopranos often appear in the bel canto, while tenors are seldom found.

The next feature is length, which is mainly used to describe the length of sound duration. Generally, fast-paced songs have a shorter length and slow-paced songs have a longer length.

The last feature is volume, also known as sound intensity or loudness, which is mainly used to describe the strength of the sound. From a physics point of view, pitch is more focused on the vibration frequency of the sound source, and volume is used to reflect the vibration magnitude of the sound source. The greater the magnitude is, the stronger the volume is. The smaller the magnitude is, the weaker the volume is (Lan and Yan,
2.2 Physical characteristics of sound

Humming signal processing methods are divided into three categories: the time-domain-based method, the frequency-domain-based method and the mixed method. The fundamental frequency analysis method based on time-domain, the earliest and the most widely used analytical method, directly applies the time-domain waveform of acoustic signals to analyze the characteristics of the average short-term energy, the average short-term magnitude, the zero-crossing rate, the short-time autocorrelation function and the average short-time average magnitude difference function of speech signals. The method based on frequency domain analyzes the frequency domain characteristics of speech, including the spectrum, power spectrum, cepstrum, etc. of speech signals. The mixed rule is to analyze the acoustic perception characteristics of speech by using the time-domain or frequency-domain characteristics of sound signals, but the amount of calculations is usually complicated and computationally intensive. The time-based analysis method is simple and intuitive, and the extracted feature information is accurate and widely used (Sun and Lu, 2011). This article applies the method of time domain analysis and the following characteristics of the sound signal.

2.2.1 Average short-term energy

It is assumed that the signal is \( x(n) \) and the window length of the window function is \( N \), and its short-term energy formula is as follows:

\[
E_n = \sum_{m=n}^{n+N-1} x^2(m)
\]

(1)

In the calculation of short-term energy, the main step is to derive the square of the signal value. Therefore, in the obtained calculation results, the signal can exert a more profound effect on short-term energy. In addition, in practice, logarithmic energy is frequently applied, and its formula is as follows:

\[
E_n = \sum_{m=n}^{n+N-1} 10 \times \log_{10} (x^2(m))
\]

(2)

By logarithmically calculating the above formula, the unit of the finally obtained value is decibel. The strength of the value obtained in this way is similar to how the human ear perceives the sound (Yan and Nan, 2014).

2.2.2 Average short-term magnitude

The short-time average energy obtained in the above formula still holds some shortcomings, which are mainly reflected in the overly high sensitivity of the function to the electrical level value. Besides, after the square operation, the gap between the high and low signals is added by human action, which greatly increases the computation burden. In the actual application, the average magnitude is usually adopted to reflect the volume size of the music. The average short-term magnitude is calculated as follows:

\[
M_n = \sum_{m=n}^{n+N-1} |x(m)|
\]

(3)

Average short-term magnitude and average short-term energy can be used to indicate the size of the volume, also known as the energy, strength, intensity, etc. of music. Its role is mainly reflected in the following areas.

First, it can be used to distinguish voiced sound, unvoiced sound and noise. In physics, the energy of a sound can be used to differentiate between different types of sound. For example, voiced sound is much higher in energy than unvoiced sounds, and the energy of voiced sound is far higher than that of noise. According to this nature, a reasonable threshold can be set among voiced sound, unvoiced sound and noise during voice processing to distinguish different types of voice through intuitive data.
Secondly, in many cases, people are in a fairly noisy environment. The above content can be applied to effectively distinguish the noise so as to ensure the sound accuracy to the maximum extent.

In the end, it can be used for endpoint detection to estimate in accord with the beginning and the ending positions of the initial consonant or the simple or compound vowel of a Chinese syllable (Cai and Zhao, 2014).

2.2.3 Average short-term zero-crossing rate

The average short-term zero-crossing rate refers to the number of zero-crossing values in each frame, which is a highly significant characteristic parameter in the time domain analysis of speech signals. In terms of a continuous speech signal with a time horizontal axis, it can be observed that the speech waveform crosses the horizontal axis. In case of discrete signals, the sample values obtained after the observation are used. If adjacent samples have different algebraic signs, zero crossing is assumed to occur. The zero-crossing rate over a period of time is called the average zero-crossing rate. If the sinusoidal signal is a signal, the average zero-crossing rate is divided by twice the sampling frequency (Bai, 2016). For this reason, the zero-crossing rate can reflect the signal’s frequency information to a certain extent. But only when the speech signal is not a simple sinusoidal sequence, the result of the average zero-crossing rate is extremely inaccurate. Meanwhile, the zero-crossing rate is highly sensitive to noise. If random noise repeatedly passes the coordinate axis in the background, massive false zero crossings would emerge (Hou and An, 2016). The formula for the average zero-crossing rate is:

\[ Z_n = \frac{1}{2} \sum_{m=0}^{N-1} \text{sgn}\{x_n(m) - \text{sgn}\{x_n(m-1)\}\} \]  \hspace{1cm} (4)

where \(x_n(m)\) represents the voice signal of the \(n\)-th frame and \(N\) stands for the length of the frame. \(\text{sgn}[]\) refers to the symbol function, and its formula is as below:

\[ \text{sgn}[x_n(m)] = \begin{cases} 1, & x_n(m) \geq 0 \\ -1, & x_n(m) < 0 \end{cases} \]  \hspace{1cm} (5)

In general, the zero-crossing rate of a voiced sound is much lower than that of an unvoiced sound, and a vowel signal has a much lower zero-crossing rate than a consonant signal. By combining the above three sections, power failure detection can be performed to further determine the start position and the final position of the sound.

3. FUNCTIONAL FRAMEWORK OF THE MUSIC MELODY RETRIEVAL SYSTEM

Figure 1 illustrates the functional framework of the music melody retrieval system.

![Figure 1. System Functional Structure Diagram](image-url)
First, the user humming system is the user interface (UI) of the entire system, namely the part provided to users. Users can collect music symbols and upload them to the database through the humming system, thereby realizing the interaction between users and the system.

Secondly, in the music database management system, with the increasing quantity of music data in the Internet, the music database is also constantly expanding, which leads to the difficulty of arranging and managing the traditional music database. To this end, this system needs to establish a complete music database management module. The information contained in the music database mainly includes various information of the music itself. The music database management module registers, updates, deletes and plays this part of information and also includes the automatic extraction function of music features of MIDI music files. After the administrator registers a new file in this module, it is necessary to collect the feature information in the MIDI and to analyze the contour of the music melody.

Thirdly, in the function of playing music files, users can store the music signals in the format of wav by humming, and music files in the music library are stored in the format of mid (Zhang, 2015). Hence, a music player is needed, which should be able to play files in both the wav format and the mid format at the same time.

Fourthly, the humming retrieval system, one of the most complex and the most crucial module in this system, consists of three sub-modules: the preprocessing of sounds and signals, the presentation and expression of the fundamental frequency and the matching engine. Specifically, the basic frequency extraction adopts the classic ACF algorithm. By analyzing the periodic features of the signal, period detection of the fundamental tone can be realized.

Fifthly, after finishing humming, the user uploads it to the cloud and searches the related results by retrieving the music database. The system performs a series of algorithm operations, locates several similar results according to the features of humming in the background, and sorts them in ascending order based on the similarity or relevance. Users can identify the song by listening to the song to find out the desired results (Li, 2013).

REFERENCES

Li P. (2013). The music skills course of pre school education in Jiangxi science and technology normal university and its implementation status analysis, Music Time and Space, (14), 150-151.
Liu H.F, Ma D.Y. (2012). In the integrated curriculum for preschool education major means to improve students' music accomplishment -- five years of preschool education specialty curriculum integration and optimization on music education research, (04), 51-52.
Normal University preschool education (art direction) as an example, Journal of Yangtze Normal University, 28(02), 5-9.

