Computer Aided Music Structure and Hierarchical Feature Analysis

Ying Qin[1]* and Xiaoyu Xu[2]

[1]Department of Art and Sports, Huanghe S&T University, Zhengzhou 457000, China, wq1998317@sina.com
[2]Department of Art and Sports, Huanghe S&T University, Zhengzhou 457000, China, Xxy381190267@163.com

Corresponding author: Ying Qin, wq1998317@sina.com

Abstract. With the rapid popularization of the Internet and the wide application of audio compression technology, the effective combination of computer electronic audio processing technology and music has become the development trend of music analysis. In this paper, the instrument coding of music is realized through computer analysis coding method, and the analysis method of network model is constructed to realize the hierarchical feature analysis of computer music. Based on audio signal processing, Fourier transform, short-time Fourier transform and Meir transform are used to extract the acoustic features of music directly, and analyze the computer music structure features, which can save the analysis time and improve the accuracy of analysis. Finally, the music spectrum analysis test method is used to effectively realize the rationality of feature analysis, and the performance of the model is accurate.

Keywords: Music structure, music structure analysis, computer aided music, music hierarchy characteristics, feature analysis
DOI: https://doi. org/10.14733/cadaps.2022.S3.45-54

1 INTRODUCTION

With the rapid development of digital technology and the continuous update of the new media technology environment, the structure and hierarchical characteristics of music are gradually combined with the digital technology, and computer music arises at the historical moment. Computer music is a new music form which combines modern computer technology, electronic audio processing technology and music effectively. Computer music is attracting more and more attention and plays an increasingly important role in various fields of music. Based on the digital characteristics of computer technology, computer music has incomparable advantages over traditional music forms [1]. But most of the characteristics of music art still only stay in the way of parallel listing of the characteristics of music art, almost no music art features describe the internal hierarchical relationship and logical relationship between the characteristics of music art.

Chen et al. [2] proposed to separate the various elements of the overall musical style, and then quantify the separated elements, and then find out the corresponding relationship between each quantitative parameter and people's subjective feelings. Through the comprehensive application of computer to automatically extract music features and analyze music features, Kaleli [3] classifies the whole music level. Jiang [4] analyzed Rutoslavsky's technique of controlled accidental creation and the creation concept of macro rhythm structure, and explored the enlightening examples in computer music creation. Huang [5] puts forward the basic concept of algorithmic composition and leads to the most important hierarchical structure control in algorithmic composition. Brunkan [6] combines music with computer to complete music creation through computer music production technology.

In order to analyze the structure and characteristics of music effectively, the traditional way of analysis is to classify the music according to the acoustic characteristics such as timber, rhythm and mode of music. Although this classification method can achieve a high accuracy, it cannot cope with the high consumption of time and space resources brought by the increasing number of music libraries. Moreover, in the process of feature extraction, it needs to use very complex mathematical operations to process music signals, so as to extract valuable features. These features are often designed for a single task and cannot be used to analyze a variety of music of the same type.

2 HIERARCHICAL CHARACTERISTICS OF MUSIC

2.1 Characteristics of Music

The material of music is selected according to the needs from the many sounds that the human ear can feel. Sequence music determines in advance one or more parameters of pitch, intensity, time value and its arrangement sequence , and then repeats the arrangement sequence or the variation of the arrangement sequence in the music. Sequential music has abandoned all kinds of conventions of traditional music. Due to the numerous contingencies of sound selection, the music formed in the end contains so many contingencies that the music has become a musical equation arranged by functions and calculators [7]. Music has a standard for the choice of sound, but there is no fixed standard. The standard is open, amorphous, and constantly in flux. The perception of music is done by the ear, which means that music cannot be looked at by the eyes like painting, nor can it be touched by the hands like sculpture.

The art of music as sound will become the first feature of music as art; Other features of musical art will be derived gradually from this feature of music as sound art, in the form of movement in time to develop. As the art of sound, the timeliness and motility of sound naturally become the dependent characteristics of music based on the fundamental characteristics of sound. Timeliness of music is realized through the rhythm of music and the movement of beats. Rhythm is achieved by breaking and changing pitches. Music presents timeliness through the movement of rhythm and beat and points to the future based on this moment. On the staff, it is the forward flow of melody line. The melody line is exactly the numerical simplified graph of horizontal length and vertical pitch on the coordinate axis. The features of music art have a kind of internal logic level inference between them, rather than parallel and juxtaposed.

2.2 Computer Music

Computer music refers to the work and activities related to music made by computers or digital circuits with CPU, all of which belong to the scope of computer music. In the narrowest sense, computer music works are regarded as computer music only. Computer Music created using computers and MIDI technology.

Networked collaboration refers to connecting multiple sets of computer music working systems through the network, connecting each set of working systems to form a working system, just like working on a music production system [8]. To a great extent, it reduces the calculation pressure of each computer, thus improving the quality and efficiency of music production. The specific connection mode can be divided into MIDI interface connection mode, Internet connection mode and Ethernet connection mode. The MIDI interface is the simplest and most common way to connect. The convenient digital computer music system lays a foundation for the analysis of music structure, which makes the analysis of film music characteristics step towards the road of digitization and normalization.

2.3 Music Analysis Encoding Method

In the analysis of music, music notes are abstracted into network nodes, which requires digital coding of music score. Implement music machine encoding by encoding. Encoding is mainly based on the composition of the notes to encode. The format is: rhythm + high bass symbol + medium pitch, that is, with three digits to represent. The music analysis model uses Rhythm-Tonal coding mode to conveniently represent the nodes of each note. According to the characteristics of the designed Rhythm-Tonal coding mode, the Rhythm network and the tone network are established respectively for the characteristics of Rhythm and tone, and the analysis and mining are carried out according to Rhythm-Tonal coding.

2.4 Network Model Analysis Method

Music is based on a series of sound elements that are recorded by musical notes to form a continuous stimulus that is meaningful or attractive to the human brain. The two basic elements of a network are nodes and edges. The relationship between music and the network is as follows: A piece of music consists of several different notes. The notes in the music are regarded as nodes in the network, and the edges in the network are the connections of two adjacent notes in the music. The relationship between notes is described through the establishment of a complex network model, so as to reflect the nature of the interaction between notes in the music through analysis.

A network of notes, nodes being individual notes and adjacent notes connected by edges. The number of edges that come from a node and converge on that node is called the degree of that node, denoted by k. The degree distribution of the artificial connection network follows the power law. In the log-log coordinate system, it is a straight line, and the slope is the characteristic index. For music to build the network need to define the nodes and edges, a music can be seen as the sequence of notes, each individual in the music notes as a node, according to the sequence of events, when music playing from a connection to another notes each node is connected to a certain number of other nodes, as shown in Figure 1.

3 MUSIC STRUCTURE FEATURE EXTRACTION

Music analysis mainly consists of two modules. The key part of the system is the processing of audio data, which determines the accuracy of the system. There are usually three ways to deal with audio: extracting and analyzing the acoustic features of music directly; The music samples are transformed into spectral images to analyze them indirectly. Use raw audio directly for analysis. With the rapid development of deep learning, the application of convolution neural network image recognition in the field of MIR is gradually increasing.

3.1 Audio Signal Processing

Music is essentially audio, and its audio with a lot of detail. There are identifiable features in pieces of music that contain thematic melodies. The classification of electronic music is to analyze the identifiable features contained in the audio. In the process of analysis and recognition, audio signal processing comes into being because the human brain and computer perceive audio signals in different ways.

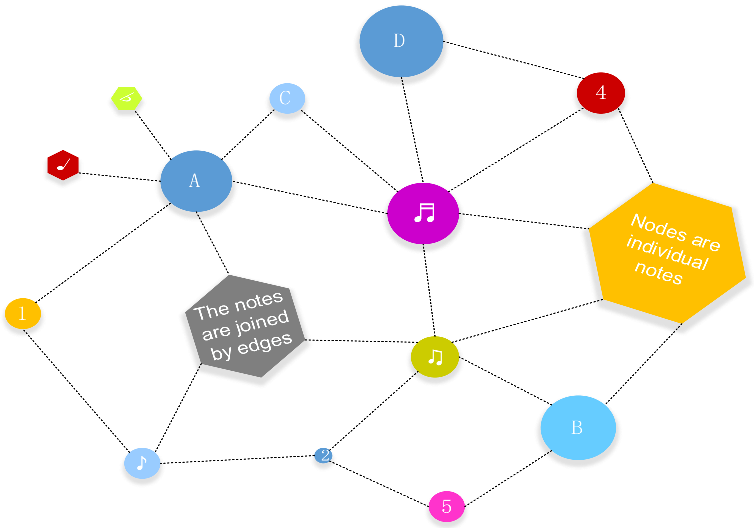

Figure 1: Model diagram of note network.

3.2 Fourier Transform

Waveforms of sound cannot be directly observed with the naked eye, but they are always present in people's life. A very intuitive example of this is that when a music player plays a piece of music, it displays a waveform of that piece of music. The waveform shows the relationship between time and loudness. The horizontal axis represents time and the vertical axis represents loudness. This is shown in Figure 2.

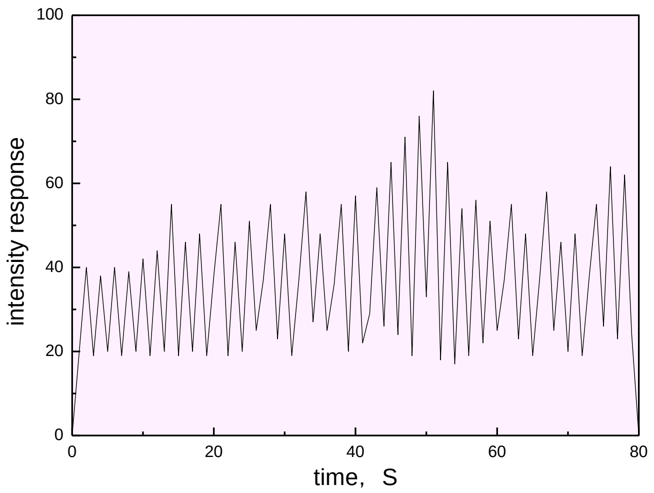

Figure 2: Music waveform.

Fast Fourier Transform (FFT) of the audio signal shown in the waveform can get the spectrum of this signal. The spectrum diagram shows the relationship between the frequency and amplitude of a signal, with frequency on the horizontal axis and amplitude on the vertical axis. This is shown in Figure 3.

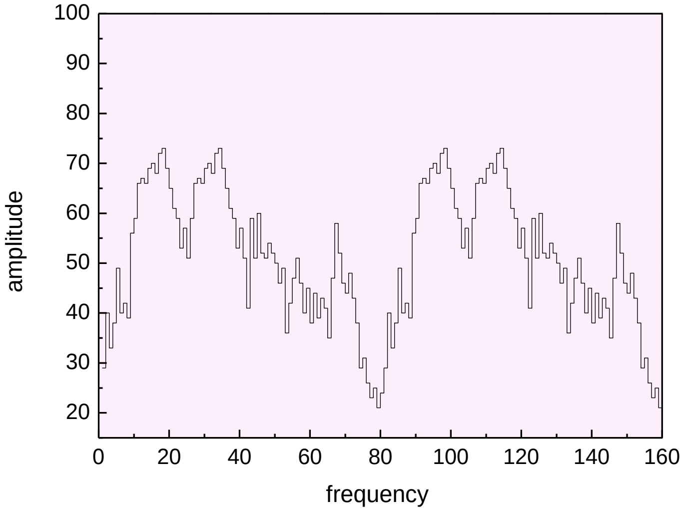

Figure 3: Signal spectrum diagram.

Waveform can reflect the characteristics of the signal in the time domain, while spectrum diagram reflects the changes of the signal in the frequency domain. The idea of serial time-domain and frequency-domain conversion was put forward by French scientist Fourier, and the method of signal time-frequency domain information conversion is called Fourier transform. The formula of Fourier transform is as follows:

$$f(x)=\int_{0}^{\infty} f(x)\log \pi r dx+\int_{-\infty}^{0} f(x) r^{3} dx \tag{1}$$

Where x is any real number, denoted by frequency, and the independent variable x represents time. F of x is the Fourier function of the original function, x of the spectrum of the signal.

3.3 The Short-time Fourier Transform

Fourier transform is to transform the information corresponding to the time domain into the frequency domain, so as to conduct subsequent analysis of the frequency characteristics in the frequency domain. Although a spectrum diagram shows the distribution of frequencies more clearly, the element of how long a frequency is held is missing. The Fourier transform is difficult to deal with for such an extended problem as trying to figure out exactly when frequencies begin and end. In order to solve this problem, the short time Fourier transform is generated to join the time-frequency domain information.

The specific method of short-time Fourier transform with Fourier transform as the core is to frame the original signal, add Windows and transform it again [9]. That is, you cut the whole signal into short signals of the same length, and at each of those short signals you apply the Fourier transform. This method increases the feasibility of accurate analysis in the frequency domain, and at the same time has the necessary information in the time domain, which enriches the research means of time-frequency analysis.

Short-time Fourier transform signal of the frame, the entire length of the signal can be divided into sub signal with the same size, the next operation to add a window, is each child signal with a non-zero window function multiplication, then the results according to the dimension of a stack, and stacked the spectrum drawings can clearly show the time-frequency characteristics of information. The flowchart of short-time Fourier transform of the audio signal is shown in Figure 4.

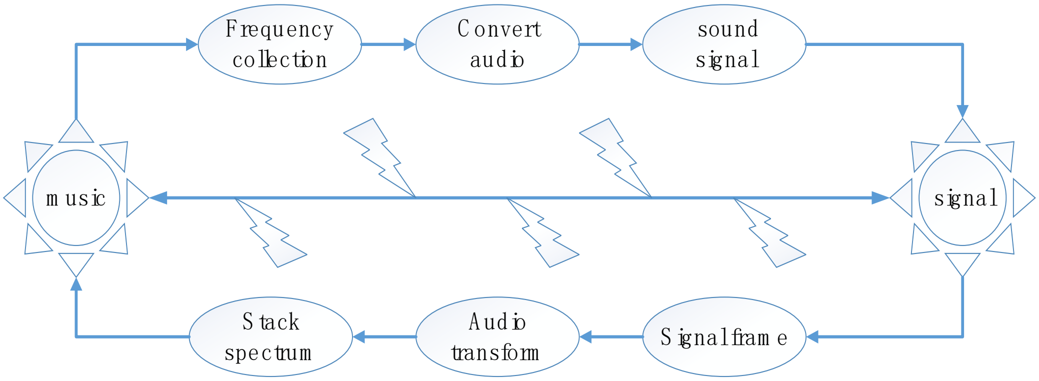

Figure 4: Short-time Fourier transform flowchart.

The short-time Fourier transform form can be expressed by the formula:

$$f(x) = \int_0^\infty (2+\pi) r x^3 dx \tag{2}$$

As it changes, the window function shifts along the time axis, the signal only leaves the part cut off by the window function for the final Fourier transform, and the result obtained is a complex function, which represents the magnitude and phase of the signal changing with time and frequency.

3.4 MEL Transformation

The human ear is not able to receive the sound of all frequency bands, just like a filter bank, only pay attention to some specific frequency components, let some frequency signals through, and directly ignore some frequency signals that do not want to perceive. The human ear usually receives sound information in the range of 20 to 20, 000Hz. In other words, human auditory nerve's perception of frequency is selective, without linear correlation [10]. It can be understood as follows: when the human ear hears a sound of 1000Hz, when the frequency of the sound suddenly increases by 1000Hz, the human ear will not notice the big fluctuation, but only notice a part of the increase in frequency rather than the doubling of the increase in frequency. By means of Mohr scale, the relationship between the ear's perception of frequency and Mohr frequency is linearly correlated. In the above example, when the frequency of the sound is doubled, the sensation is also doubled. Common frequency scale corresponding to Mohr scale conversion formula:

$$f_m = \lg(6x + 5x^2 + x^3) \tag{3}$$

Where x is the frequency value of the audio at the frequency scale, and F is the frequency value of the audio at the Myer scale.

4 ANALYSIS OF THE CHARACTERISTICS OF MUSIC STRUCTURE

Audio processing method, through the music signal for audio signal processing, we can get such as spectrum diagram, Meier cestrum coefficient and other characteristics in different kinds of music greatly different. The use of different audio signal processing means, the corresponding will get different characteristics, characteristics reflected by the characteristics of nature will be focused on.

4.1 Acoustical Signal

Music is also a form of sound. The music that people hear is also sound. Sound travels in nature in the form of waves, which can be thought of as continuous signals. If sound is viewed in a mathematical framework, then sound is a continuous function with independent variable as time and dependent variable as amplitude of vibration. Loudness is often referred to as volume, in decibels as a unit, the value of loudness reflects the amplitude of the sound wave, the amplitude of the sound wave directly reflects the volume, which is the size of the decibels. The amplitude of the audio signal in the waveform corresponds to the loudness. Pitch: Pitch corresponds to the pitch of a pitch. Tone is also used as a brief expression of frequency, so the unit of pitch is the same as frequency, which is Hertz (Hz). As the frequency changes, so does the pitch. People are sensitive to changes in pitch, but behind them is a change in frequency. Timbre: Timbre is a unique feature of all kinds of sound equipment, and every kind of equipment has a different timbre. The melody of music is a stack of notes of different pitches on the same timeline, even if the loudness and pitch parameters are the same for different types of voices.

4.2 Music Characteristics

Music analysis divides musical characteristics into short-term characteristics, long-term characteristics, semantic characteristics and component characteristics. The acoustic characteristics are divided into two parts: timbre characteristics and spatial characteristics. Tone, rhythm and harmonic features are the main components of perceptual features.

Short-time zero crossing rate: refers to the number of zero crossing of the music signal waveform. The higher the frequency is, the greater the short-term zero crossing rate is, the greater the change in the signal frequency is, and it can also indicate that the frequency of alternating between high frequency and low frequency is also greater. Short-time zero crossing rate can be calculated by the formula:

$$z = \sum_{n=5} f(x+1)(nx+1) + \sum_{n=3} f(x) \tag{4}$$

The cestrum features are converted according to Mohr scale. Meier scale is widely used because it can reflect the human ear's perception of sound change and Meier scale has a linear relationship. Meier cestrum coefficient is a further extension of Meier scale, which can also realize the mutual conversion between linearity and non-linearity. The extraction process of Mohr cestrum coefficient is shown in Figure 5.

4.3 Music Spectrum Analysis

According to the Meier scale, the audio wave signal is mapped to the Meier scale, and then shown by the image spectrum, the Meier spectrum is formed. Mohr spectrum is actually a re-transformation form of short-time Fourier spectrum, and the transformation form can be simplified to make nonlinear calculation of frequency axis of short-time Fourier spectrum. Thus, the transition from HZ scale to Mohr scale is achieved. The Mohr scale highlights low-frequency details, selectively abandoning high-frequency details such as noise. Such a transformation is used because the human ear is sensitive to changes in low-frequency output. Analytical features on the Mayer scale are close to human ear perception, which means that it is similar to the artificial discrimination of expert systems. Mohr spectrum diagram is shown in Figure 6.

4.4 Test Analysis

Computer music is a stage of digital music development, the computer occupies the development of science and technology in music creation, the processing technology and processing performance of the computer are getting higher and higher, the accuracy of sampling and quantification is also continuously improved, and the efficiency is improved. The analysis accuracy of the classification model reached 86. 17% by using the feature spectrum feature combined with ensemble learning method. The test results are shown in Figure 7. The analysis method is effective and reasonable, and the performance of the model is accurate.

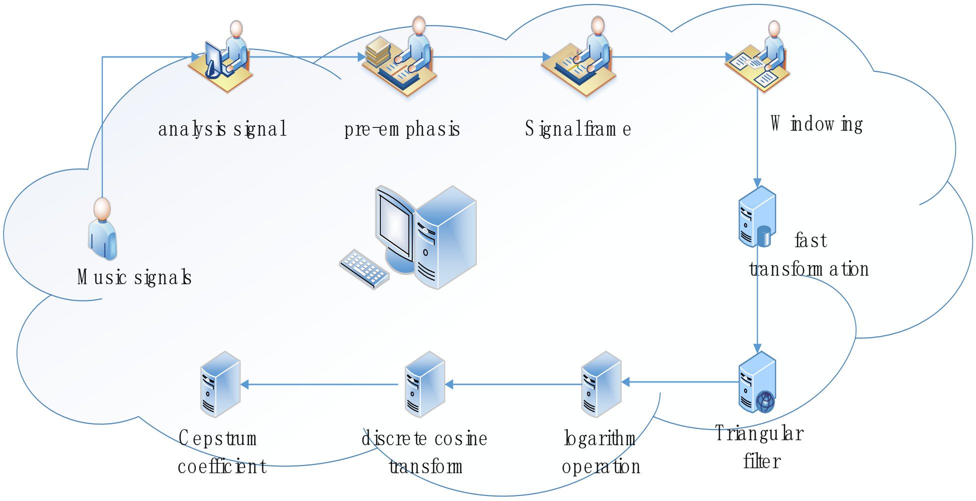

Figure 5: Calculate short time zero crossing rate process.

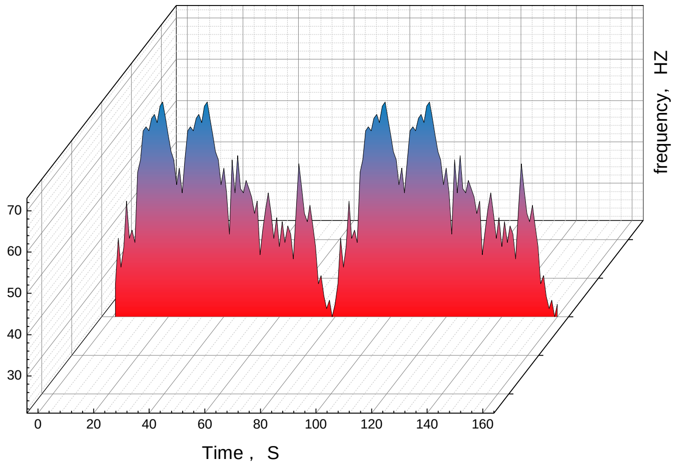

Figure 6: Meyer spectrum diagram.

5 CONCLUSION

Based on computer technology, this paper constructs a network model analysis method by using music analysis coding method, and summarizes the characteristics of computer music and music itself. The audio system discusses the method of extracting the acoustic features of music and the method of converting music samples into spectral images.

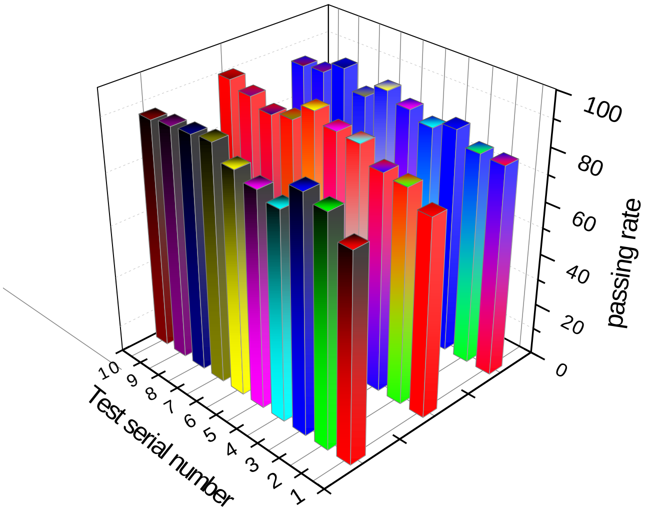

Figure 7: Test result diagram.

Music analysis is processed by Fourier transform, short-time Fourier transform and Meir transform to obtain music characteristic parameters and make feature analysis quantified. By analyzing the characteristics of computer music structure, we can improve the processing technology and performance of computer-aided music, improve the accuracy of analysis and quantification, and improve the efficiency of analysis.

Ying Qin, https://orcid.org/0000-0003-0244-1320
Xiaoyu Xu, https://orcid.org/0000-0003-3767-4140

REFERENCES

[1] Bi, X.; Shi, X.: On the Effects of Computer-assisted Teaching on Learning Results Based on Blended Learning Method, International Journal of Emerging Technologies in Learning, 14(01), 2019, 58-70. https://doi.org/10.3991/ijet.v14i01.9458

[2] Chen, Y.-T.; Chen, C.-H.; Wu, S.; Lo, C.-C.: A two-step approach for classifying music genre on the strength of AHP weighted musical features. Mathematics, 7(1), 2019, 19. https://doi.org/10.3390/math7010019

[3] Kaleli Y.-S.: The Effect of Computer-Assisted Instruction on Piano Education: An Experimental Study with Pre-Service Music Teachers, International Journal of Technology in Education and Science, 4(3), 2020, 235-246. https://doi.org/ 10.46328/ijtes.v4i3.115

[4] Jiang H.: Computer-Assisted Interactive Platform Design for Online Music Teaching Based on Cloud Computing, International Journal of Engineering Intelligent Systems, 27(2), 2019. https://doi.org/10.1007/10.14733/cadaps.2021.S2.92-101

[5] Huang Y.: Research on Interactive Creation of Computer-Aided Music Teaching, Solid State Technology, 64(01), 2021, 577-587. https://doi.org/10.1109/ICMT.2011.6001675

[6] Brunkan, M.-C.; Mercado E.-M.: A Comparison of Laboratory and Virtual Laryngeal Dissection Experiences on Preservice Music Educators' Knowledge and Perceptions, Journal of Voice, 33(6), 2019, 872-879. https://doi.org/10.1016/j.jvoice.2018.06.012

[7] Bhalke, D.-G.; Rao, C.-R.; Bormane, D.-S.: Automatic musical instrument classification using fractional Fourier transform based-MFCC features and counter propagation neural network.

Journal of Intelligent Information Systems, 46(3), 2016, 425-446. https://doi.org/10.1007/s10844-015-0360-9

[8] Perez,G.; Manuel, T.-J.; Morant, R.: Cantus: Construction and evaluation of a software solution for real-time vocal music training and musical intonation assessment, Journal of Music Technology and Education, 9(2), 2016,125-144. https://doi.org/10.1386/jmte.9.2.125_1

[9] Hamanaka, M.; Hirata, K.; Tojo, S.: Implementing methods for analysing music based on lerdahl and jackendoff' s generative theory of tonal music. In Computational Music Analysis, 2016, 221-249. https://doi.org/10.1007/978-3-319-25931-4_9

[10] Nam, J.; Choi, K.; Lee, J.; Chou, S.-Y.; Yang, Y.-H.: Deep learning for audio-based music classification and tagging: Teaching computers to distinguish rock from Bach. IEEE signal processing magazine, 36(1), 2018, 41-51. https://doi.org/10.1109/MSP.2018.2874383