International Journal of Web-Based Learning and Teaching Technologies
Volume 18 • Issue 1

Solfeggio Teaching Method Based on MIDI Technology in the Background of Digital Music Teaching

Shuo Shen, School of Contemporary Music, Shandong University of Arts, China*
Kehui Wu, School of Physical Education and Music, Qilu University of Technology, China

ABSTRACT

This research aims at teaching solfeggio and ear training in college music and proposes a teaching method for college music note recognition that combines the musical instrument digital interface (MIDI) and hidden Markov models (HMM). The experiment showcases that after preprocessing the music frequency sample signal using HMM model, it achieves the target accuracy after 20 times of training. From the HMM transition probability matrix diagram estimated from all training data sets, it can be seen that the transition matrix is close to the diagonal matrix. This indicates its high transfer efficiency. This study compares the HMM model with the other two algorithms, and the results show that its accuracy rate is about 99.56%. The probability of insertion errors and elimination errors is 0.52% and 2.58%. This is superior to the other two algorithms. In summary, the HMM model proposed in the study has extremely strong performance in the teaching of music note feature recognition in universities and can provide better teaching methods.

KEYWORDS

Feature Recognition, Fundamental Frequency Identification Method, HMM Model, MIDI Technology, Music Teaching

1. INTRODUCTION

With the development of technology, digital music teaching has become an increasingly popular method in modern music teaching. The digitization of musical works is also receiving more and more attention. Due to the increasing cost of piano teaching in universities, the teaching tasks of university music teachers are also becoming more and more difficult. Therefore, it is important to improve the efficiency of teaching and creation (Abeysinghe et al., 2021; Zeng et al., 2020). Today, with the continuous improvement of phonetic feature recognition, the accuracy rate of phonetic feature classification using traditional methods is not high. Traditional music teaching requires specialized teachers to guide and improve students' music level through repeated practice. Repetitive work not only greatly reduces the efficiency of teachers, but also increases the cost of one-on-one tutoring.

DOI: 10.4018/IJWLTT.331085
*Corresponding Author

This makes it difficult for low-income families to systematically learn music. In addition, in teaching, musicians' judgment of pitch is based on their rich teaching experience, and is based on the pitch heard by the human ear. However, this method of judgment is too subjective, inaccurate and prone to error. Introducing computer algorithm technology into music teaching can help students to reduce work pressure on the one hand, and can also help students to escape the influence of teachers to a certain extent. This can reduce the cost of learning (Lee et al., 2020). Among them, MIDI (Musical Instrument Digital Interface) serves as a digital interface for musical instruments, which uses the digital control signals of notes to record music. It can achieve the minimum size recording of music files, greatly reducing the cost of composition and orchestration. This MIDII technology is known as the "computer-understandable score" and can improve the efficiency of music creation. In traditional music teaching, the effectiveness of solfeggio and ear training is inevitably limited by classroom teaching techniques and individual differences among students, making it difficult to demonstrate good application effects. Therefore, this study aims to propose a teaching method for college note feature recognition that combines MIDI and HMM. The approach proposed in the study can combine traditional music teaching methods with modern technology, providing a more efficient and flexible learning approach. MIDI technology can achieve real-time performance, real-time feedback, and automatic grading functions, enabling students to learn musical knowledge and skills by interacting with electronic devices. At the same time, the multimedia function of digital music teaching software can effectively display the playing sounds of various instruments, deepen students' understanding of music theory and skills, and stimulate students' interest in learning while enriching teaching content, improving learning motivation. The Solfeggio teaching method based on MIDI technology can be quantified and customized according to students' performance, provide targeted guidance, accelerate the learning process, realize the personalization, interaction and intelligence of music teaching, provide students with a wider learning space and more efficient learning methods, and improve the quality and effect of music teaching. The exploration of music teaching methods under MIDI technology can provide new technological means and tools for university teaching in universities and personalized learning for students. From a social perspective, the Solfeggio teaching method in the context of digital music teaching can help cultivate more music talents and improve the popularity and effectiveness of music education. The widespread application of digital music technology provides new tools, means, and approaches for music teaching. The Solfeggio teaching method based on MIDI technology can effectively combine digital technology with traditional teaching methods, improve students' mastery and understanding of music knowledge, enhance their interest and learning motivation in music, and cultivate more outstanding music talents. In addition, the promotion and application of digital music teaching can also help promote the development of music education and enhance the country's cultural soft power. From a theoretical point of view, the Solfeggio teaching method based on midi technology can better integrate music theory and practice, providing more specific and visual learning materials and teaching methods. Through MIDI technology, students can intuitively see the symbols, melodies, intervals and other elements of music, and deepen their understanding of music theory and practice by combining actual instrument performance and arrangement. In addition, the application of MIDI technology to Solfeggio teaching can also expand the research field of music education and explore new teaching modes and methods.

2. RELATED WORK

There are various types of music, and the characteristics of various notes are also different. In music creation, many composers have used speech recognition software for note features to achieve automatic recognition of note features. However, in the recognition process, the classification accuracy of musical notes is relatively low. Therefore, an effective note teaching and recognition technology needs further exploration (Yl et al., 2020; Fonseca et al., 2020). Sosiawan et al. (2021) discussed the implementation of using HMM-GA in time series data. This research aims to improve the accuracy of prediction by developing an algorithm that combines Hidden Markov Model (HMM) and Genetic Algorithm (GA). This research has provided valuable contributions to the prediction field of time series data. Dinesh K et al. studied the implementation and analysis of FAR and FRR for face and speech recognition (multimodal) using KNN classifiers. Compared to other recognition methods, the KNN classifier achieves higher overall accuracy when used in face and speech recognition. The research results also reflect that FAR and FRR can be improved by using KNN classifiers. This indicates that KNN classifier is a reliable and effective method for face and speech recognition (Dinesh et al., 2020). Lg et al. (2020) explored the relationship between the estimation of voice similarity between human listeners and automatic speaker recognition systems that include speech features. The results illustrate that the use of speech features improves the accuracy of the system's speech similarity estimation compared to the use of acoustic features alone. Pandeya Y.R. et al. proposed a visual object detector for cow sound event detection. The proposed visual object detector is based on the ResNet-50 deep learning model for cow detection in video. The model was trained on a dataset consisting of cow and non cow images. The results demonstrate that the proposed visual object detector can accurately detect cows in the video, with an accuracy rate of 97.5% (Pandeya et al., 2020). Wu et al. (2021) analyzed the acoustic near-field characteristics of acoustooptic modulators (AOMs). They used numerical simulation methods to analyze the acoustic field distribution of AOM. The results indicate that there is an acoustic near-field that depends on the aperture size and the angular frequency of the acoustic wave. This study provides valuable information that can be used to improve the performance of acoustooptic devices.

Mittapalle et al. (2022) observed the laryngeal flow characteristics of vowels emitted by speakers with heart failure. The author collected voice samples from 16 male speakers with heart failure and 16 healthy male speakers. Both sets of samples were analyzed using throat flow analysis software. The results demonstrate that people with heart failure have significantly lower throat flow rates, steeper slopes, and larger area values. Zhao et al. (2021) studied how aircraft noise affects the spectral temporal acoustic characteristics of frog species. The research has shown that aircraft noise may have a greater impact on frog species that with long calls than on frog species that communicate with short, long calls. This study provides an insight into how aircraft noise affects the vocal behavior of amphibians and helps to further understand the potential impact of human activity on amphibian populations. Ding et al. (2020) proposed the problem of genomic privacy inference attacks using Hidden Markov Models (HMMs) and Recursive Convolutional Neural Networks (RCNN) models for unrelated individuals. Experimental results indicate that the proposed improved HMM model and RCNN model achieve excellent performance compared to traditional HMM models and other deep learning models. These findings provide useful insights into privacy inference attacks on genomic data and suggest possible directions for further research. Vioria et al. (2020) studied the segmentation process and spectral characteristics to determine music genres. They use a segmentation process to segment music into segments, and then extract spectral features from the segmented music to identify music genres. Experiments illustrate that the proposed method can successfully identify music genres with an accuracy of up to 86%. The results of this study provide important significance for understanding the characteristics of music genres and developing more effective genre recognition systems. Richard et al. (2020) studied the characteristics of frequency following responses to speech in newborns and their potential applicability in clinical practice. The results showcase that newborns have a frequency-following response to speech and are enhanced by sound cues. The frequency-following response of newborns to speech can be used to assess auditory temporal resolution and higher-order temporal processing. The authors conclude that the frequency-following response to speech can be a useful tool in the clinical evaluation of newborns.

Through the analysis of the application of HMM model in the teaching of music note feature recognition in universities by scholars at home and abroad, there are currently many algorithms for music note feature recognition and other feature recognition. However, there are relatively few HMM models used in the teaching of music note recognition in universities. Therefore, this study mainly focuses on the application analysis of HMM model combined with MIDI technology in music note recognition teaching in universities. Compared with other methods, the recognition rate is improved, and its feasibility and optimization are well verified.

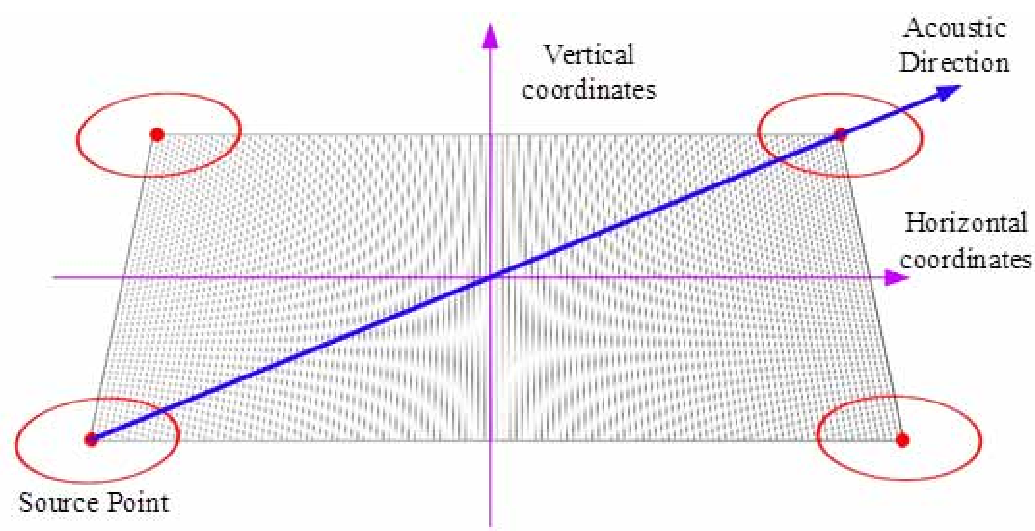
3. CONSTRUCTION OF A NOTE FEATURE RECOGNITION MODEL IN COLLEGE SOLFEGGIO AND EAR TRAINING TEACHING

Strengthening the effectiveness of visual singing and ear training teaching in music classrooms is an important teaching content. Among them, staff notation recognition is currently the main means of note feature recognition in universities. However, the staff notation covers a lot of content, and some of the content still has some ambiguity. The position and pitch of each note corresponding to the staff notation together form music audio data. To improve the accuracy of note feature recognition, research is conducted on computer simulation audio streaming based on MIDI technology, which includes digital signal transmission and reception, computer simulation of instrument sound, and mixing all simulated sounds to present the final music effect. At the same time, in order to ensure the accuracy of the feature extraction of digital signals, the acoustic characteristics of the human auditory system are satisfied with the help of Mel-frequency cepstrum coefficient and first-order Markov chain model, and the accurate recognition of the fundamental frequency characteristics of notes is realized, thus effectively improving the teaching effect of solfeggio.

3.1 Computer Simulation Based on MIDI Technology to Realize Audio Streaming Method

Computer simulation of acoustic wave propagation can be accomplished using digital frequency propagation technology (DFT). The Fourier transform can be used to transform sound signals from the time domain to the frequency domain, and then the frequency-domain signal can be propagated in the air. When sound propagates through air, it is affected by environmental factors, such as air density, airflow velocity, and sound reflection and absorption. These factors can affect the propagation path of the sound, as well as the intensity and frequency characteristics of the sound. Therefore, digital frequency propagation techniques can be used to simulate the propagation process of sound waves. In this way, the propagation of sound can be better predicted. A simple computer simulation of sound waves is shown in Figure 1.

Figure 1. Computer simulation of sound wave demonstration diagram

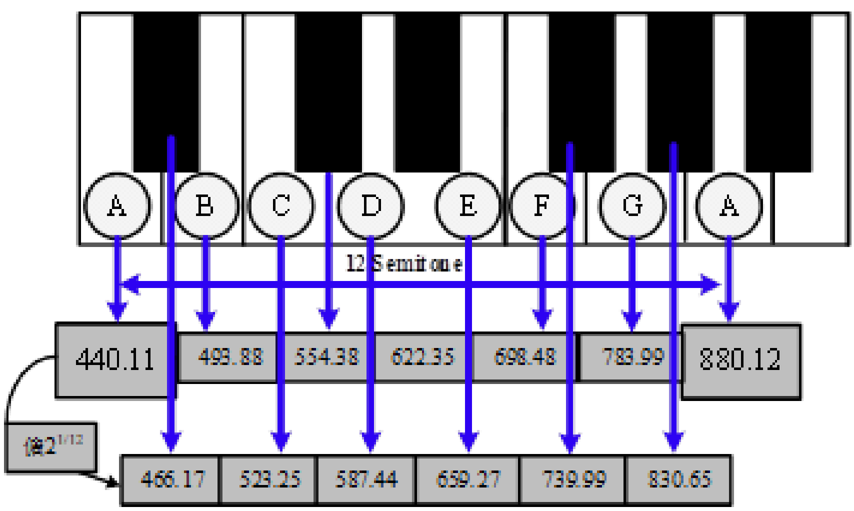


A certain number of digital particles are evenly distributed in a plane; The four corners of the matrix are distributed as sound sources, and the sine wave emitted by the sound source is fixed in both the transverse and longitudinal directions of the particles in three-dimensional space-time. The vertical coordinate is moved a new height by the sound wave, and the same row of particles moves up and down at the same time as the sound wave. This forms a wave. Computer simulation of computer sound simulation streaming can be divided into three parts: receiving MIDI signals, analyzing and processing MIDI signals, and mapping MIDI data to simulate audio streaming experimental variables. First of all, this study first receives MIDI signals: MIDI signals are a technology used to exchange digital music information. It can transfer music data from one device to another, such as a keyboard, musical instrument, mixer, etc; Secondly, the experiment simulates sound: Once a MIDI signal is received, the computer uses complex algorithms to simulate the sound of an actual musical instrument. The computer then uses predefined sound frequencies and environmental parameters to simulate the sound of an actual musical instrument; Finally, this study mixes sounds: The computer mixes all the simulated sounds to produce the final musical effect. This step can use predefined sound parameters such as EQ (Equalizer), reverberation, delay, and so on. This can change the final effect of the music. Sound is caused by the vibration of objects. The human ear can sense many sounds, but not all of them can become the basis of music. The sound used in music can convey people's lives and thoughts, and form a fixed system. This can be used to express people's thoughts and musical imagery. Tones are composed of single tones arranged in chronological order, with many individual tones in a continuous musical sound. In a physical sense, the basic components of a single note include three parts: fundamental frequency, amplitude, and octave (Bt et al., 2020). The frequency ratio of each adjacent key on a piano is the same, with their pitch frequency ratio being 1/12 times that of 2. The details are showcased in Equation (1):

$$f_2 / f_1 = 2^{1/12} \quad (1)$$

In Equation (1), f_n is the frequency of the n th key; f_{n+1} is the frequency of the n+1 th key. Taking the piano as an example, the variation of its pitch frequency is demonstrated in Figure 2.

Figure 2. Piano note frequency relationship



The twelve-tone average rhythm provides a scientific basis for quantitatively describing the tones of musical sounds. Intervals in different ranges can cause differences in pitch levels. The difference between the sweet and vigorous sounds of musical instruments is that they are composed of different materials; Commonly used vibration materials include steel strings, reeds, diaphragms, etc. Due to the differences in the sound quality of the components, their timbre is different (Wang et al., 2021). Original music without preprocessing is generally divided into two categories: one is power frequency interference caused by electromagnetic radiation; The other is that the energy of the frequency component is uniformly distributed in a specific auditory region. For power frequency interference, a high pass filtering method can be used to improve the high-frequency component in the fundamental frequency. Then, a low-pass filter is used to directly eliminate Gaussian noise. To determine the basic frequency of each note and calculate the instantaneous energy of each frame, it is necessary to divide it into a single note. The calculation equation is indicated in Equation (2):

$$E(i) = \sum_{i=1}^{m} x_i^2(k) \tag{2}$$

In Equation (2), m is the frame length. Research divides it into k frames, where i serves as a specific frame to identify a note; $x_i^2(k)$ is expressed as a functional function for calculating the energy of a single note. In a single frame note signal, its instantaneous energy has a significant impact on the sound. This experiment uses the average amplitude of the instantaneous energy to divide the silent and musical basebands. The expression is showcased in Equation (3):

$$E(i) = \sum_{i=1}^{m} \left| x_i(k) \right| \tag{3}$$

The instantaneous zero crossing rate reflects the frequency of a note. The study set the sampling frequency as f_s and the sampling frequency as f_0; Its zero crossing rate is $\frac{2f_0}{f_s}$. It sets the instantaneous zero crossing rate as illustrated in Equation (4):

$$ZCR(i) = \sum_{i=1}^{m} \left| x_i(k) - x_i(k+1) \right| \tag{4}$$

In practical applications, when the difference between two sampled note signals is greater than the set threshold, the instantaneous energy and zero-crossing rate of the note are normalized to facilitate the identification of fundamental frequencies (Nasrabadi et al., 2022). The normalization equation is shown in Equations (5) and (6):

$$E(i) = \begin{cases} 1, \sum_{i=1}^{m} \left| x_i(k) \right| \geq \delta \\ 0, \sum_{i=1}^{m} \left| x_i(k) \right| < \delta \end{cases} \tag{5}$$

$$ZCR[i] = \begin{cases} 1, x(i) * x(i+1) \leq 0 \\ 0, x(i) * x(i+1) > 0 \end{cases} \tag{6}$$

In Equation (5), δ is its threshold value. It detects a subsequent t frame. If it is found that all tones contain tones, then take this frame as the starting point and move it out of the basic frequency containing the music tone; If an instantaneous minimum occurs, subsequent t frames must also be detected; If there is no sound, the frame is used as the endpoint. After filtering, the signal is pre-emphasized. The pre-emphasis processing uses a digital pre-emphasis filter with a frequency of 6 dB/times. This is typically a first order digital filter (Sevilgen et al., 2022). Equation (7) showcases the details:

$$H(z) = 1 - \mu z^{-1} \tag{7}$$

In Equation (7), μ is its frequency coefficient, with a value close to 1. For single tone signals, continuous segmentation can be used for windowing and framing. However, to ensure a smooth transition between frames, overlapping segmentation is often used. The overlap between the previous frame and the subsequent frame is called frame offset, and its frame offset ratio is usually 1/2.

3.2 Solfeggio Ear Training Note Feature Recognition Model Based on HMM Model

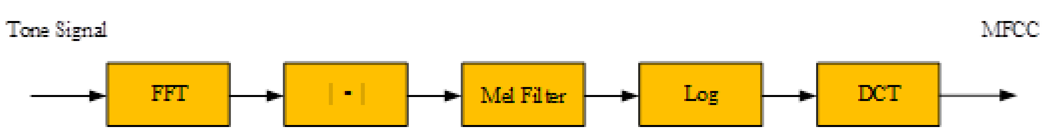
The Mel Frequency Cepstrum Coefficient (MFCC) is an acoustic characteristic that was created from the research results of the human auditory system. The Mel scale is a way to measure this critical bandwidth. The conversion relationship between Mel frequency (in mils) and linear frequency f (in hertz) is shown in Equation (8):

$$Mel(f) = \frac{1000 \ln(1 + \frac{f}{700})}{\ln(1 + \frac{1000}{700})} \approx 1125 \ln(1 + \frac{f}{700}) = 2595 \log(1 + \frac{f}{700}) \tag{8}$$

Generally, MFCC features are effective in speech recognition and recognition of various musical instruments in music. This feature has been studied as a method for piano single tone recognition. Because this method can reflect the sound energy distribution at different frequencies, and the energy of different piano single tones is concentrated in a specific frequency band. Therefore, MFCC can well describe the characteristics of monophonic sounds (Ham et al., 2020). The feature extraction principle of MFCC is shown in Figure 3.

In the study, the order of the cepstrum is selected as 16, thus obtaining a cepstrum feature with 16 dimensional output features. Due to the different length of each syllable, it is necessary to normalize the characteristic parameters of each syllable in time domain and amplitude for better processing. Finally, the unified characteristics of 16 * 4 dimensional monosyllables were obtained

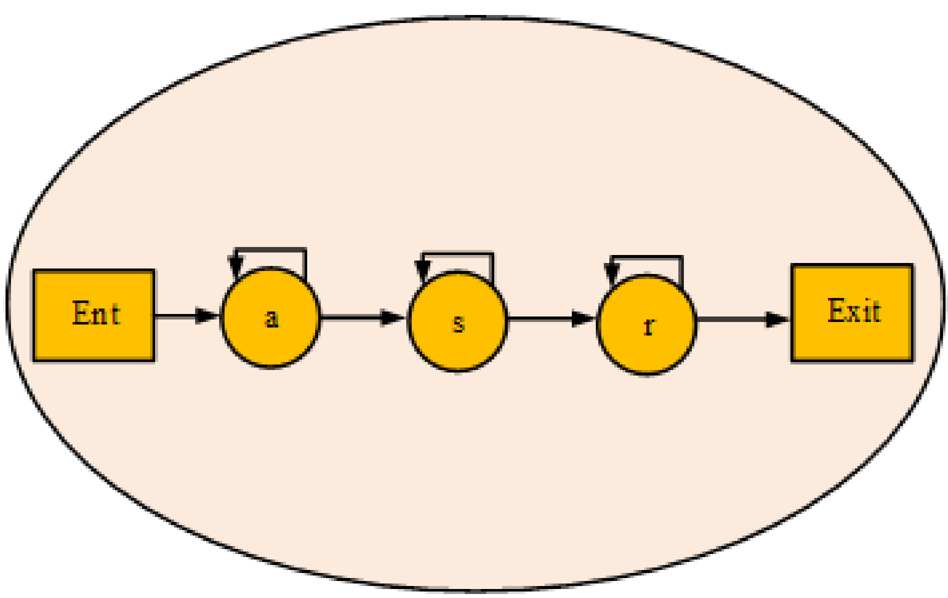
Figure 3. Block diagram of MFCC feature extraction



in the experiment. The basic frequency information of a note signal is reflected by the parameters of its basic frequency characteristics. To accurately identify the fundamental frequency of a note, it is necessary to extract its corresponding fundamental frequency characteristics. The notes in multi-part music are related to each other, and the frequency of the notes, which determines their pitch, is an important aspect of music development. Music audio includes both meaningful and meaningless parts, with the vocal part carrying more information and the silent part containing more fundamental features. The application of common feature recognition methods based on filter adaptation and differential washing of fundamental frequency features is relatively limited, and cannot be applied to the fundamental frequency recognition of notes in multiple parts. The HMM model proposed in the study is based on the first-order Markov chain model, and is mainly realized through the two Stochastic process of state description and the description between state and observation value. With the help of the HMM model, the note pitch recognition of multi-vocal score can realize the frame processing of fundamental frequency, parameter calculation, and the recognition of feature weight vector. To a certain extent, it has application value and effectiveness in music teaching. The topological structure of the notes is shown in Figure 4.

As shown in Figure 4, ent is an input state that has no physical significance. This marks the beginning of the mode, where a is the beginning of (vocalization), the instantaneous characteristic of a person speaking; "S" refers to the duration of the (delayed) sound. At the beginning, the sound will enter a stable duration; "R" is the last sound, and at the end, the sound has a brief attenuation until it disappears; Exit is an output state that has no physical meaning and is marked at the end of the model. In this experiment, the HMM mode is used to identify the fundamental frequency of multi vocal music scores. The HMM mode must be established for the fundamental frequency of each note. This process requires the use of multiple basic frequencies in HMM mode learning and training. The fundamental frequency characteristic is superior to the energy characteristic. According to the characteristics reflected by the music score characteristic, multiply each fundamental frequency characteristic parameter by the corresponding weighting. Then, the fundamental frequency characteristic parameters of each frame are combined to form a five-dimensional fundamental frequency characteristic vector. The existence relationship is shown in Equation (9):

Figure 4. Topology of a note model



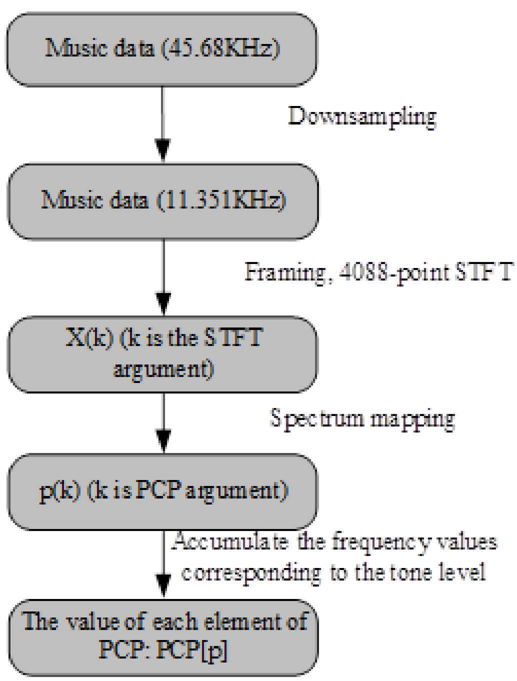
$$a_1 + a_2 + a_3 + a_4 + a_5 = 1 \tag{9}$$

In Equation (9), a_1, a_2, a_3, a_4, a_5 serves as its weight. Typically, empirical values are taken and the characteristic parameters are multiplied by corresponding weighting coefficients, as shown in Equation (10):

$$M_i = \begin{pmatrix} a_1 \times F_0, a_2 \times \frac{dF_{0i}}{dt}, a_3 \times \frac{d^2F_{0i}}{dt^2} \\ a_4 \times \frac{dF_i}{dt}, a_5 \times \frac{d^2F_i}{dt^2} \end{pmatrix} \tag{10}$$

In Equation (10), M_i is the vector of the fundamental frequency characteristic parameter between frames; F_{0i} serves as the fundamental frequency of the tone of the *i* th frame; E_i serves as the note energy at frame *i*. The characteristic of scale contour (PCP) is used by Fujishima as a characteristic of musical chords. The perception of music involves two completely different characteristics, namely pitch and scale. "Tone level refers to when a tone circulates within a certain octave (octave) period, while pitch refers to the increase in the sound of music as the frequency increases.". Therefore, the sound levels for two different octave ranges are the same. The sampling rate for music signals is typically 45.68 KHz. This sampling rate can well guarantee the quality of music and extract its characteristics. Therefore, the sampling rate is not required to be too high, and all music data is reduced to 11.351 kHz using the cool edit 3.0 software. Figure 5 is a flowchart for calculating the PCP characteristics.

Figure 5. Flowchart of PCP feature calculation



The specific steps are as follows: First, dividing it into frames, and dividing it into overlapping frames of 4088 sampling points. Subsequently, Fourier transform is performed, as shown in Equation (11):

$$X_{STFT}(q,n)=\sum_{m=0}^{N-1}x(n-m)\cdot w(m)\cdot e^{-j2\pi km/N} \tag{11}$$

In Equation (11), q serves as the coordinate frequency, and $0\le k\le N-1$; n serves as the center of the Fourier transform; $w(m)$ serves as the Hanning window at N=4088. Secondly, map $X_{STFT}(k,n)$ to $p(k)$. It typically includes 12 dimensional vectors, each representing the intensity of a chromatic scale. The mapping from frequency to scale is calculated using a logarithmic method in steps of 100 milliseconds, which is 10 PCP frames per second. Map k to p in PCP in STFT. The corresponding relationship is shown in Equation (12):

$$p(k)=[12\cdot\log_2(k\,/\,N\cdot f_{sr}\,/\,f_{ref})]\bmod 12 \tag{12}$$

In Equation (12), f_{ref} is the reference frequency corresponding to PCP; f_{sr} is the sampling rate. Finally, the frequency values of all frequency points corresponding to a particular scale are accumulated to the corresponding PCP unit values at each moment. The specific equation is shown in Equation (13):

$$PCP(p)=\sum_{k:p(k)=p}\left|X(k)\right|^2 \tag{13}$$

In this study, accuracy, insertion errors (by detecting the proportion of musical notes in a non tonal fundamental frequency), and elimination errors (by eliminating the proportion of fundamental frequencies of notes containing musical notes) were selected as the method evaluation indicators. As shown in Equation (14):

$$\begin{cases} Accuracy=\dfrac{\alpha}{\chi_1} \\ insert\ \ error=\dfrac{\beta}{\chi_2} \\ e\lim inate\ \ errors=\dfrac{\varepsilon}{\chi_1} \end{cases} \tag{14}$$

In Equation (14), α is the number of points whose fundamental frequency identification is correct and not zero; χ_1 is the number of points where the actual fundamental frequency is not 0; χ_2 is the number of points where the actual fundamental frequency is 0; ε is the number of elimination points. The insertion point refers to a point where the actual base frequency is 0 and the budgeted base frequency is not 0. The elimination point refers to a point where the actual basic frequency is less than 0 and its basic frequency is 0.

4. ANALYSIS OF THE APPLICATION EFFECT OF THE NOTE FEATURE RECOGNITION MODEL IN COLLEGE SOLFEGGIO TEACHING

The research selected the teaching tracks from the music courses of University A. It uses multi vocal music with a strong rhythm. The study used four hours of humming recordings, with one hour of bootstrap training and three hours of embedded training. The data is approximately 500 M, with nearly 20000 notes in the entire dataset; The entire test sample includes four female students and three male students as testers, trying to meet different student characteristics as much as possible. Each user hummed out 10 recordings at a sampling rate of 45.68 KHz, quantized at 16 bits, and filtered out 6 bass segments. Then this study obtained an experimental sample group, with a total of 64 humming segments, including 1073 notes. Then, the note data is pre-processed to remove Gaussian noise and normalize it. The comparison of the effects of the audio signal dataset before and after preprocessing is shown in Figure 6.

From Figure 6 (a), it can be seen that the maximum value of the signal frequency exhibited by the audio signal before preprocessing reached 9000Hz, and the overall fluctuation was more obvious, especially when the tag numbers were 0-500. When the sequence number is between 500 and 900, most of the signals exhibit concentrated high and low frequencies, and the signal data contains a lot of noise data. The frequency variation of the signal shows obvious abnormal fluctuations. After preprocessing the audio signal, the abnormal frequency change of the signal data in Figure 6 (a) has been significantly improved, with a signal frequency change amplitude range of (5800Hz, -5000Hz) and a smoother signal change. Figure 6 shows that the preprocessing not only removes the power frequency noise in the notes, but also removes the Gaussian white noise, making it close to the non-interference state. Based on this, an HMM model suitable for the fundamental frequency characteristics of a large number of notes was selected in the experiment; The fundamental frequency was identified through the filtered frequency domain. This reduces the impact of noise on the fundamental frequency interference, thereby improving recognition accuracy.

Figure 7 shows that the slope of the error curve of the pre processed music frequency sample signal is relatively large, and after 100 training sessions, the accuracy of the music frequency sample signal still does not reach the target value. The error value reaches 0.23 when the number of iterations is 10. After preprocessing the music frequency sample signal and 20 training sessions, the accuracy of the music frequency sample signal reached the target value; Moreover, the overall curve changes relatively smoothly, indicating a relatively stable data processing ability. The above results show that

Figure 6. Comparison diagram of sound frequency signal before and after preprocessing

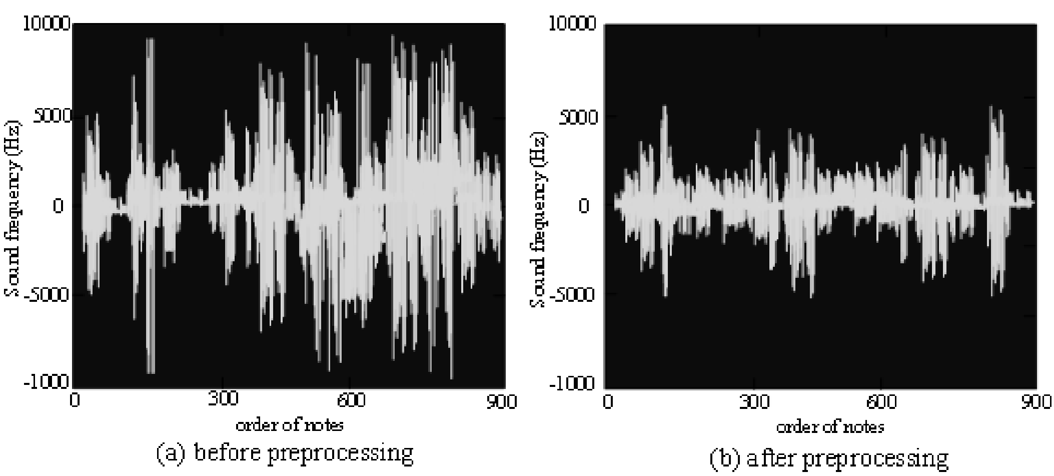
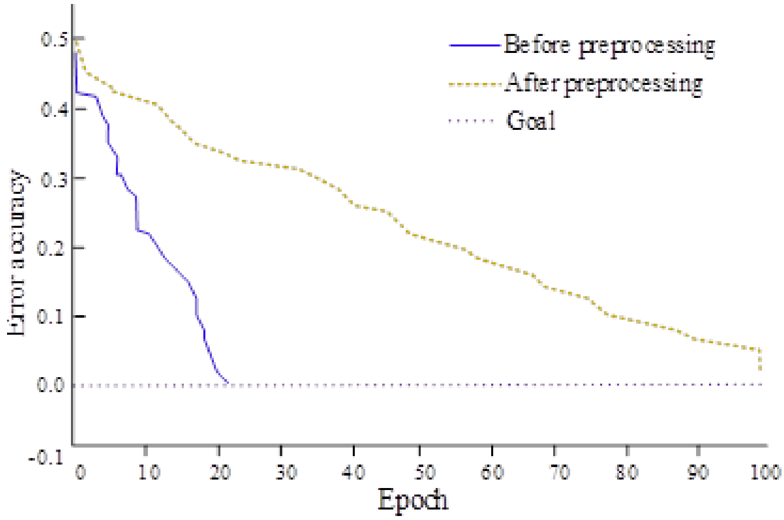


Figure 7. Comparison of training results



the music frequency sample signal is preprocessed. This can effectively improve the efficiency and accuracy of music frequency sample signal recognition, and save time and cost.

Figure 8 indicates an estimated HMM transition probability matrix from all training data sets. From the observation results, it can be seen that the transfer matrix is very similar to the diagonal matrix. Since its time length is generally larger than the frame length, it will not change in the next few screens. This leads to a high probability of being transmitted to oneself.

Figure 8. Transition probability matrix for HMM

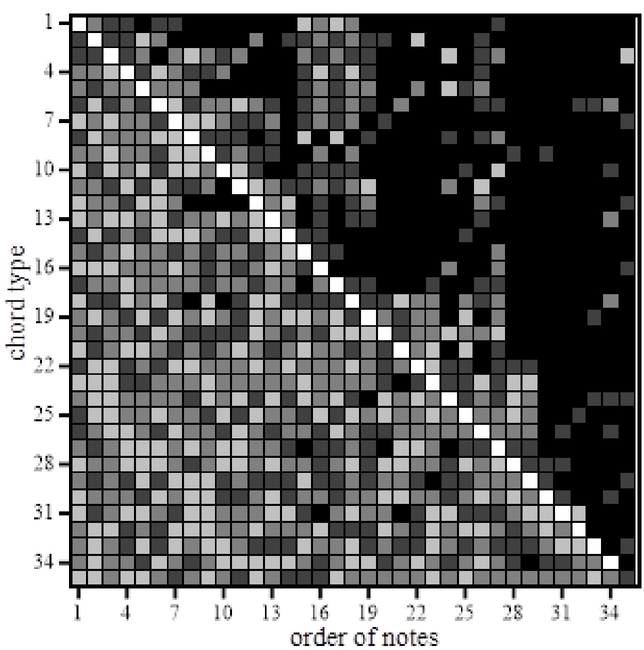


Figure 9 shows that changes in chords based on musical theory can also be identified by the possibility of transmission. For example, the large C chords mentioned above are most likely to remain in the same condition. The reason for keeping the large C chord unchanged is that the frame changes much faster than the changing chord. In addition, the possibility of moving from a large C chord to a large G or small F chord is higher than that of other chords, and the fundamental scale of a chord is four levels up. The major chord of the C key is 1, and its four ascending levels are 5, which is also called G. In Western music, it is very common for them to switch to each other. The big C chord is also the main chord of the bass F, so there are many times to tune from big C to big F. The non-diagonal area shown in Figure 7 also indicates a deviation of 4 to 5 semitones between the tonic tone and its tonic tone.

Figure 10 reflects the observed distribution parameters of the large C-chord estimated from the training data. Figure 9 (a) shows the average PCP vector of major C chords in the HMM, with a relatively uniform distribution of the overall average vector. There are obvious peaks on the three major

Figure 9. HMM major C chord transition probability

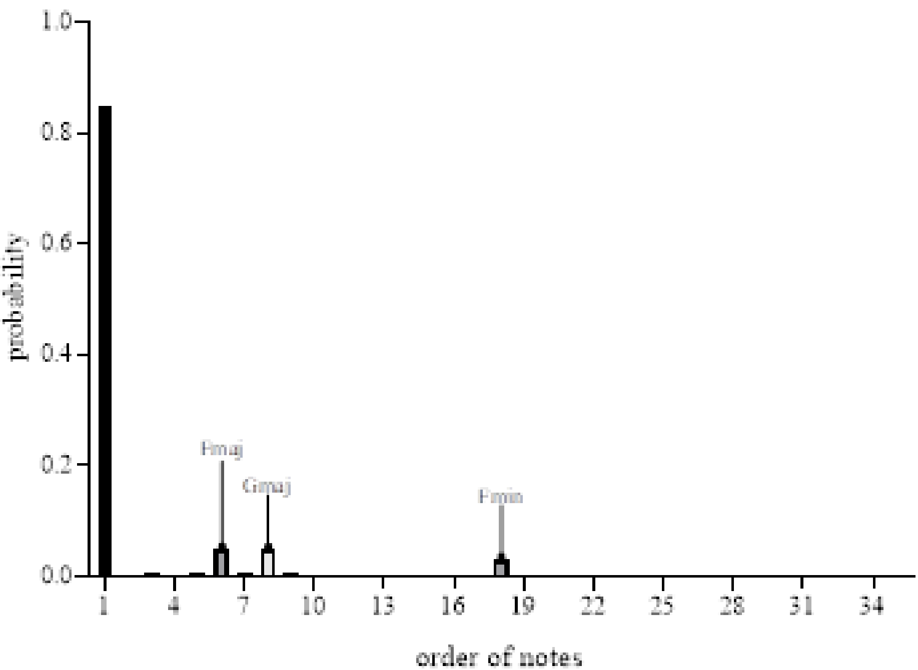
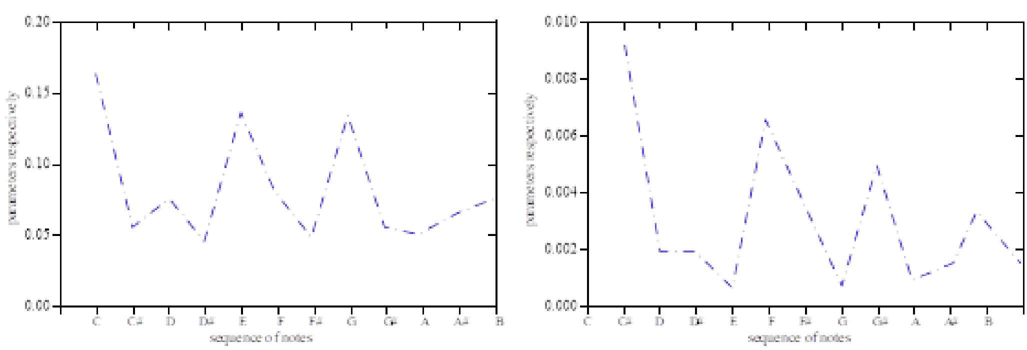


Figure 10. HMM algorithm vector comparison chart: (a) Major C chord PCP means vector, (b) large C chord covariance vector



tones C, E, and G, with peaks of 0.17, 0.14, and 0.13, respectively. Figure 9 (b) shows the covariance vector of the large C-chord in HMM, which has a high degree of autocorrelation. The autocorrelation of D #, E #, and G # is very small, with values of 0.002, 0.0007, and 0.0008, respectively, which proves the good performance of the HMM model. For ease of description, the collected note fragments are ordered in this article. They are A, C, B, and D. The fundamental frequency of the fundamental spectrum is identified through adaptive filtering, fundamental frequency feature identification, and HMM mode identification methods. Then it compares the recognition effects of the three methods, and the comparison results are shown in Table 1.

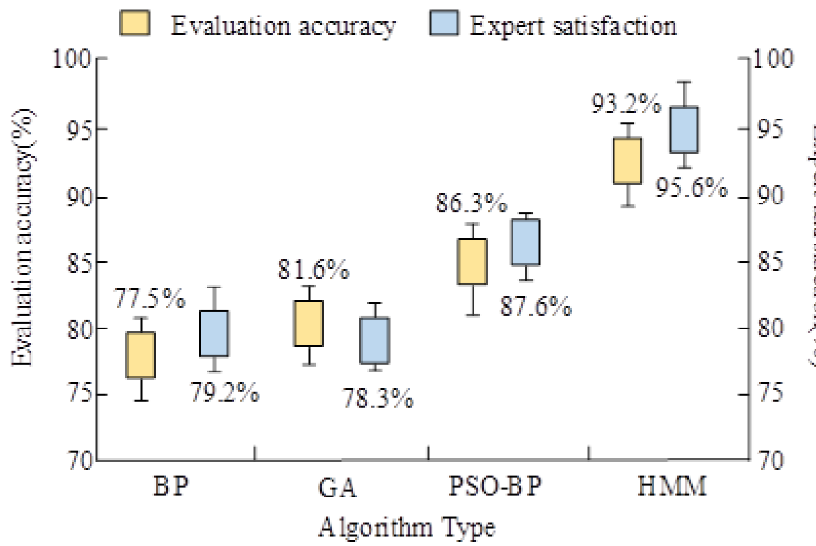
According to the analysis in Table 1, the recognition results of the filter adaptive method have significant errors, specifically manifested as the elimination errors of 89.67%, 84.89%, 86.45%, and 93.65% on the recognition results of four music segments, respectively. Moreover, the recognition accuracy of this recognition method on music segments does not exceed 70%, and the maximum recognition accuracy is only 68.52%. The elimination errors of the recognition method based on the fundamental frequency features of musical notes are 35.98%, 43.86%, 24.58%, and 16.97%, respectively, with an overall average accuracy of 67.04%. The recognition accuracy of the HMM model proposed in the study on music segments A, C, B, and D is 99.65%, 99.96%, 99.86%, and 98.95%, respectively, with an overall elimination error of less than 5%. On music segment C, the insertion error of the HMM model is higher than the insertion error adjusted by the filter, which effectively achieves error elimination and recognition of the fundamental frequency characteristics of notes. However, the insertion error of this method is relatively large, indicating that the method for detecting notes is incorrect. Therefore, the HMM model is superior to the other two methods in terms of insertion error and insertion error, and has a higher accuracy rate. Then, the algorithm is used to construct a project-based teaching method reform evaluation model. Moreover, the actual evaluation effect of different algorithms used in project-based teaching method reform software engineering teaching remains to be observed. By selecting 1000 music major students from a certain music university to participate in the experiment on a voluntary basis, the experimental subjects will learn the two teaching modes before and after the improvement, and there is no significant difference in the individual data of the students. Data on their satisfaction and learning performance under different teaching systems were collected and organized, and the results are shown in Figure 11.

Figure 11 indicates the accuracy and student satisfaction of the four algorithms in the evaluation model for solfeggio teaching reform. In the research, four algorithms are applied to the evaluation

Table 1. Recognition result comparison chart

Result		Music Clip			
		Fragment A	Fragment B	Fragment C	Fragment D
Filter Adaptive Method Identification Results	Eliminate errors/%	89.67	84.89	86.45	93.65
	Insertion error/%	0.95	0.68	0.98	1.59
	Correct rate/%	43.30	39.68	50.98	68.52
Recognition result of note fundamental frequency feature recognition method	Eliminate errors/%	35.98	43.86	24.58	16.97
	Insertion error/%	1.85	2.56	1.96	1.48
	Correct rate/%	54.98	79.56	44.26	89.36
HMM model recognition results	Eliminate errors/%	2.65	3.79	2.56	1.65
	Insertion error/%	0.35	0.98	0.19	0.73
	Correct rate/%	99.65	99.96	99.86	98.95

Figure 11. Accuracy and student satisfaction of different algorithms in the evaluation model of MIDI technology for the reform of sight-singing and ear-training pedagogy



model of solfeggio teaching reform. According to the evaluation results, the HMM algorithm has the best evaluation accuracy. The algorithm can accurately evaluate the parameters and indicators that affect the teaching reform, with an evaluation accuracy rate of 93.2%. The evaluation accuracy rates of BP neural network, GA algorithm, and PSO-BP algorithm are 77.5%, 81.6%, and 86.3%, respectively. In addition, the experiment also collected the satisfaction of music majors with the evaluation results of the four algorithm models. Student satisfaction can provide more effective ideas for the future direction of curriculum reform. The evaluation satisfaction of music majors with BP neural network, GA algorithm, PSO-BP algorithm, and HMM algorithm was 79.2%, 78.3%, 87.6%, and 95.6%, respectively. According to the satisfaction results, students are most satisfied with the evaluation effect of using HMM algorithm in the evaluation model.

5. CONCLUSION

Traditional music teaching in universities requires specialized teachers to provide guidance, and its repeated practice and repetitive characteristics greatly reduce teachers' work efficiency and invisibly increase teaching costs. At the same time, this teaching method mainly requires teachers to judge notes and pitches through human ear listening, which requires high comprehensive quality and professional ability of teachers, and has strong personal subjectivity, which is not conducive to students' self-learning and growth, and is also not conducive to the cultivation of professional music talents. Therefore, the study proposes the application of MIDI technology and HMM model in the teaching of note recognition in universities. The results illustrate that the algorithm achieves the expected accuracy after 20 times of training after preprocessing the music frequency sample signal. The study then used all training data for estimation. The results illustrate that the method is very close to the diagonal matrix, indicating that the self transfer efficiency is extremely high and the algorithm performance is extremely strong. It is significantly superior to the other two algorithms. This also indicates that the HMM model proposed in the study has extremely strong performance in the teaching of music note feature recognition in universities and has practical significance. The model is also helpful for the development of music teaching in universities and the improvement of teaching efficiency. Strengthening the Solfeggio teaching of midi technology in practical performance and arrangement is one of the important research contents needed in the future. At the same time, developing midi teaching materials and tools suitable for students of different ages and degrees is necessary to study the direction and ideas for future improvement. It is necessary to strengthen the personalized recommendation and customization functions of music performance and arrangement under the support of MIDI technology, strengthen the expansion of sample size for this method, and more practical results that need to be strengthened in the future.

COMPETING INTERESTS

The authors of this publication declare there are no competing interests.

REFERENCES

Abeysinghe, A., Fard, M., Jazar, R., Zambetta, F., & Davy, J. (2021). Mel frequency cepstral coefficient temporal feature integration for classifying squeak and rattle noise. *The Journal of the Acoustical Society of America*, *150*(1), 193–201. doi:10.1121/10.0005201 PMID:34340510

Bt, A., Nr, A., Emt, A., Cal, B., Zt, C., Bw, A., Bk, D., & Spb, E. (2020). Characteristics of Ophthalmology Trials Registered in ClinicalTrials.gov, 2007-2018. *American Journal of Ophthalmology*, *211*(1), 132–141. PMID:31730839

Dinesh, K., & Rao, P. V. (2020). Implementing and analysing FAR and FRR for face and voice recognition (multimodal) using KNN classifier. *International Journal of Intelligent Unmanned Systems, 8*(1), 55-67.

Ding, H., Tian, Y., Peng, C., Zhang, Y., & Xiang, S. (2020). Inference attacks on genomic privacy with an improved HMM and an RCNN model for unrelated individuals. *Information Sciences*, *512*(1), 207–218. doi:10.1016/j.ins.2019.09.036

Fonseca, E., Hershey, S., Plakal, M., Ellis, D., Jansen, A., Moore, R. C., & Serra, X. (2020). Addressing Missing Labels in Large-scale Sound Event Recognition using a Teacher-student Framework with Loss Masking. *IEEE Signal Processing Letters*, *27*(99), 1235–1239. doi:10.1109/LSP.2020.3006378

Ham, Y., Lee, J. H., & Lee, S. G. (2020). A Study on Evaluation in College Mathematics Education in the New Normal Era. *Communications of Mathematical Education*, *34*(4), 421–437.

Lee, C. H., Jwo, J. S., Hsieh, H. Y., & Lin, C. S. (2020). An Intelligent System for Grinding Wheel Condition Monitoring Based on Machining Sound and Deep Learning. *IEEE Access : Practical Innovations, Open Solutions*, *8*(99), 58279–58289. doi:10.1109/ACCESS.2020.2982800

Lg, A., MD, B., Fk, C., Aa, D., & Fn, E. (2020). Exploring the relationship between voice similarity estimates by listeners and by an automatic speaker recognition system incorporating phonetic features - ScienceDirect. *Speech Communication*, *124*(1), 85–95.

Mittapalle, K. R., Pohjalainen, H., Helkkula, P., Kaitue, K., Minkkinen, M., Tolppanen, H., Nieminen, T., & Alku, P. (2022). Glottal flow characteristics in vowels produced by speakers with heart failure. *Speech Communication*, *137*(1), 35–43. doi:10.1016/j.specom.2021.12.001

Nasrabadi, A. M., & Moghimi, M. (2022). Energy analysis and optimization of a biosensor-based microfluidic microbial fuel cell using both genetic algorithm and neural network PSO. *International Journal of Hydrogen Energy*, *47*(7), 4854–4867. doi:10.1016/j.ijhydene.2021.11.125

Pandeya, Y. R., Bhattarai, B., & Lee, J. (2020). Visual Object Detector for Cow Sound Event Detection. *IEEE Access : Practical Innovations, Open Solutions*, *8*(1), 162625–162633. doi:10.1109/ACCESS.2020.3022058

Richard, C., Neel, M. L., Jeanvoine, A., Connell, S. M., & Maitre, N. L. (2020). Characteristics of the Frequency-Following Response to Speech in Neonates and Potential Applicability in Clinical Practice: A Systematic Review. *Journal of Speech, Language, and Hearing Research: JSLHR*, *63*(1), 1–18. doi:10.1044/2020_JSLHR-19-00322 PMID:32407639

Sevilgen, G., Bulut, E., Albak, E. S., & Zturk, F. (2022). Prediction and optimization of the design decisions of liquid cooling systems of battery modules using artificial neural networks. *International Journal of Energy Research*, *46*(6), 7293–7308. doi:10.1002/er.7637

Sosiawan, A. Y., Nooraeni, R., & Sari, L. K. (2021). Implementation of Using HMM-GA In Time Series Data. *Procedia Computer Science*, *179*, 713–720. doi:10.1016/j.procs.2021.01.060

Viloria, A., Lezama, O., & Cabrera, D. (2020). Segmentation process and spectral characteristics in the determination of musical genres. *Procedia Computer Science*, *175*(1), 96–101. doi:10.1016/j.procs.2020.07.016

Wang, J. H., Wang, L., Xu, Q., Hou, Y., & Shi, X. M. (2021). Characteristics of consonant among children with speech sound disorder. Zhonghua er ke za zhi. *Chinese Journal of Pediatrics*, *59*(6), 478–483. PMID:34102821

Wu, J., Xu, Z., Li, K., Lv, G., & Wang, C. (2021). Analysis of Acoustic Near Field Characteristics in Acousto-Optic Modulator. *IEEE Photonics Technology Letters*, *33*(4), 201–204. doi:10.1109/LPT.2021.3051648

Yl, A., Rp, B., Td, B., Am, A., Lf, A., Md, B., Ac, A., & Bt, A. (2020). Music processing deficits in Landau-Kleffner syndrome: Four case studies in adulthood. *Cortex*, *129*(1), 99–111. PMID:32442777

Zeng, S., & Du, X. M. (2020). Theoretical analysis and experimental research of non-cavitation noise on underwater counter-rotation propellers. *Progress in Computational Fluid Dynamics*, *20*(1), 51–58. doi:10.1504/PCFD.2020.104713

Zhao, L., Wang, T., Guo, R., Zhai, X., Zhou, L., Cui, J., & Wang, J. (2021). Differential effect of aircraft noise on the spectral-temporal acoustic characteristics of frog species. *Animal Behaviour*, *182*, 9–18. doi:10.1016/j.anbehav.2021.09.014