Asia-Pacific Journal for Arts Education

Special Issue

Unfold the Future of Music Education through Technology

Co-editors:

Prof. Bo Wah LEUNG

Dr. Cheung On TAM

Dr. Chi Hin LEUNG

Dr. Richard Guy WHITBREAD

The Education University of Hong Kong

http://www.ied.edu.hk/ccaproject/apjae/apjae.htm ISSN: 1683-6995

Volume 22 Number 1 December 2023

Design of a College Teaching Model for Ethnic Vocal Singing Based on Artificial Intelligence Digital Algorithms

ZengYu Liu, MuXi Tang

Shenyang Conservatory of Music

tenorlou520@gmail.com

SiHao He

Shanghai University of International Business and Economics

sihaohe@126.com

Abstract

In recent years, there has been a surge in the development possibilities of ethnic vocal music performances and higher education teaching within the context of online education. The authors conducted a comprehensive investigation, taking into account the mixed drawbacks of conventional online education, in order to assess the current state of Internet technology development, a process leading to the development of a systematic model for summarizing the teaching system. In order to create a standardized model that integrates performance practice, teaching methods, and feedback mechanisms, the results of a statistical questionnaire survey conducted with 1657 individuals in the field of music in China were analyzed, as well as 22 Master's degree students in the Department of Ethnic Vocal Music at the Shenyang Conservatory of Music. The teaching mode is an integrated design of artificial intelligence, digital algorithm, vocal cord protection, vocal performance, applied music psychology, and vocal music teaching method, that supports the standardization and conscientization of teaching, as well as the visual display of evaluation records. To sum up, this model uses digital equipment, automatic management systems, and cyberspace as the carrier to naturally integrate national vocal music with the network using its software and extensive database, integrate resources in order to thoroughly improve teaching and performance practice, establish a professional ethnic vocal music teaching system, and address the drawbacks of online national vocal music teaching.

Key words

ethnic vocal music, vocal music performance, vocal music in higher education, vocal cord protection, artificial intelligence, digital technology

Introduction

Artificial intelligence (AI) has been used in many different industries due to the rapid expansion of the Internet. With its exceptional capacity for learning, AI has produced a large number of incredible artworks and melodies, including piano playing, in a short period of time (Civit et al., 2022). As a result of the multi-dimensional increase in the quality of teaching and singing, as well as the evaluation of teachers' instruction and the comprehension of course material by students, ethnic vocal music singing and instruction have also benefited from effective AI algorithms.

Modern online ethnic vocal music singing and instruction may be significantly improved by utilizing the real features of artificial intelligence. For example, AI can swiftly adapt to the performance and teaching environment by setting up an operating system that blends simulation, computation, learning, practice, feedback, and improvement. The flexible use of an offline teaching system, as discovered by Zhe (2021), is ideal for teaching vocal music in colleges and universities. Online instruction, on the other hand, is a relatively recent development that is still in the exploratory phase. The creation of a scientific, effective, and long-lasting online teaching model that can ensure the delivery of the educational material and even produce superior outcomes while utilizing resources sensibly (Wang, 2022) is thus an important first step.

The most effective approach to the transfer of knowledge is face-to-face, and vocal learning begins with imitation before moving on to secondary composition. Oral instruction, on the other hand, is less successful within the framework of online learning since students are unable to experience a genuine performance setting and are thus unable to completely comprehend what the instructor is trying to convey (Herbert, 2007). An AI-based education system thus offers advantages, such as the flexibility to study at any time and from any location, the capacity to provide expert services, and the assurance that outcomes obtained online can be reliably compared to those obtained offline (Lei, 2020).

The following research is presented in this paper:

(1) The teaching model, which supplements and adapts existing models, is based principally on the development of AI and is intended for Chinese folk vocal singing (Wei et al., 2022). This paradigm offers a strong basis for digitizing the instructional approach and course material for folk vocal music courses at universities, making it possible to create a "digital+" discipline and using it as a disciplinary instrument to support teaching and learning (Zhang, 2023).

(2) Comments on, and developments of, the features that users want from online teaching and learning tools, analyzing the opinions in a focused way. This feature involves the flexible use of AI (Yan, 2022), as well as the creation of easy-to-use instructional models for teachers and students.

(3) The development of particular concepts, the evaluation of findings from earlier studies, and an assessment of the viability of the ideas put forward in this study (Yu et al.,

2023).

The next four parts of the paper outline the research that was done. The first part covers the study's history and its guiding principles. The educational model design is described in the second part, together with outlining the prerequisites for performing and teaching folk vocal singing in schools, as well as describing the teaching methodologies that may be used for online delivery. It also explains the practical applications of vocal cord anatomy, music psychology, and vocal cord protection (Henrich et al., 2011).

Questionnaire analysis lays the foundation for AI design by examining what people are most interested in. The third part delves into the design concepts of artificial intelligence and an introduction to computers, while the fourth part summarizes the essential content of the text and introduces the teaching methods.

Related work

Since the development of the Internet, universities have experimented with different forms of teaching aids. For example, with the help of multimedia technology, the use of computers to produce class content, audio and video lectures, investigations into digital science as a means of managing tasks, and the development of small programs in chat software to assign homework (Webster, 2017), have all been explored.

In recent years, AI has made a splash in the arts, whether it be through conducting automatic piano playing or else writing a perfectly good opera excerpt in three seconds. In education, AI and the metaverse are also very much in the spotlight. As a result, there are a variety of opportunities for development within the realm of ethnic vocal

singing and education.

Liu (2018) proposed the “twenty-four character formula” as the main design principle for teaching ethnic vocal music singing. Adopting this method to divide the teaching mode into dynamic, behavioral and cognitive modes, the voices of two tenors and two sopranos, a collection of singing data from national vocal music, and some music data from the piano and electric piano, were all collected as part of the research process. It was decided that these data would serve as the foundation for AI analysis and its design, the aims being to improve the audio and video quality of online instruction (Roads, 1985), examine users’ vocal cord maintenance, and compare the collected audio data as the standard reference for singing (Elghaish et al., 2022). On completion, it was hoped that AI could help teach national vocal music singing using its algorithm and imitation capability.

Liu et al. (2021) developed a comparative study between control and experimental groups based on the watch-summary-question (WSQ) vocal learning assessment model (Hsia, 2022), while Valle et al. (2016) found peer assessment (PA) to be a learning strategy that had the potential to engage students in knowledge building and skill development through understanding teachers’ guidelines for assessing, learning and reflecting on peers’ vocal practice. Aimed specifically at studying the impact of WSQ-based mobile PA on students’ vocal skills and learning perceptions, and founded on the hypothesis that utilizing electronic devices improves students’ abilities to better evaluate their peers’ vocal technical abilities (while also improving their own performance), vocal learning software was used to compare WSQ-based mobile PA and conventional mobile PA by focusing on the five learning

units of vocal music (pitch accuracy, pitch centering, rhythm, expression or phrasing, and timbre quality).

The goal of a more recent study (Liu, 2022) was to determine whether or not mobile PA had any noticeable effect on students' vocal performances and learning attitudes, the ultimate aim being to guarantee that all students could engage emotionally, behaviorally and cognitively. This article focuses on the research methods used in that study while simultaneously using AI to design an effective and comprehensive vocal learning model.

Chinese ethnic vocal singing and teaching

A key singing concept for students majoring in ethnic vocal music is, "The articulation are driven by the cavity, enunciation can be driven by voice, then voice and prosody combination, and the voice is clear and strong." (Liu, 2021). Bel-canto perhaps comes closest in terms of style, but the singing effect is more natural than that of Chinese folk music, which favors "overtones" and whose timbre is more pronounced. Users' vocal cords are subjected to the collection of stroboscopic, perceptual, acoustic, and aerodynamic data in order to prevent vocal nodules (Fu et al., 2015).

Liu's "twenty-four character formula" stipulates that ethnic vocal music demands concurrent training in voice, singing exercises, pitch, and rhythm adjustments (Luo et al., 2020), underscoring the necessity for features such as automatic accompaniment. Moreover, it is crucial to extract noise based on the natural singing voice's vibration frequency and evaluate its potential impact on vocal cord health. The primary challenge in performing

national vocal music lies in conveying “emotion”. A variety of these, including “nationality, patriotism, love, labor, motivation, and positivity” (Liu, 2022), are integral to the genre, meaning that different song themes require specific, adept guidance in order to bring certain nuances to the fore.

Acoustic analysis and vocal cord protection

Breathing is very important for vocal cord protection (Valley, 2010), and is the prerequisite for vocal cord relaxation. In a study conducted by Hanyang University, Hwang et al. (2016) documented the effectiveness of basic breathing exercises in restoring laryngeal resonance, determining that during inhalation the strength of the breath and the correctness of breathing form the basis for determining the self-maintenance of the laryngeal vocal cords.

The auxiliary function of AI for ethnic vocal music learners is to serve as a voice pedagogue, defined as a teacher who has “scientific knowledge of the vocal mechanism as well as how to apply current research and valid techniques to maximize the efficiency of the voice” (Gutshall, 2006). The practical application of AI software in ethnic vocal music can effectively guide the user towards the correct vocalization method and singing rhythm. In a survey of the vocal cord structure of 376 college students, 20.75% of vocal music majors suffered laryngeal damage due to excessive voice use and poor voice practice methods, representing the largest source of laryngeal injury besides lesions. Among the symptoms of laryngeal cavity discomfort, hoarseness, throat discomfort, and air leakage during singing scored the highest proportions (Guo, 2006).

A mixed sample of 50 Chinese laryngeal cavities showed that through laryngoscope monitoring, the pronunciation frequency ranged from low to high. The dynamic laryngoscope automatic frequency measurement device showed that the frequency range was between 200 and 300HZ, and that low-frequency sounds increased the sliding degree of vocal cord mucosa (Lu, 1995). It was also determined that low-frequency sound exercises can play a role in vocal cord muscle massage.

"Glottal fry", also called "vocal fry" or "creaky voice" (Plexico, 2017), has been described as sounding similar to the "popping of corn" or a "motorboat engine" (Colton, 1996). Brought about by a low airflow rate related to the long-closed phase of the vocal fold vibratory cycle (Blomgren, 1998), and existing as a separate phonational register lying below the modal register on the frequency continuum (Hollien, 1968), it has a transparent feeling, giving the sensation of particles that are very dense but within which the gaps can clearly be distinguished. The texture is acceptable to the human ear and feels stimulating but not uncomfortable. Pulse register phonation or strohbass (Appleman, 2005) is a frequently-used exercise in ethnic vocal music that helps to rest the voice and restore singing function. It is characterized by low frequency and low decibel vibrations to make the sounds. "Opening the mouth wide" and "Grooving the tongue" simultaneously relax the entire facial musculature, while the weak airflow from the abdomen impacts the closed vocal cords, causing them to vibrate and produce a continuous bubbling sound (Guo, 2015).

The sound source tested the three processes of singing under fatigue, vocal fry recovery and high-quality singing. (See Figure 1.)

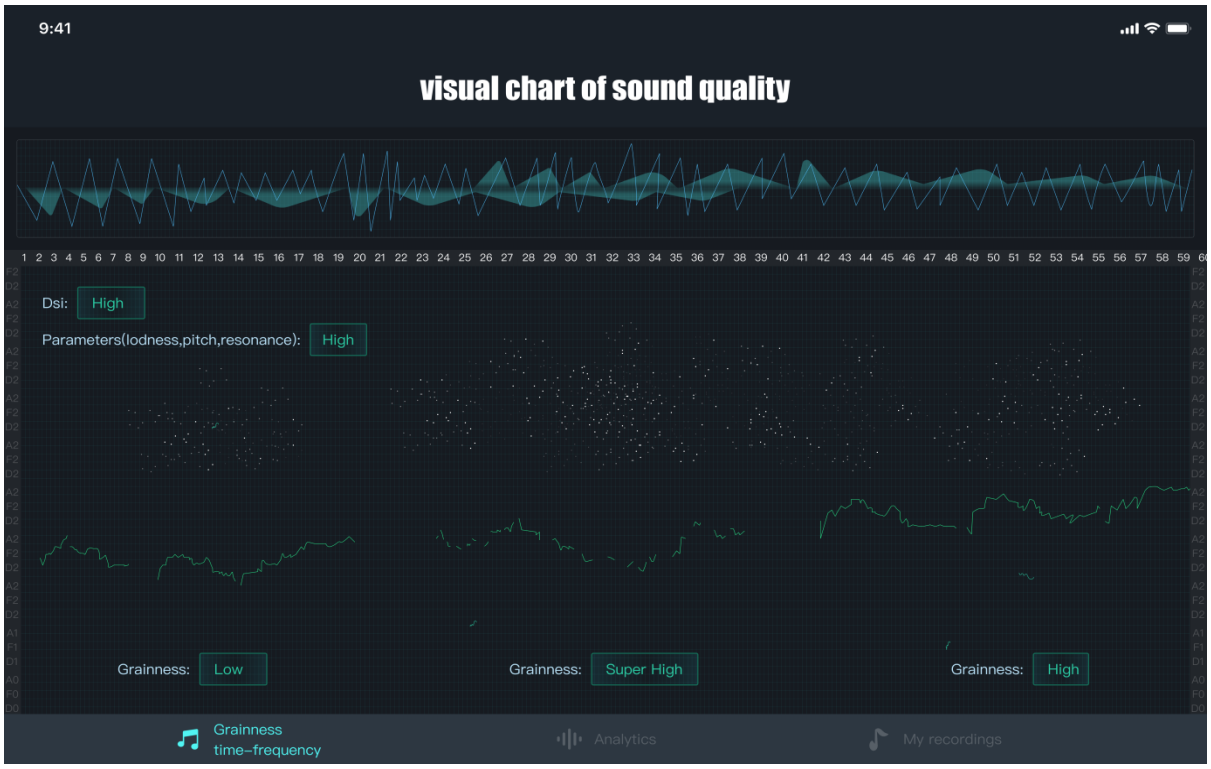

Figure 1 *Glottal fry data in AI application (Beta Version)*

Vocal core and laryngeal cavity health monitoring are shown in Figure 2.

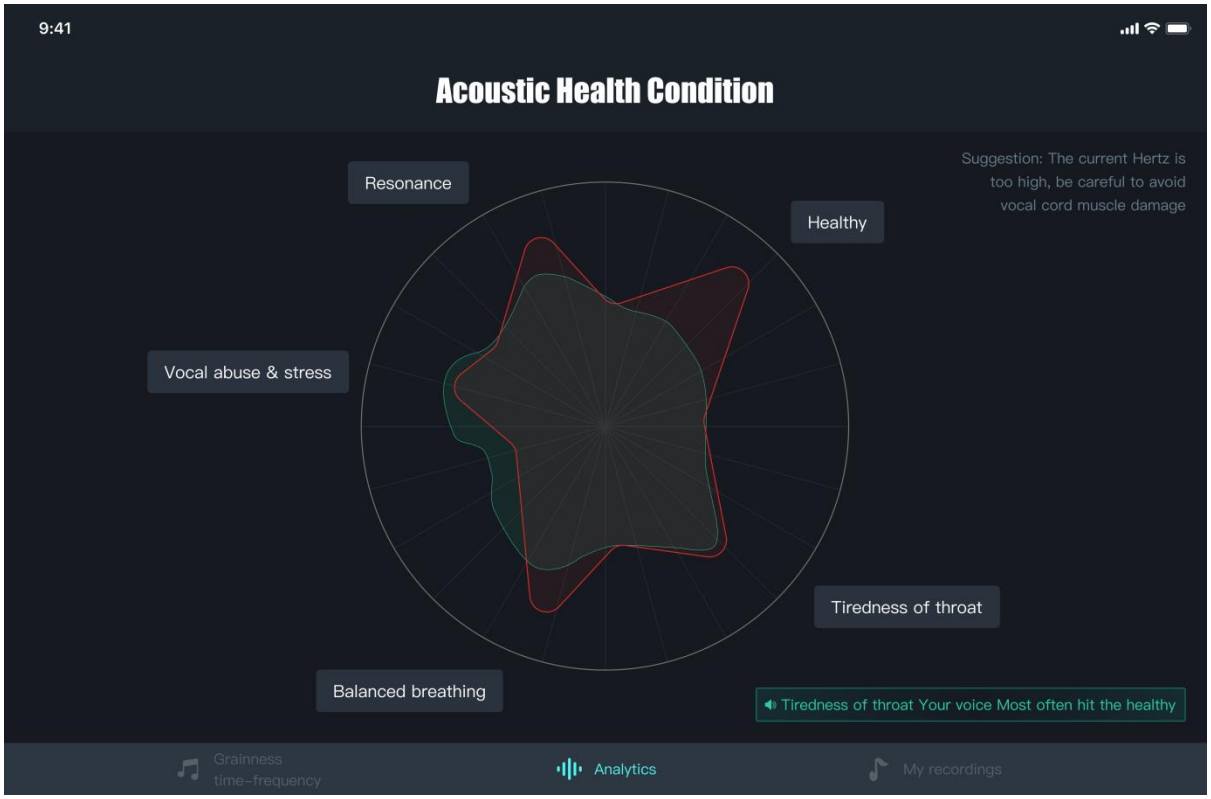

Figure 2 *Vocal cord and laryngeal cavity health monitoring (Beta Version)*

AI can analyze the hoarseness, roughness and breathiness in the voices of ethnic vocal music singers. In addition to the four standard acoustic parameters (F_0, Jitter, Shimmer, NNE), two additional acoustic parameters suitable for ethnic vocal music singing voices can be added: glottal spectrum frequency, and formant frequency perturbation (Huang, 2008). The glottal spectrum slope is a measure of the singer's voice signal in the high-frequency region and a quantity of intensity difference in the low-frequency region, thereby accurately monitoring the current voice state of the singer. AI also provides other exercises for vocal cord relaxation, such as producing the voice more naturally; warm-up and cool-down voice; and the relaxation of jaw and pharynx while producing nasal, vowel and humming sounds (Hazlett, 2011).

Audio processing software for online courses

As the output device typically employed for online teaching, it is a common experience for mobile phone users that the input voice is excessively loud. Liu et al. (2022) proposed a novel time-warping approach for pitch correction: shape-aware dynamic time warping (SADTW), which ameliorates the robustness of existing time-warping approaches, to synchronize the amateur recording with the template pitch curve. In order to improve the clarity of ethnic vocal singing on mobile devices, it is necessary to collect data on live vocals, musical instruments, and environmental sounds, and then make the necessary adjustments. Normally, when faced with mobile devices, the vocal effect will suffer a reduction in quality during the transmission process, or even be covered by current and ambient sound. A bel-canto tenor, a Chinese ethnic vocal tenor, a bel-canto soprano, and a Chinese ethnic vocal soprano were

chosen as the four subjects to test their vocalization, examine their frequency of singing, and record and examine their singing data using “AudioTool” APP as the recording tool. (See Figure 3.)

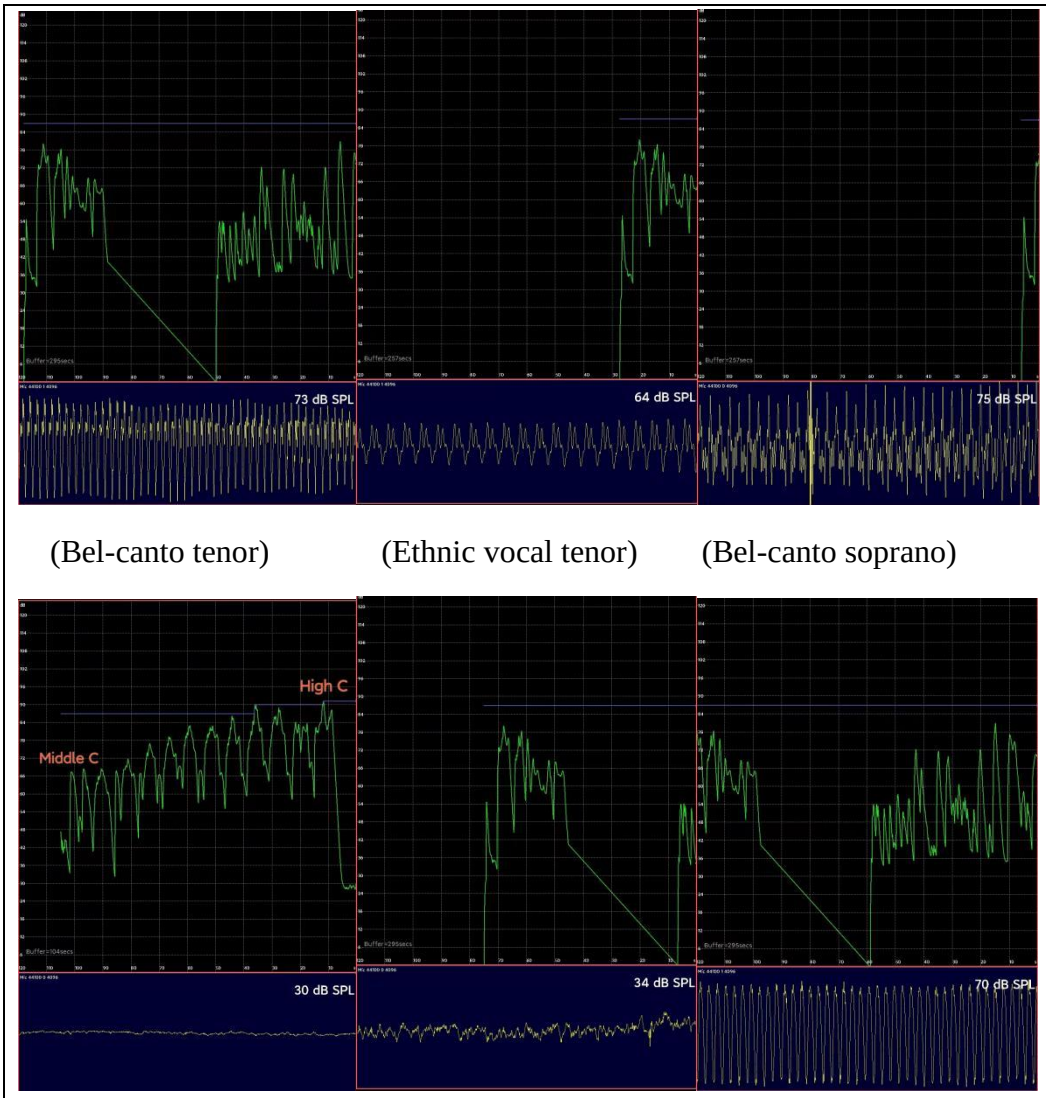

Figure 3 *Vocal records on Hertz and dB SPL*

Distinct sound data is presented in Table 1 for the easy understanding of each singer's performance. This information aids the AI in tailoring the sound-carrying capacity of mobile devices to match the singer's current situation.

Table 1 *Sound data of individual singers*

Gender	Age	Vocal Part	Pitch Name	Herz	Decibel	Result
Male	23	Bel-canto tenor	G5	479HZ	75dB	High dB SPL,Low HZ,Voice Sound output is normal
Male	23	Ethnic Vocal Tenor	G5	511HZ	64dB	Low dB SPL,High HZ,Voice Sound output disappears
Female	23	Bel-canto soprano	A5	698HZ	73dB	High dB SPL,Low HZ,Voice Sound output is normal

Female	23	Ethnic vocal soprano	A5	783HZ	70dB	Low dB SPL,High HZ,Voice Sound output disappears
		Piano	C4	440HZ	54dB	When the chord above A5 is reached, the sound becomes weaker
		Normal talk		390HZ	31dB	The sound does not disappear

As can be seen, the online teaching software must modify voice input and output in keeping with the high decibel capacity of the ethnic vocal tenor and soprano, and the piano's

sound amplitude above A5 (Chandna et al., 2020). By adjusting this fundamental setting, it is possible to guarantee that the folk song's high voice will be adequately transmitted throughout the online lesson, while also depending on AI's control and management to analyze the audio and make the necessary dynamic modifications. Such problems can be solved with a system capable of transmitting a low-bandwidth text / control stream over a network, or a similarly low-data-rate channel which facilitates the rendering of spoken or sung text at the receiving end of the transmission channel (Cook, 1993). Finally, the human voice, environmental sound and musical instrument sound data is transmitted to the device. After AI calculations, the device adjusts the input and the priority of the output sound data, so as to present the human voice completely and clearly.

The application of vocal psychology

Research has demonstrated that the importance value of vocal technology recognition increases with the level of improvement of psychological resources (Zhang, 2022). Some degree of harm can be done to the quality of a performance when pupils experience anxiety while singing, such as a rapid heartbeat, increased blood pressure, quick and shallow breathing, and dry mouth (Spahn, 2015), so teaching effectiveness needs to be optimized by ensuring that individuals are as emotionally "stable" as possible.

Information obtained via online instruction demonstrates the importance of motivation, encouragement, and the desire to study (Theiler & Lippman, 1995). Anxiety may be successfully managed within the context of the ethnic vocal performance process using slight, focused relaxation. During the 2022-2023 school year, the Shenyang Conservatory of Music's second-year graduate students used a combination of online and face-to-face engagement as part of their completion of a professional course in national voice music. The provide courses tuition in ethnic vocal music singing, including stage performance, opera, and national vocal music theory.

Under the author's guidance, 22 students convened to discuss how to address stage singing anxiety following the final national vocal performance examination. Of the participants, 17 expressed the view that, "Adaptability is crucial in mitigating anxiety related to stage performances." According to Zakaria (2013), the concept of "adaptability" not only bolsters self-confidence through an understanding of, and adjustment to, the song itself, but also extends to acclimating to the performance environment, keenly observing the audience's reactions and effectively appraising oneself psychologically, all of which ultimately contribute to the cultivation of a positive mental state.

Continued cross-modal research will undoubtedly provide further insights into the expressive aspects of vocal expression and music performance (Juslin, 2003). In the meantime, we may draw the conclusion that ethnic vocal music performance and training

should be founded upon a psychological model that carefully takes into account various factors, as shown in Figure 5.

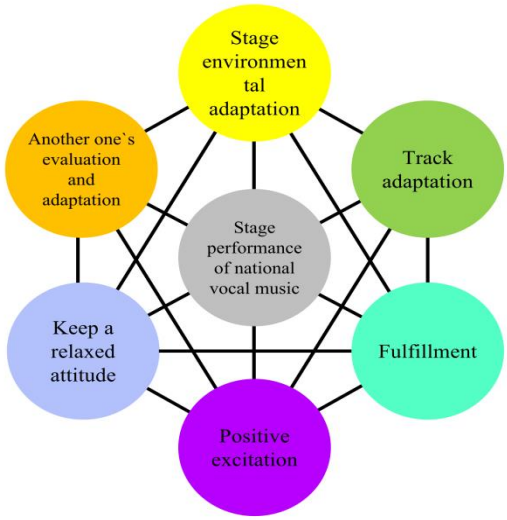

Figure 5 *Psychological model of national vocal music stage performance (Tang, XXX)*

Questionnaire investigation

Online surveys were distributed as the main data collection component of the present study, to be assessed based on predetermined criteria. While a total of 1,657 were completed, those with response times of less than 30 seconds were excluded. This led to 1,243 valid surveys being obtained, which translated to a successful recovery rate of 75%. According to the formula used to calculate individual user satisfaction, the outcome reflects a high level of satisfaction:

$$Sj = \frac{100}{1657}(\sum 10) \quad (1)$$

The basic information regarding the 1,243 survey participants, as derived from the sample's primary data, is outlined in Table 2. The participants' ages range from 18 to 35, and predominantly comprise undergraduate and Master's degree holders (81.6% of the student body), a demographic characterized by a high level of knowledge and a strong openness to new concepts. Additionally, vocal music was the primary major for the majority of students.

Table 2 *Basic information of the sample*

User information	**The content**	**Quantity**	**Percentage (%)**
Age	Over 18	915	86.5
	Over 35	168	13.5
Identity	Student	950	76.4
	Teacher	293	23.6
Record of formal schooling	Junior college or below	125	10.1
	Undergraduate	673	54.1
	Master's degree	342	27.5
	Doctoral degree	103	8.3

Major	National vocal music	261	21
	Bel-canto vocal music	245	19.7
	Popular music	269	21.6
	Music education, Instrumental music performance, Musicology, Music therapy, etc.	250	20.1
	Other major	218	17.5

According to Table 3, a significant majority of individuals engage in daily vocal practice sessions lasting between one and three hours. This data reinforces the idea that most individuals find this duration comfortable, as well as aligning with the initial advice provided by the AI regarding recommended practice times. As a result, the author felt that recognizing the frequency of daily training was of paramount importance in shaping a model for the teaching and practice of ethnic vocal singing (LeBorgne, 2002).

Table 3 *Analysis of the current situation*

Sum of daily vocal training hours	**Frequency**	**Percentage (%)**
1 to 3 hours	875	70.4

Over 3 hours	368	29.6

In order to comprehend user intentions in terms of the employment of supplementary software functionalities, and with the aim of establishing a blueprint for the development of AI applications, the authors centered their attention on the “open throat vocalization” app developed by Beijing Xihang Technology Co., Ltd. This decision garnered strong endorsement from the software users, as the results of the survey results shown in Table 4 indicate.

Table 4 *Situational analysis on the use of the “open throat vocalization” app*

		Quantity	**Percentage (%)**
User satisfaction	Extremely satisfied	843	67.8
	Very satisfied	162	13
	Satisfied	148	11.9
	So-so	28	2.3
	Dissatisfied	62	5
Favorite function	Music scale training	260	20.9
	Intonation device	147	11.8

	Following songs	196	15.8
	Solfeggio & ear training	198	15.9
	Breath practice	165	13.3
	Virtual piano	72	5.8
	Metronome	115	9.3
	Another type	90	7.2
Would you like to use the app regularly?	Yes	1067	85.8
	No	176	14.2

The statistics reveal that most users are content with the software's user-friendliness, particularly favoring the "music scale training" feature. They also indicated a strong desire to use it in the future, in so doing underscoring the potential for the refinement of vocal practice software. The valuing of the incorporation of interactive elements and an expression of willingness to integrate tools that can enhance voice training in the future are both considered positive signs (Ng et al., 2013).

From an analysis of the data presented in Table 5, it is evident that users' expectations for each AI function are relatively consistent. Nevertheless, when viewed from a broader perspective, there is a significant surge in expectations for AI development, especially with

regard to a demonstration of singing. As the most eagerly anticipated feature, this concept aligns with the ongoing endeavor to enhance human-AI interaction, particularly through advancements in human speech emotion recognition (Lawson, 2023).

Table 5 *Expected value of AI*

		Response	
	What functions do you expect from AI?	**Number of cases**	**percentage (%)**
1.	Auto accompaniment	992	18.7
2.	Course record	1095	20.6
3.	Singing evaluation	1055	19.9
4.	Suggestions for improvement	1062	20.0
5.	AI singing demonstration	1104	20.8

Teaching modes based on AI

Teaching in the fourth Industrial Revolution will aim at different ways of integrating advanced digital technologies into learning environments (Arel, 2010). Previous teaching models have been limited by time and space, making it difficult to meet the needs of different

personalities. Using modern online educational technology to establish an efficient, fast and user-friendly application system has become an essential trend (Ma, 2021) as the technical basis of the “digital campus”. This study adopts the perspective of AI and information technology within the online teaching environment to subdivide and expand the tasks of the application scenarios in the teaching and personal learning of ethnic vocal music within universities. When intelligent technology is applied to various tasks, data consumers, in the form of students, teachers, and network platforms, can make more comprehensive and efficient use of data information, improve the quality of classroom teaching, and cultivate students’ music literacy.

In the field of music, the AI processing tasks are used mainly to identify and generate the logic and theory of music creation through the objective modeling of data and algorithm in the three main parts of music (Liang, 2022), performance (Li, 2021) and sound (Chen et al., 2019). In the acoustic model, the modeling needs to obtain fundamental frequency information, such as timbre (Pons et al., 2017), speaker information (Xue et al., 2021), pause information (Martin et al., 2022), language information (Schulze-Forster et al., 2021), facial features (Konstantinidis et al., 2018), breath control (Fukuda et al., 2018), sentiment analysis (Abburi et al., 2016), and other features. For different styles of music, the algorithm needs to analyze the composition and variation of the chords, rhythms, and instrumentation. Different from ordinary speech processing, deep learning not only needs to ensure sound quality

(Bochner et al., 2022), but also needs to pay attention to sampling rate (Gao et al., 2022), as well as continuity and fluency (Huang et al., 2022).

In the imitation and creation of songs, the characteristics of lyrics and melody can be simulated by a large number of data and models, but the weak relationship between them requires the strict monotony of tone and word pairing and connection (one-to-one to many-to-many), which in turn requires the quality assurance of data and the strengthening of prior knowledge. For example, if the tone of the lyrics is not consistent with the melody, the problem of homophone confusion will occur. The ability of models to align too closely does not create more diverse artistic music, especially in folk music (Savage et al., 2022). In fact, when it comes to expressing the emotion of the music more vividly and strongly, the rhythm of the lyrics and the melody are often exaggerated and over-dramatized. The melody corresponding to important lyrics will be enhanced by emotion (BAO, 2021), so emotion analysis needs to go beyond lyrics to incorporate melody and other factors. In the future, AI will cooperate with human beings to complete music in the ways outlined above, rather than replace their roles as artists (Feldman, 2017). As things currently stand, AI is widely used in aesthetic music applications. If AI music creation is also to be regarded as an information tool for producers, there are large-scale AI applications that can serve to promote the dissemination of musical works and enhance audience experiences. For example, the music recommendation system, as the mainstream personalized technology for audiences, can

expand the number of music platforms and enhance the experiences of returning users. The combination of high quality and popular music platforms, in conjunction with the music recommendation system technology, requires an effective combination of audio characteristics, user interaction, user differences and commonalities, and user and platform annotation under music works (Liu & Liu, 2020). As a media for digital music communication, music platforms are more inclined to expand users' listening recommendations, thereby avoiding the "echo chamber" problem. At the same time, they also need to pay attention to copyright issues in the process of music transmission. In personal works, such as NFT digital products, privacy, security and other features need to be guaranteed by blockchain technology (Zhao & O'Mahony, 2018). For the audience, AI also requires research approaches that are multi-dimensional and multi-field.

Online education, as a teaching platform for universities in the information age, is typically composed of cloud classrooms and electronic resources; as such, it has become something of a mainstream teaching methodology, especially as a result of the pandemic. As the basis of online education, the network communication environment, sited between and where the devices of teachers and students are located, is crucial to ensuring the quality of cloud-based classroom teaching and the usage experiences of teachers and students. The cloud-based classroom should ensure a low-delay, highly reliable network communication environment (Bojović et al., 2020), as well as clear, three-dimensional conditions for high

income recording with the removal of environmental noise. Even allowing for the establishment of a stable, cloud-based classroom, the cost and time limitations of university equipment, as well as the variations among students' equipment, should be taken into account in order to enhance specific user experiences (McRoy., 2020). Consideration also needs to be given to support students' ability to listen to lectures online in real time, to independently and flexibly select past courses, exchange discussions, and take online tests. At the same time, when applied to vocal music teaching, the cloud-based classroom needs to consider more diversified and three-dimensional forms of screen courseware that can effectively present the teaching content, as well as students' compositions, and the playing of musical instruments and singing (Fang, 2021).

A combination of AI and the digital curriculum system within the training cycle, with a musical instrument in the form of a harp from a specific teaching institution (http://s.immusician.com/static/web/index.html), is given here as an example. Its AI technology is mainly embodied in interactive video teaching based on voice and video recognition technology and customized personalized courses based on big data analysis, and the intelligent adjustment of learning paths based on data collected from users' actual practice. Students master the basic techniques of playing and singing in the early course; from the intermediate to the senior course, students can play and sing 80% of popular songs on the market without any difficulties. In the advanced course, students can master the principles of

songwriting and try to create songs with the aid of voice and video recognition technology (called “audio-visual speech recognition”). In this form of deep learning, the recorded video and voice are used as the input data of the model for multi-mode feature fusion, extraction and coding. The construction of the algorithm model to achieve various specific tasks is shown in Figure 6 below.

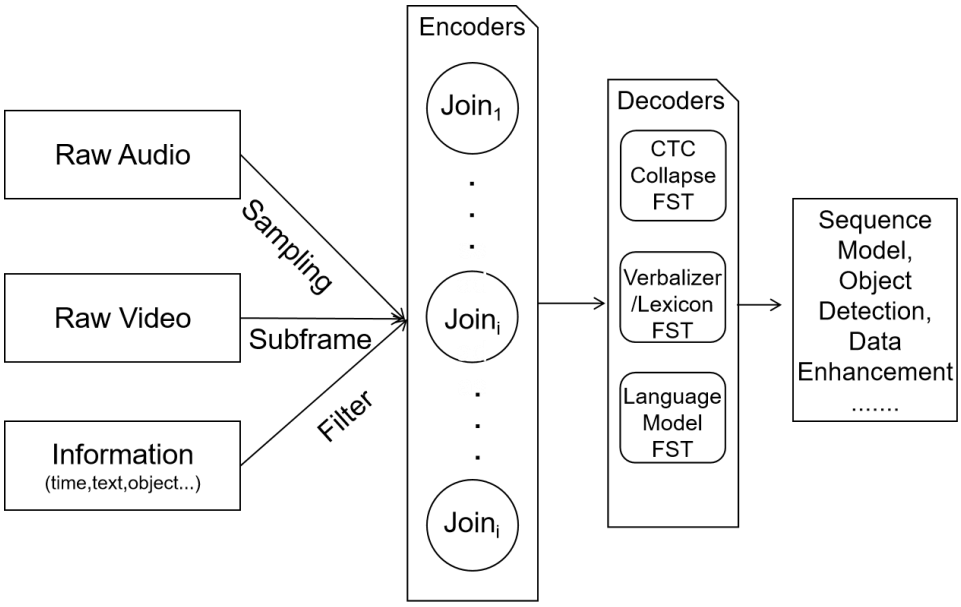

Figure 6 *Algorithm model*

In the university online ethnic vocal singing teaching model, AI can be subdivided into the following tasks, all of which complement each other:

1. According to the acquisition of students’ physiological signals (voice, ECG, eye movement, facial expressions, and body changes) in the process of learning and practicing, information feedback, evaluation, advice and emotional analysis can be carried out. In

addition, the relationship between a student's personality and music can also be explored. Combined with the capture and analysis of physiological signals, multimodal fusion technology is used to perform representation, fusion, conversion, and alignment of cross-modal data, the similarity, dissimilarity, freshness and other indicators being used to score, provide feedback, and evaluate students' singing and performance through combining the existing scores with the evaluation data. To further stimulate students' imagination, the sound of musical instruments and human voices are mapped into images in a visual way, i.e. the audio space is transformed into a visual space, thereby assisting the teaching process (Khulusi et al., 2020), as shown in Figure 7 below. Optical music recognition (OMR), which enables the detection of staff notation and the visualization of various musical elements (Calvo-Zaragoza, & Rizo, 2018), is one example. Audio-visual separation is used to distinguish the targets of different sound sources, while the detection and classification of acoustic scenes are closely related to multi-channel signals, e.g., microphone array multi-channel signals, conversion methods, e.g., Fourier transform and wavelet transform, and audio scaling. These processing elements correspond to the heterogeneous distribution and discretization of data features; group voice processing between teacher and student, or in a chorus, corresponds to speaker recognition (SR) in the speech domain (Li et al., 2022).

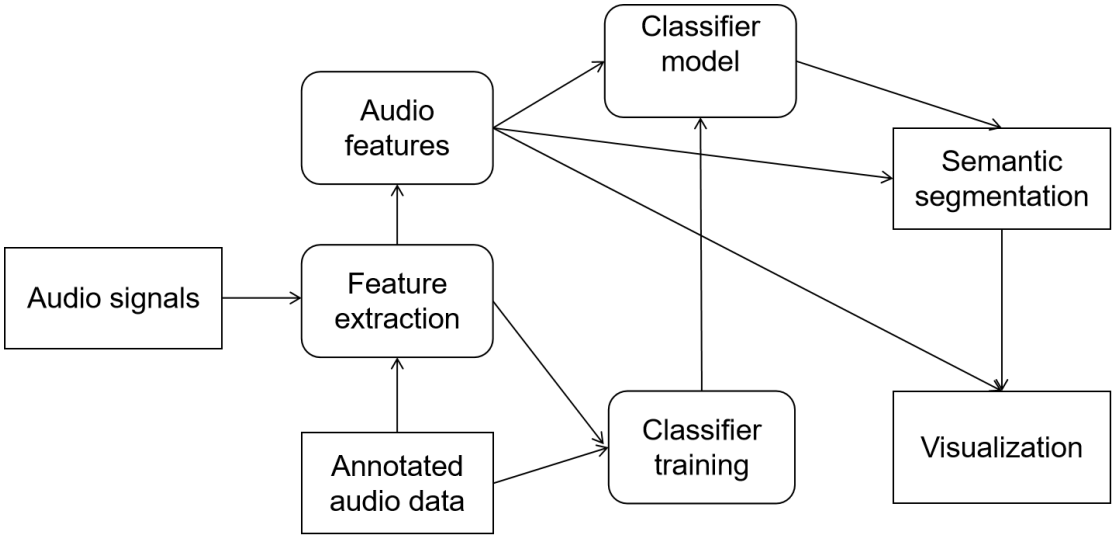

Figure 7 *Operation diagram*

2. Enhancing the experience of using multimedia intelligence tools, such as intelligent musical instruments and virtual musical teaching aids, the sound quality of online classes, and the immersion experience. The targeted, combined optimization of platform tools, such as audio and video recording, privacy processing, blackboard writing, and teaching content thought maps, are divided amongst teachers and students in line with the recommendation system. For the acquisition and overview of mainstream aesthetic works and information, “web crawler” and dynamic theme models can be used. Real-time and recorded data of teachers and students, along with distributed storage and privacy processing of computing using deep learning, together correspond to dynamic distributed messaging systems, blockchain and federated learning.

3. Exploring effective methods and strategies to improve students’ comprehension

quality and cultivate innovative consciousness in the development of electronic music creation and performance training and promoting the reform of electronic music online education. Teacher and student data can be used to judge learning models, and match professional knowledge, teaching content, progress and rhythm, and course and students' learning effect evaluations. For example, mapping the radiating function of wireless networks into online education and establishing the intimacy and immediacy of social learning systems can improve the effectiveness of the AI electronic music creation learning process. Scholars have analyzed representative online learning platforms, highlighting the problems and trying to use the structure of personalized learning systems and online learning content – constructed by using the improved optimization strategy, and automatically arranging sequences through a data-driven recommendation system, respectively – to build models.

4. The imitation and generation of musical scores and playing skills corresponds to the tasks in natural language processing (NLP), and assists students with opening their voices (performance preparation) and singing (performance). The generation of sounds and instrumental tones corresponds to speech understanding and speech synthesis in the field of speech, while generating a song according to one's own lyrics or melody can correspond to transfer learning and continual learning in the field of deep learning. It is also possible to generate models from text by giving arbitrary formats and templates in order to generate content that fits.

When a deep learning algorithm is applied to a specific scene, one of the necessary and particularly important problems is how to evaluate the effect of the model results. Subjective and objective perspectives should be combined to conduct a comprehensive evaluation, the aim being to make the effect of the algorithm more sustainable and practical. The judgement of musical works may be highly subjective, so the quality assessment mechanism should always guide the music processing direction and ability of the algorithm.

Teaching model

Among the methods of teaching vocal music, that of ethnic vocal music is characterized by being effective, intuitive, operational, open, and targeted. Focusing on the emotional, behavioral, and cognitive modes helps pupils learn based on their unique abilities (Kashina et al., 2020). The teaching model shown in Figure 8 may be constructed with the intention of assisting with the performance and teaching of ethnic vocal music (Wang, 2022); as such, it uses the performance techniques, instructional strategies, and artificial intelligence algorithm to collectively support this study's vocal teaching model.

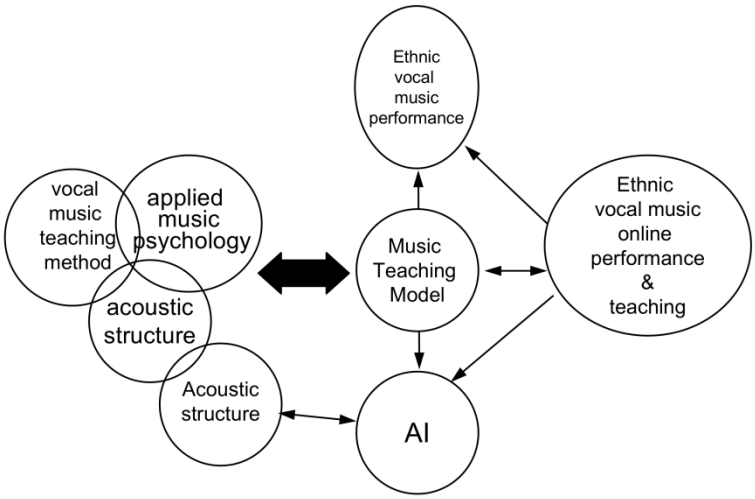

Figure 8 *Teaching model*

Table 6 summarizes the teaching mode of this article, As shown in the figure, the online performance and education of ethnic vocal music is based on other disciplines. At the same time, AI analyzes learning and teaching methods suitable for distance teaching based on data, in the process forming a complete teaching model (Su, 2022). Combined with the previous analysis, the WSQ model employs the observation method as a means of AI's fundamental assessment of vocal learners. Additionally, it utilizes the PA approach to facilitate a comparison between two software objects that do not occupy the same space (Body, 2021). This comparative evaluation yields valuable data, which in turn aids with individual assessment.

Table 6 *Model information summary*

The content of this teaching model	**Approach**	**Significance**
National vocal performance	Liu Hui “twenty-four character formula”	Plays a normative role in singing
Teaching methods of national vocal music	Emotional, behavioral and cognitive modes	Plays a normative role in teaching
Artificial Intelligence	User feature capture, course record, vocal music demonstration, scene	Assists the singing and teaching of ethnic vocal music in colleges and universities through digital algorithms

	analysis, etc	
Vocal cord maintenance	Tips Correct breathing	Protects students' voices
Applied music psychology	Psychological self adjustment	Spiritually guarantees performance and teaching
Acoustic structure	Adjust audio decibel and frequency	Guarantees normal online teaching

---

It can be inferred from the study presented in this article that ethnic vocal music and teaching are interconnected and that AI algorithms may be a useful tool for classroom instruction. The inclusion of different viewpoints and the feedback received from the study as it relates to its practice suggests that the teaching model is suitable for folk vocal singing and

instruction.

References

Abburi, H., Akkireddy, E. S. A., Gangashetti, S., & Mamidi, R. (2016, January). Multimodal sentiment analysis of telugu songs. *In SAAIP@ IJCAI.*

Appleman, R., & Bunch, M. (2005). Application of vocal fry to the training of singers. *J Sing*, *62*(1), 53-9.

Arel, I., Rose, D. C., & Karnowski, T. P. (2010). Deep machine learning - a new frontier in artificial intelligence research [research frontier]. *IEEE Computational Intelligence Magazine, 5*(4), 13-18.

Bao, C. *Generating music with sentiments[J]*. 2021.

Blomgren, M., Chen, Y., Ng, M. L., & Gilbert, H. R. (1998). Acoustic, aerodynamic, physiologic, and perceptual properties of modal and vocal fry registers. *The Journal of the Acoustical Society of America*, *103*(5), 2649-2658.

Bochner, J., Indelicato, M., & Konnur, P. (2022). Effects of Sound Quality on the Accuracy of Telephone Captions Produced by Automatic Speech Recognition: A Preliminary Investigation. *American Journal of Audiology*, 1-8.

Body, M. (2021). Flipped Classroom-Perfect Mix of Face-to-face and Online Learning. *Scripta*, 15.

Bojović, Ž., Bojović, P. D., Vujošević, D., & Šuh, J. (2020). Education in times of crisis:

Rapid transition to distance learning. *Computer Applications in Engineering Education*, *28*(6), 1467-1489.

Calvo-Zaragoza, J., & Rizo, D. (2018). Camera-PrIMuS: Neural End-to-End Optical Music Recognition on Realistic Monophonic Scores. *International Symposium/Conference on Music Information Retrieval.*

Chandna, P., Blaauw, M., Bonada, J., & Gómez, E. (2020, May). Content based singing voice extraction from a musical mixture. *In ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing* (ICASSP) (pp. 781-785). IEEE.

Chartrand, R. (2016). Advantages and disadvantages of using mobile devices in a university language classroom. *Bulletin of the Institute of Foreign Language Education Kurume University*, *23*(1), 1-13.

Chen, X., Chu, W., Guo, J., & Xu, N. (2019, March). Singing voice conversion with non-parallel data. *In 2019 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)* (pp. 292-296). IEEE.

Civit, M., Civit-Masot, J., Cuadrado, F., & Escalona, M. J. (2022). A systematic review of artificial intelligence-based music generation: Scope, applications, and future trends. *Expert Systems with Applications*, 118190.

Colton, R. H., & Casper, J. K. (2001). Understanding Voice Problems: Voice Problems

Associated with the Pediatric and the Geriatric Voice.

Cook, P. R. (1993). SPASM, a real-time vocal tract physical model controller; and singer, the companion software synthesis system. *Computer Music Journal*, *17*(1), 30-44.

Elghaish, F., Chauhan, J. K., Matarneh, S., Rahimian, F. P., & Hosseini, M. R. (2022). Artificial intelligence-based voice assistant for BIM data management. *Automation in Construction*, *140*, 104320.

Fang, P. E. N. G. (2021). Optimization of music teaching in colleges and universities based on multimedia technology. *Advances in Educational Technology and Psychology*, *5*(5), 47-57.

Feldman, S. S. (2017). Co-creation: human and AI collaboration in creative expression. *Electronic Visualisation and the Arts* (EVA 2017).

Fu, S., Theodoros, D. G., & Ward, E. C. (2015). Delivery of intensive voice therapy for vocal fold nodules via telepractice: a pilot feasibility and efficacy study. *Journal of Voice*, *29*(6), 696-706.

Fukuda, T., Ichikawa, O., & Nishimura, M. (2018). Detecting breathing sounds in realistic Japanese telephone conversations and its application to automatic speech recognition. *Speech Communication*, *98*, 95-103.

Gao, Z., Tang, M., Li, A., & Chen, Y. (2022, April). An Audio Frequency Unfolding Framework for Ultra-Low Sampling Rate Sensors. *In 2022 23rd International*

Symposium on Quality Electronic Design (ISQED) (pp. 1-6). IEEE.

Guo, L., Zhang, L., & Li, Y. (2006). Analysis of 376 cases of voice diseases among vocal music students. *Music Exploration (Journal of Sichuan Conservatory of Music)*, *01*, 86-88.

Guo, X. (2015). The important role of bubble sounds in vocal singing. *Northern Music*, *35*(17), 64.

Gutshall, E. M. (2006). The twenty-first century vocal pedagogue: A mediator and translator for the vocal studio and its interdisciplinary colleagues. *ProQuest*.

Hazlett, D. E., Duffy, O. M., & Moorhead, S. A. (2011). Review of the impact of voice training on the vocal quality of professional voice users: implications for vocal health and recommendations for further research. *Journal of Voice*, *25*(2), 181-191.

Hebert, D. G. (2007). Five challenges and solutions in online music teacher education. *Research and Issues in Music Education*, 5(1), 1-10.

Henrich, N., Smith, J., & Wolfe, J. (2011). Vocal tract resonances in singing: Strategies used by sopranos, altos, tenors, and baritones. *The Journal of the Acoustical Society of America*, *129*(2), 1024-1035.

Hollien, H., & Michel, J. F. (1968). Vocal fry as a phonational register. *Journal of Speech and Hearing Research*, *11*(3), 600-604.

Hsia, L. H., Hwang, G. J., & Lin, C. J. (2022). A WSQ-based flipped learning approach to

improving students' dance performance through reflection and effort promotion. *Interactive Learning Environments*, *30*(2), 229-244.

Huang, R., Cui, C., Chen, F., Ren, Y., Liu, J., Zhao, Z., & Wang, Z. (2022, October). Singgan: Generative adversarial network for high-fidelity singing voice generation. *In Proceedings of the 30th ACM International Conference on Multimedia* (pp. 2525-2535).

Huang, Z. M., & Wan, P. (2008). Research on the correlation between voice acoustic parameters and voice quality. *Journal of Clinical Otolaryngology-Head and Neck Surgery*, *2008*(06), 251-255.

Hwang, Y. S., Shim, M. R., Kim, C. J., Choi, Y. S., Kim, S. Y., Choi, C. H., & Sun, D. I. (2016). Comparison of Vocal Cord Motion and Voice Characteristics of Applied Music Singing Students before and after Singing Voice Therapy. *Journal of The Korean Society of Laryngology, Phoniatrics and Logopedics, 27*(2)*,* 114-121.

Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological bulletin*, *129*(5), 770.

Kashina, N. I., Pavlov, D. H., & Li, H. (2020, May). Music-digital technologies as a means of developing creative independence of students of colleges and higher schools in China and Russia. *In International Scientific conference "digitalization of education:*

History, trends and prospects" (DETP 2020) (pp. 27-31). Atlantis Press.

Khulusi, R., Kusnick, J., Meinecke, C., Gillmann, C., Focht, J., & Jänicke, S. (2020, September). *A survey on visualizations for musical data. In Computer Graphics Forum* (Vol. 39, No. 6, pp. 82-110)

Konstantinidis, D., Dimitropoulos, K., & Daras, P. (2018, October). A deep learning approach for analyzing video and skeletal features in sign language recognition. *In 2018 IEEE international conference on imaging systems and techniques (IST)* (pp. 1-6). IEEE.

Lawson, F. R. S. (2023). Why can't Siri sing? Cultural narratives that constrain female singing voices in AI. *Humanities and Social Sciences Communications*, *10*(1), 1-11.

LeBorgne, W. D., & Weinrich, B. D. (2002). Phonetogram changes for trained singers over a nine-month period of vocal training. *Journal of Voice*, *16*(1), 37-43.

Lei, D. (2020, October). RETRACTED: Research on Network Teaching of Music Major Based on Artificial Intelligence Technology. *In Journal of Physics: Conference Series* (Vol. 1648, No. 4, p. 042094). IOP Publishing.

Li, H. (2021). Piano education of children using musical instrument recognition and deep learning technologies under the educational psychology. *Frontiers in Psychology*, 12.

Li, X., Liu, S., & Shan, Y. (2022). A Hierarchical Speaker Representation Framework for One-shot Singing Voice Conversion. *arXiv preprint arXiv*:2206.13762.

Liang, M. (2022). Music Score Recognition and Composition Application Based on Deep

Learning. *Mathematical Problems in Engineering*, 2022.

Liu, C., Wan, P., Tu, Y. F., Chen, K., & Wang, Y. (2021). A WSQ-based mobile peer assessment approach to enhancing university students' vocal music skills and learning perceptions. *Australasian Journal of Educational Technology*, *37*(6), 1-17.

Liu, H. (2011). Thoughts on the relationship between sounds and words in singing -24-character formula for the relationship between sounds and words in Chinese national vocal music singing teaching. *Songs of the Music Bureau (Journal of Shenyang Conservatory of Music), 04*, 9-10.

Liu, H. (2018). Inheriting beauty, expressing beauty, creating beauty, spreading beauty, and leading beauty: Thoughts on the development and construction of Chinese vocal music. *Chinese Music*, (01),7-14+193. doi:10.13812/j.cnki.cn11-1379/j.2018.01.002.

Liu, J., Li, C., Ren, Y., Zhu, Z., & Zhao, Z. (2022). Learning the beauty in songs: Neural singing voice beautifier. *arXiv preprint arXiv:*2202.13277.

Liu, N. (2022). Study on the Application of Improved Audio Recognition Technology Based on Deep Learning in Vocal Music Teaching. *Mathematical Problems in Engineering*, 2022.

Liu, S., & Liu, C. (2020). A review of music recommendation systems. *Journal of Guangzhou University: Natural Science Edition*.19(05):36-46+77.

Liu, Z. Y. (2022). My opinion on the improvement of national vocal music singing through

musical aesthetics. *Cultural Monthly*, *09*, 163-165.

Lu, Y. Y., Wang, S. Q., & Ling, L. (1995). Study on the relationship between vocalization at different frequencies and vocal cord mucosal waves under dynamic laryngoscopy. *Zhejiang Journal of Integrated Traditional Chinese and Western Medicine*, *1995*(02), 34-35.

Luo, Y. J., Lin, Y. J., & Su, L. (2020). Toward expressive singing voice correction: On perceptual validity of evaluation metrics for vocal melody extraction. *arXiv preprint arXiv*:2010.12196.

Ma, J., & Feng, B. (2021). Integrated design of graduate education information system of universities in digital campus environment. *Hindawi Limited*.

Martin, V. P., Arnaud, B., Rouas, J. L., & Philip, P. (2022). Automatic Estimation of Pauses Duration and Location in Read Speech as Biomarkers of Sleepiness. *Submitted to ICASSP*, 2021.

McRoy, C., Patel, L., Gaddam, D. S., Rothenberg, S., Herring, A., Hamm, J., & Awan, O. (2020). Radiology education in the time of COVID-19: a novel distance learning workstation experience for residents. *Academic radiology*, *27*(10), 1467-1474.

Ng, S. C., Lui, A. K., & Lo, W. S. (2013). An interactive mobile application for learning music effectively. In Knowledge Sharing through Technology: 8th International Conference on Information and Communication Technology in Teaching and

Learning, ICT 2013, Hong Kong, China, July 10-11, 2013, Revised Selected Papers 8 (pp. 148-157). *Springer Berlin Heidelberg*.

Plexico, L. W., & Sandage, M. J. (2017). Influence of glottal fry on acoustic voice assessment: A preliminary study. *Journal of Voice*, *31*(3), 378-413.

Pons, J., Slizovskaia, O., Gong, R., Gómez, E., & Serra, X. (2017, August). Timbre analysis of music audio signals with convolutional neural networks. *In 2017 25th European Signal Processing Conference (EUSIPCO)* (pp. 2744-2748). IEEE.

Roads, C. (1985). Research in music and artificial intelligence. *ACM Computing Surveys (CSUR)*, *17*(2), 163-190.

Savage, P. E., Passmore, S., Chiba, G., Currie, T. E., Suzuki, H., & Atkinson, Q. D. (2022). Sequence alignment of folk song melodies reveals cross-cultural regularities of musical evolution. *Current Biology*, *32*(6), 1395-1402.

Schulze-Forster, K., Doire, C. S., Richard, G., & Badeau, R. (2021). Phoneme level lyrics alignment and text-informed singing voice separation. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, *29*, 2382-2395.

Spahn, C. (2015). Treatment and prevention of music performance anxiety. *Progress in brain research*, *217*, 129-140.

Su, L. (2022, April). Analysis on the Application of Modern Multimedia Devices in the Current College Vocal Music Guiding from the Long Distance Signal Communication

Perspective. *In 2022 International Conference on Sustainable Computing and Data Communication Systems* (ICSCDS) (pp. 792-796). IEEE.

Theiler, A. M., & Lippman, L. G. (1995). Effects of mental practice and modeling on guitar and vocal performance. *The Journal of General Psychology*, *122*(4), 329-343.

Valle, C., Andrade, H., Palma, M., & Hefferen, J. (2016). Applications of peer assessment and self-assessment in music. *Music Educators Journal*, *102*(4), 41-49.

Valley, P. M. (2010). Generation Y and Vocal Fry.

Wang, X. (2022). Design of Vocal Music Teaching System Platform for Music Majors Based on Artificial Intelligence[J]. *Wireless Communications and Mobile Computing*, 2022.doi:https://doi.org/10.1155/2022/5503834

Webster, P. R. (2017). Computer-based technology and music teaching and learning. *In Critical essays in music education* (pp. 321-344). Routledge.

Wei, J., Karuppiah, M., & Prathik, A. (2022). College music education and teaching based on AI techniques. *Computers and Electrical Engineering*, *100*, 107851.

Xue, H., Yang, S., Lei, Y., Xie, L., & Li, X. (2021, January). Learn2sing: Target speaker singing voice synthesis by learning from a singing teacher. *In 2021 IEEE Spoken Language Technology Workshop (SLT)* (pp. 522-529). IEEE.

Yan, H. (2022). Design of online music education system based on artificial intelligence and multiuser detection algorithm. *Computational Intelligence and Neuroscience*, 2022.

Yu, X., Ma, N., Zheng L., et al. (2023). Developments and applications of artificial intelligence in music education[J]. *Technologie*s, *11*(2), 42.

Zakaria, J. B., Musib, H. B., & Shariff, S. M. (2013). Overcoming performance anxiety among music undergraduates. *Procedia-social and behavioral sciences*, *90*, 226-234.

Zhang, Y. (2022). Cultivation and interpretation of students' psychological quality: Vocal psychological model. *Frontiers in public health*, *10*, 966628.

Zhang, Y. (2023). Optimizing Ethnic Vocal Performance Using Artificial Intelligence Technology.

Zhao, S., & O'Mahony, D. (2018, December). Bmcprotector: A blockchain and smart contract based application for music copyright protection. In *Proceedings of the 2018 International Conference on Blockchain Technology and Application* (pp. 1-5).

Zhe, T. (2021). Research on the model of music sight-singing guidance system based on artificial intelligence. *Complexity*, 2021, 1-11.