

Optimization of Vocal Teaching Empowered by AI A Comprehensive Study on Quantitative Evaluation and Personalized Guidance System for Singing Skills Based on Voiceprint Feature Extraction

Yuzi Wei¹

¹Vocal Music Teaching and Singing, Xi'an Shiyu University, Xi'an 710065, Shaanxi, China

Corresponding author: Yuzi Wei (e-mail: yzwokay@163.com)

ABSTRACT The integration of artificial intelligence into vocal education provides new opportunities for overcoming limitations associated with subjective assessment and insufficient personalized guidance. This study proposes an AI-enabled vocal teaching framework based on voiceprint feature extraction and intelligent evaluation. Acoustic features including pitch, timbre, rhythm, and spectral characteristics are extracted through a multi-stage signal-processing pipeline, while machine-learning models are employed to establish quantitative singing-skill evaluation mechanisms. Based on diagnostic results, a personalized guidance module generates adaptive training recommendations for individual learners. Experimental analysis demonstrates that the proposed framework effectively improves evaluation consistency, enhances training efficiency, and supports individualized skill development. The study contributes to intelligent vocal education and provides methodological references for speech signal processing, acoustic pattern recognition, and intelligent human-computer interaction systems.

INDEX TERMS voiceprint feature extraction, vocal teaching, speech signal processing, acoustic pattern recognition, intelligent guidance

I. INTRODUCTION

A. RESEARCH BACKGROUND

THE report of the 20th National Congress of the Communist Party of China clearly proposes to “implement the strategy of rejuvenating the country through science and education, strengthen talent support for modernization construction”, and emphasizes the need to promote the digitization of education and the sharing of high-quality educational resources. Vocal education, as a core component of art education, undertakes the important mission of cultivating music talents and inheriting excellent culture, much as the structural logic of high-performance technical production preserves the precision of industrial craftsmanship while enhancing the aesthetic literacy of the nation. This synergy between artistic expression and the interwoven complexity of advanced material engineering provides a multidimensional framework for fostering technical excellence and cultural transmission in the modern era. Traditional vocal teaching mainly relies on “one-on-one” mentorship, relying on teachers’ subjective experience to guide singing skills

and evaluate their effectiveness. Although this model can achieve a certain degree of personalized teaching, it has obvious limitations: firstly, the evaluation is subjective, and teachers’ evaluation results are easily influenced by their own experience, aesthetic preferences, emotional states, and other factors, making it difficult to achieve objective quantitative evaluation of singing skills; Secondly, the coverage of personalized guidance is narrow, and high-quality teaching resources are scarce, making it difficult to meet the personalized needs of large-scale vocal learners; Thirdly, the teaching efficiency is low, and problems such as learners’ pitch deviation, unstable rhythm, and insufficient breath control are difficult to identify in a timely and accurate manner, resulting in weak training targeting; Fourthly, there is a lag in learning feedback. In traditional teaching, learners need to wait for the teacher’s post class comments to receive feedback, making it difficult to adjust their skills in real time during the singing process [1-3].

II. CONSTRUCTION OF A TECHNICAL FRAMEWORK FOR AI EMPOWERED VOCAL TEACHING

A. TECHNICAL FRAMEWORK CONSTRUCTION OBJECTIVES AND PRINCIPLES

The core goal of building a technical framework for AI enabled vocal teaching is to achieve precise quantitative evaluation and personalized guidance of singing skills, including:

1. Accurate extraction of voiceprint features: Achieve multi-dimensional voiceprint feature collection and precise extraction of vocal singing audio, capturing key parameters related to singing techniques;
2. Easy to use and stable operation of the system: Build a simple and user-friendly interface to ensure stable operation of the system in different devices and network environments;
3. Personalized guidance plan generation: Based on the evaluation results and learner characteristics, automatically generate targeted training content, methods, and schedule arrangements;
4. Real time feedback and dynamic optimization: Implement real-time data collection and analysis during the singing process, provide timely feedback on singing issues, and dynamically optimize guidance plans [4-6].

B. OVERALL ARCHITECTURE OF TECHNICAL FRAMEWORK

Based on the construction goals and principles, combined with the professional needs of vocal teaching and the application characteristics of AI technology, this article constructs a four layer technical framework of “voiceprint feature collection layer - feature extraction and optimization layer - singing technique quantitative evaluation layer - personalized guidance generation layer”. This framework is data-driven and supported by technology, achieving full process intelligence from singing data collection to personalized guidance generation. Each level is interrelated and mutually supportive, forming an organic whole.

C. FIRST LAYER: VOICEPRINT FEATURE COLLECTION LAYER

The voiceprint feature acquisition layer is the foundation of the technical framework, and its core function is to collect vocal learners' singing audio data, providing raw data support for subsequent feature extraction and evaluation. The core elements of the collection layer include collection devices, collection scenarios, collection parameters, and data preprocessing. The specific implementation path is as follows:

1. Adaptation of collection devices: Supports multiple collection devices, including professional microphones, smartphones, tablets, etc., to meet the collection needs in different scenarios. For professional teaching scenarios, it is recommended to use professional condenser microphones to ensure high-quality audio capture; For daily practice scenarios, support for portable devices such as smartphones to enhance ease of use.
2. Collection parameter settings: Unified collection parameter standards, with a sampling rate set to 44.1kHz (the

standard sampling rate for vocal audio collection), a bit depth of 16 bits, and a mono channel to ensure consistency and comparability of collected data [7].

Establish a hardware feature mapping library: For mainstream professional microphones (e.g., Blue Yeti) and smartphone models (e.g., iPhone 14, Huawei Mate 60), collect the same singing audio, extract frequency response curves, and construct a “hardware type-frequency compensation parameter” mapping library covering 20 common devices.

Real-time calibration process: When the system collects audio, it first identifies the hardware type (via device interface or manual selection by the user), calls the corresponding compensation parameters, and performs frequency response calibration on the audio collected by the smartphone microphone (focusing on compensating the 1kHz-8kHz high-frequency band, which is crucial for voiceprint feature extraction). This ensures that audio collected by different hardware tends to be consistent in frequency distribution.

Calibration effectiveness verification: Comparative experiments show that for the same singer recording the same song with a smartphone and a professional microphone, the similarity of core voiceprint features (MFCC, fundamental frequency) reaches 96.8% after calibration, compared with only 78.3% before calibration, proving that the calibration mechanism effectively ensures data comparability.

Professional condenser microphones are recommended for core scenarios such as phased evaluations and skill assessments to ensure high-precision feature extraction;

Portable devices such as smartphones are only used for daily practice feedback. Data must be processed through the system's built-in calibration module before being included in cross-period comparative analysis, avoiding uncalibrated data from directly participating in quantitative evaluation to ensure comparability from a process perspective.

D. SECOND LAYER: FEATURE EXTRACTION AND OPTIMIZATION LAYER

The feature extraction and optimization layer is the core of the technical framework, which extracts multidimensional voiceprint features from preprocessed audio data and performs feature optimization to provide high-quality feature data for quantitative evaluation. The core elements of this layer include feature extraction module and feature optimization module, and the specific implementation path is as follows:

1. Feature extraction module: Extract multi-dimensional voiceprint features related to singing techniques, covering three major categories of features: time domain, frequency domain, and time-frequency domain

Time domain characteristics: including fundamental frequency (F0), sound pressure level (SPL), pitch length, fundamental frequency standard deviation (reflecting pitch stability), sound pressure level change rate (reflecting breath stability), etc;

Frequency domain features: including spectral peak, spectral slope, resonance peak frequency (F1-F4), resonance peak bandwidth, Mel frequency cepstral coefficients (MFCC, extracting 13 dimensional core coefficients), linear predictive cepstral coefficients (LPCC, extracting 12 dimensional core coefficients), etc;

Time frequency domain features: including wavelet transform coefficients (using db4 wavelet basis functions), short-time Fourier transform (STFT) coefficients, etc. [8].

2. Feature optimization module: Optimize the extracted raw features to enhance their effectiveness and robustness

Feature screening: ReliefF algorithm is used to screen core features highly related to singing skills, remove redundant features, and reduce computational complexity;

Feature normalization: Z-score normalization method is used to normalize feature data, eliminate dimensional differences, and improve model training effectiveness;

Feature fusion: Using attention mechanism to fuse multidimensional features, highlighting the importance of core features, and enhancing the representational ability of features [9-10].

E. SUPPLEMENTARY CALCULATION LOGIC OF NEW INDICATORS

Vibrato Rate: Calculated by extracting periodic features of audio through Short-Time Fourier Transform (STFT);

Jitter: Calculated using the sliding window standard deviation of the fundamental frequency sequence;

Shimmer: Statistically analyzed based on inter-frame differences of Sound Pressure Level (SPL).

F. THIRD LAYER: QUANTITATIVE EVALUATION LAYER OF SINGING SKILLS

The quantitative evaluation layer of singing skills is the key to the technical framework, and its core function is based on optimized voiceprint features. Through a quantitative evaluation index system and algorithm model, it achieves objective evaluation of the five core singing skills. The core elements of this layer include evaluation index system, evaluation model, and evaluation standards. The specific implementation path is as follows:

1. Construction of evaluation index system: Construct a quantitative evaluation index system covering the five core singing techniques of pitch accuracy, rhythm, timbre, breath, and resonance, and clarify the voiceprint characteristics and calculation methods corresponding to each index (see Chapter 4 for details);

2. Evaluation model training: Build a multi task quantitative evaluation model based on deep learning algorithms, with Transformer as the backbone network, and achieve parallel evaluation of the five major techniques. The training data for the model includes standard singing audio of professional singers (labeled as high-quality samples) and singing audio of learners at different levels (labeled as samples of different levels), and the model is trained through supervised learning;

3. Evaluation criteria formulation: Referring to the professional standards and expert experience of vocal teaching, quantitative evaluation criteria for each indicator are formulated, and the evaluation results are divided into five levels: "excellent (90-100 points), good (80-89 points), moderate (70-79 points), passing (60-69 points), and failing (below 60 points)" [11].

G. TRAINING DATASET DETAILS

Dataset scale: Contains standard singing audio of 1200 professional singers, with 8600 valid samples and a total duration of 1280 hours. Each sample lasts 15-60 seconds, covering singing segments with different pitches, rhythms, and emotional expressions;

Language diversity: Covers 5 mainstream vocal singing languages - Chinese (Mandarin, Cantonese, minority languages, 65%), English (20%), Italian (8%), French (4%), German (3%), meeting vocal teaching needs of different languages;

Vocal range and voice type coverage: Fully covers 6 core voice types - Soprano (220 singers), Mezzo-Soprano (180), Contralto (100), Tenor (250), Baritone (230), Bass (220). Each voice type includes singing samples from high, medium, and low ranges to ensure the model's generalization ability to different voice types.

H. FOURTH LAYER: PERSONALIZED GUIDANCE GENERATION LAYER

The personalized guidance generation layer is the goal of the technical framework, and its core function is to generate personalized teaching guidance plans based on quantitative evaluation results and learner characteristics. The core elements of this layer include learner portrait construction, guidance scheme generation, real-time feedback, and dynamic optimization. The specific implementation path is as follows:

1. Learner Profile Construction: Collect learners' basic information (such as singing style, learning foundation, learning goals), historical evaluation data, and learning behavior data, construct multidimensional learner profiles, and clarify learners' strengths and weaknesses;

2. Guidance plan generation: Based on the learner profile and quantitative evaluation results, generate targeted guidance plans, including:

Recommended training content: Recommend suitable exercises and skill training modules (such as pitch training and breath training);

Training method guidance: Provide specific singing technique improvement methods (such as breath control techniques, resonance regulation methods);

Training schedule arrangement: Develop a reasonable training plan, clarify training frequency, duration, and stage goals;

3. Real time feedback: During the learner training process, real-time collection of singing data, analysis of skill

improvement, and timely provision of positive feedback and adjustment suggestions;

4. Dynamic optimization: Based on the changes in learners' training effectiveness and evaluation results, regularly update learner profiles, dynamically adjust guidance plans, and ensure the pertinence and effectiveness of guidance [12-13].

III. DESIGN AND DEVELOPMENT OF PERSONALIZED GUIDANCE SYSTEM

A. SYSTEM DESIGN OBJECTIVES AND PRINCIPLES

The weights of timbre indicators are reallocated using the Analytic Hierarchy Process (AHP): Spectral Feature Similarity (30%), Harmonic Richness (25%), Vibrato Rate (20%), Jitter (15%), Shimmer (10%). This ensures that professional evaluation dimensions match the weights of quantitative indicators, consistent with the professional cognition of vocal teaching.

B. DESIGN OBJECTIVES

The core goal of designing a personalized guidance system is to provide accurate, convenient, and personalized learn-

ing guidance for vocal learners based on the quantitative evaluation results of singing skills, and to improve learning efficiency and singing level. Specific goals include:

1. Fully functional: covering core functions such as singing data collection, quantitative evaluation, personalized guidance, training tracking, and resource recommendation;

2. Convenient operation: Designed with a simple and user-friendly interface to ensure that learners and teachers of different technical levels can easily use it;

3. Rapid response: Real time collection and analysis of singing data to ensure the rapid generation of evaluation results and guidance suggestions;

4. Strong personalization: able to generate tailored guidance plans based on individual differences of learners and dynamically optimize them;

5. Stable and reliable: Ensure the stable operation of the system in different devices and network environments, and guarantee data security [14-15].

TABLE 1. Core Indicators of Quantitative Evaluation System for Singing Skills

Primary Indicators	Weight	Secondary Indicators	Corresponding Voiceprint Parameters	Evaluation Standards (Excellent)
Intonation	30%	Pitch Deviation Rate	Fundamental Frequency (F0)	$\leq 1\%$
		F0 Stability	F0 Standard Deviation	≤ 5 Hz
		Transposition Accuracy	F0 Change Rate	$\geq 98\%$
Rhythm	20%	Note Duration Deviation Rate	Note Duration	$\leq 2\%$
		Beat Stability	Beat Interval Standard Deviation	≤ 0.05 s
		Breathing Rhythm Rationality	Breathing Duration/Interval	$\geq 95\%$
Timbre	15%	Spectral Feature Similarity	MFCC, Spectral Features	$\geq 90\%$
		Harmonic Richness	Harmonic/Fundamental Energy Ratio (H1/H2)	≥ 3.5
Breath Control	20%	SPL Stability	SPL Standard Deviation	≤ 2 dB
		Breath Endurance	Long-tone Duration, SPL Attenuation Rate	$\geq 95\%$
		Breath-Pitch Coordination	Correlation between SPL and F0 Change Rates	≥ 0.85
Resonance	15%	Formant Frequency Deviation	Formant Frequencies (F1-F4)	$\leq 3\%$
		Formant Bandwidth Adaptability	Formant Bandwidth (B1-B4)	$\geq 92\%$
		Resonance Intensity	Resonance/Fundamental Intensity Ratio	≥ 2.5

C. DESIGN PRINCIPLES

The system design follows the following five principles:

1. User centered: Fully consider the usage needs of learners and teachers, design convenient operation processes and user-friendly interfaces;

2. Data driven: All guidance plans are based on quantitative evaluation data and learner profiling to ensure the scientific nature of guidance;

3. Modular design: adopting a modular architecture for easy expansion and maintenance of functions;

4. Strong compatibility: Supports multiple terminal devices (computers, phones, tablets) and operating systems (Windows, iOS, Android);

5. Data Security: Adopting technologies such as data encryption and permission management to ensure the security of learners' singing data and personal information [16-17].

D. OVERALL SYSTEM ARCHITECTURE DESIGN

Based on design goals and principles, adopting the architecture pattern of "front-end and back-end separation", the overall architecture of the personalized guidance system is designed, including three parts: front-end layer, back-end layer, and data layer

E. FRONT END LAYER

The front-end layer is the window for user system interaction, using Vue.js as the core development framework and Element UI as the UI component library, to build two major user terminals for learners and teachers:

1. Learner end: The core functions include:

Singing data collection: supports two collection methods: audio upload and real-time recording. Real time recording supports microphone permission calling and audio quality adjustment (sampling rate and bit rate can be configured).

Audio upload supports mainstream formats such as MP3 and WAV;

Quantitative evaluation results viewing: Display quantitative scores and detailed analysis of pitch, rhythm, timbre, breath, and resonance in visual forms such as radar charts and bar charts, and mark the strengths and weaknesses of each technique;

Personalized guidance plan viewing: View training content (such as pitch training modules, breath control exercises), specific methods (such as abdominal breathing training steps, resonance adjustment techniques), and schedule (daily training duration, stage goals) for one's own shortcomings;

Training tracking: Record the singing data and evaluation results of each training session, generate a learning progress curve, and visually display the improvement trend of each skill;

Resource recommendation: View the system's recommended adaptive exercises, teaching videos, audio demonstrations, and other resources, supporting online playback and download of resources;

Feedback and Questions: Submit learning questions and system usage feedback to teachers, and view their responses and supplementary guidance.

2. Teacher end: The core functions include:

Student Management: View the basic information of the students being taught (singing style, learning foundation, learning objectives), learning progress, and previous evaluation results, and support filtering students by dimensions such as class, skill shortcomings, etc;

Evaluation result review: The quantitative evaluation results of the review system can be manually adjusted and supplemented based on teaching experience to generate the final evaluation report;

Customized guidance plan: Based on the guidance plan automatically generated by the system, personalized guidance plans are customized according to the individual situation of students (such as physiological conditions, study habits);

Teaching statistics: View the overall learning situation of students in the class (such as average scores and progress in various skills) and individual differences, generate teaching effectiveness analysis reports, and provide a basis for optimizing teaching methods;

Resource management: Upload and manage teaching resources (such as original exercises, teaching course ware, demonstration audio), and provide targeted recommendations based on students' learning needs [18-20].

TABLE 2. Performance Comparison Between Experimental and Control Groups

Evaluation Dimensions	Experimental Group (System-Assisted)	Control Group (Traditional Teaching)	Improvement Amplitude
Comprehensive Singing Skill Score (Post-Experiment)	82.3±6.5	74.1±7.2	11.1%
Skill Improvement Amplitude	15.6±4.2	8.2±3.8	90.2%
Learning Efficiency (Score/Week)	0.98±0.15	0.51±0.12	92.2%
Learning Satisfaction (10-Point Scale)	8.7±1.1	6.2±1.5	40.3%
Weekly Active Training Duration (Hours)	6.8±1.3	4.2±1.6	61.9%

Revised conclusion statement: "Through a 16-week controlled experiment, it was found that the experimental group using the system achieved an 11.1% improvement in comprehensive singing skills, 90.2% higher than the control group's 5.3%. The weekly skill improvement rate reached 2.13%, significantly higher than the control group's 1.15%."

Table 2 Note: "All indicators are calculated based on the 16-week experimental period. 'Comprehensive Singing Skill Score Improvement' is the absolute improvement ratio; 'Skill Improvement Amplitude' is the relative improvement ratio compared to the control group; 'Learning Efficiency' is the average weekly score improvement."

F. BACKEND LAYER

The backend layer is the core processing module of the system, using Spring Boot as the development framework and providing API interface services, including:

1. Data collection interface: Process the singing audio data uploaded by the front-end, call the voiceprint feature extraction module for feature extraction, support batch data processing and real-time streaming processing;

2. Quantitative evaluation interface: Receive feature ex-

traction results, call a multi task quantitative evaluation model based on Transformer, generate quantitative scores and detailed analysis reports for each singing technique;

3. Learner Profile Interface: Integrate learners' basic information, historical evaluation data, and learning behavior data (such as training frequency and resource access records) to construct and dynamically update multi-dimensional learner profiles;

4. Guidance plan generation interface: Based on quantitative evaluation results and learner profiles, call personalized recommendation algorithms to generate targeted training content, methods, and schedule arrangements;

5. Resource recommendation interface: Based on learners' skill shortcomings and learning preferences, select suitable teaching resources from the resource library to achieve precise integration of "shortcomings resources";

6. Data statistics interface: Perform statistical analysis on learning data and evaluation data, generate learning reports for individual students and the entire class, and support data export function.

G. DATA LAYER

The data layer is responsible for the storage and management of system data, using a hybrid storage architecture of “relational database+non relational database”:

1. Relational database (MySQL): stores structured data, including user basic information (learners, teachers, administrators), quantitative evaluation results, guidance plans, teaching resource metadata, etc;
2. Non relational database (MongoDB): stores unstructured data, including singing audio raw data, voiceprint feature data, teaching videos, audio resources, etc;
3. Cache database (Redis): Cache high-frequency access data, such as learner profiles, recent evaluation results, and popular teaching resources, to improve system response speed;
4. Data backup and recovery: A combination of scheduled backup and remote backup is adopted to ensure data security and support rapid recovery after accidental data deletion.

H. SYSTEM CORE FUNCTION MODULE DESIGN

I. SINGING DATA COLLECTION MODULE

The core function is to collect vocal learners' singing audio data, providing raw data support for subsequent evaluation and guidance. Specific functions include:

1. Real time recording function: supports real-time recording of singing audio through a microphone, displays audio waveform during the recording process, monitors volume changes in real time, and avoids data distortion caused by excessive or insufficient volume;
2. Audio upload function: Supports uploading locally recorded audio files, compatible with mainstream formats such as MP3, WAV, FLAC, etc., with a maximum support for uploading 500MB files;
3. Audio preprocessing function: automatically denoise, clip, normalize the collected audio, remove environmental noise and invalid segments (such as blank space before recording and noise after recording);
4. Singing task association function: Supports learners to select specific singing tasks (such as specified exercises, skill training modules) for recording or uploading, and the system automatically associates corresponding standard audio and evaluation criteria.

The calibration module is based on the Least Mean Square (LMS) algorithm. By comparing with the frequency response benchmark curve of professional microphones, it dynamically adjusts the gain and filtering parameters of the audio. The calibration process takes ≤ 0.2 seconds, which does not affect the user experience, ensuring both ease of use and comparability.

J. QUANTITATIVE EVALUATION MODULE FOR SINGING SKILLS

The core function is based on voiceprint feature extraction and quantitative evaluation model, which achieves objective evaluation of the five core singing techniques. Specific functions include:

1. Automatic evaluation function: After uploading or recording audio, the system automatically extracts voiceprint features, calls a quantitative evaluation model, generates quantitative scores and detailed analysis of pitch accuracy, rhythm, timbre, breath, and resonance;
2. Visualization function of evaluation results: Display the score distribution of the five major techniques in radar charts, display detailed data such as pitch deviation and rhythm fluctuations in line charts, and analyze strengths and weaknesses in text form;
3. Historical evaluation comparison function: supports viewing the trend of changes in previous evaluation results, and visually displays the improvement of various skills;
4. Manual review function: Teachers can review the automatic evaluation results of the system, adjust scores and analysis content, and supplement personalized comments.

K. PERSONALIZED GUIDANCE PLAN GENERATION MODULE

The core function is to generate targeted learning guidance plans based on quantitative evaluation results and learner profiles, including:

1. Shortboard positioning function: automatically identify the core short board skills of learners and clarify the key areas that need to be improved;
2. Training content recommendation function: Based on short board skills, recommend suitable training content, such as pitch practice for short board recommendations, abdominal breathing training for breath short board recommendations, etc;
3. Training method guidance function: Provide specific techniques for improvement, including written instructions, step decomposition, precautions, etc. For example, resonance training provides specific operation methods for “opening the throat” and “adjusting the cavity”;
4. Training plan development function: Based on the learner's learning time and basic level, develop a reasonable training plan, clarify the daily training duration, frequency, and stage goals (such as reducing the pitch deviation rate from 5% to 3% within 2 weeks);
5. Dynamic optimization function: Based on the learner's training effectiveness and new evaluation results, regularly update the guidance plan, adjust the training content and progress.

L. REINFORCEMENT LEARNING-DRIVEN PERSONALIZED GUIDANCE MODULE

1. State space: Defined as an 18-dimensional state vector including “learner's current skill state (5 core skill scores) + learning behavior state (training duration, practice frequency, resource preference) + stage goal state”.
2. Action space: Includes 12 executable actions such as training content selection (3 difficulty levels and 4 practice methods for pitch training), training duration allocation, and feedback timing adjustment.

3. Reward function: A composite reward mechanism is set, including immediate rewards (+10 points for single skill improvement), stage rewards (+50 points for achieving stage goals), and penalty items (-5 points for unsatisfactory training, -10 points for incomplete resource learning), guiding the model to generate longterm optimal guidance strategies.

4. Algorithm selection: Deep Q-Network (DQN) is adopted, with iterative optimization of action selection strategies through experience replay and target network updates. The model undergoes 100,000 training iterations, and the converged strategy adaptation rate reaches 88%.

M. TRAINING TRACKING AND FEEDBACK MODULE

The core function is to track the training process of learners, provide real-time feedback on training effectiveness, and achieve closed-loop learning of “training evaluation feedback optimization”. Specific functions include:

1. Training recording function: Automatically record learners’ training time, training content, and singing data, forming detailed training logs;

2. Stage evaluation function: After completing each stage of training tasks, the system automatically performs stage evaluation, generates stage evaluation reports, summarizes improvement effects and existing problems;

3. Goal achievement reminder function: When learners achieve the preset stage goals, the system sends reminder messages to provide encouragement and enhance learning

motivation.

N. TEACHING RESOURCE RECOMMENDATION MODULE

The core function is to recommend suitable teaching resources for learners and assist in personalized training. Specific functions include:

1. Resource classification management: Teaching resources are divided into four categories: etudes, instructional videos, audio demonstrations, and skill documents. Each category of resources is classified according to singing style (bel canto, ethnic, popular), skill type (pitch accuracy, rhythm, etc.), and difficulty level (beginner, intermediate, advanced);

2. Personalized recommendation function: Based on the learner’s skill shortcomings, learning foundation, and singing style, recommend suitable resources, such as recommending simple pitch practice songs and breath training videos for beginners of popular singing techniques;

3. Resource preview and download function: Supports online preview of resources (such as video clip playback, audio listening), supports downloading of high-quality resources, and facilitates offline learning;

4. Resource evaluation and sharing function: Learners can rate and comment on resources, share their usage experience, and provide reference for other learners.

TABLE 3. System Adaptability Across Different User Groups

User Groups	Skill Improvement Amplitude	Learning Satisfaction (10-Point Scale)	System Usability Rating (10-Point Scale)
Singing Style			
• Bel Canto	14.8±3.9	8.6±1.0	8.9±0.8
• National	15.3±4.1	8.8±0.9	9.0±0.7
• Pop	16.1±4.3	8.9±0.8	9.1±0.6
Learning Level			
• Primary (≤1 Year)	18.2±4.5	8.5±1.2	8.7±0.9
• Intermediate (1-3 Years)	15.1±3.8	8.8±1.0	9.0±0.7

IV. EMPIRICAL TESTING AND RESULT ANALYSIS

A. EMPIRICAL TEST DESIGN

B. PURPOSE OF INSPECTION

The core purpose of empirical testing is to verify the effectiveness, practicality, and adaptability of the vocal technique quantitative evaluation and personalized guidance system based on voiceprint feature extraction, including:

1. Verify the accuracy of the quantitative evaluation of the system: Verify the consistency between the quantitative evaluation results of the five core singing skills by the system and the expert evaluation results;

2. Verify the personalized guidance effect of the system: examine whether learners using the system have significantly better singing skills, learning efficiency, learning satisfaction, and other aspects than learners under traditional teaching modes;

3. Verify the adaptability of the system: Test the adaptability of the system among learners with different singing

styles (bel canto, ethnic, popular) and different learning foundations (beginner, intermediate, advanced);

4. Collect user feedback: Understand the user experience and improvement suggestions of learners and teachers on the system, and provide a basis for system optimization.

C. INSPECTION OBJECT

Using a stratified random sampling method, 300 vocal learners from 2 art colleges (A college is a comprehensive art college, and B college is a music major college) were selected as the empirical test objects. Among them:

Distribution of singing styles: 100 bel canto singers, 100 ethnic singers, and 100 popular singers;

Basic distribution of learning: 100 beginner learners (learning duration ≤ 1 year), 100 intermediate learners (learning duration 1-3 years), and 100 advanced learners (learning duration>3 years);

Gender distribution: 142 males and 158 females;

Age distribution: 215 people aged 18-22, 65 people aged 23-25, and 20 people aged 26 and above. 300 test subjects were randomly divided into an experimental group (150 people) and a control group (150 people). There were no significant differences ($P>0.05$) between the two groups in terms of singing style, learning foundation, gender, age, etc., indicating comparability. This study was approved by the Ethics Committee of Xi'an Shiyou University, and was performed in accordance with the principles of the Declaration of Helsinki. All eligible participants signed an informed consent form.

D. INSPECTION METHODS AND TOOLS

1. Control group experimental method: The experimental group uses the personalized guidance system developed in this article for learning, while the control group adopts the traditional vocal teaching mode ("one-on-one" mentorship teaching without system assistance) for learning. The experimental period is 16 weeks;

2. Singing skills test: Before and after the experiment, two groups of learners will be tested for singing skills, including specific exercises for singing and skill testing (pitch accuracy, rhythm, breath, resonance, timbre). Five vocal education experts will be invited to conduct blind evaluation and scoring (out of 100 points);

3. Quantitative evaluation comparison: Compare the systematic evaluation results of the experimental group learners with the expert evaluation results, and calculate the degree of consistency (using Pearson correlation coefficient);

4. Questionnaire survey method: After the experiment, a questionnaire survey was conducted on two groups of learners to understand their learning efficiency, satisfaction, and initiative. A total of 300 questionnaires were distributed, and 292 valid questionnaires were collected, with an effective response rate of 97.3%;

5. In depth interview method: After the experiment, 30 learners and 10 teachers from the experimental group were selected for in-depth interviews to understand the user experience, advantages, and disadvantages of the system;

6. Statistical analysis method: Using SPSS 26.0 and Python data analysis tools, descriptive statistics, independent sample t-test, analysis of variance, correlation analysis, etc. are conducted on test data and questionnaire data to verify research hypotheses.

E. INSPECTION INDICATORS

Construct a two dimensional inspection index system, including system performance indicators and learning effectiveness indicators:

1. System performance indicators: including quantitative evaluation accuracy, system response speed, operational convenience, stability, used to evaluate the technical performance of the system;

2. Learning effectiveness indicators: including comprehensive scores for singing skills, scores for each individual skill, learning efficiency (skill improvement rate/learning duration), learning satisfaction, and learning initiative, used to evaluate the teaching effectiveness of the system.

F. EMPIRICAL TEST RESULTS AND ANALYSIS

G. SYSTEM QUANTITATIVE EVALUATION ACCURACY VERIFICATION RESULTS

The correlation analysis between the systematic quantitative evaluation results of 150 learners in the experimental group and the blind evaluation results of 5 experts showed that:

The Pearson correlation coefficient for pitch assessment was 0.892 ($P<0.01$);

The Pearson correlation coefficient for rhythm evaluation was 0.876 ($P<0.01$);

The Pearson correlation coefficient for timbre evaluation was 0.835 ($P<0.01$);

The Pearson correlation coefficient for breath assessment was 0.858 ($P<0.01$);

The Pearson correlation coefficient for resonance evaluation was 0.823 ($P<0.01$);

The Pearson correlation coefficient for overall singing skill evaluation was 0.887 ($P<0.01$).

The results indicate that the quantitative evaluation results of the system are highly correlated with the expert evaluation results, proving that the system has good accuracy in quantitative evaluation and can objectively reflect the singing skills level of learners.

After expanding the timbre indicator system, the Pearson correlation coefficient between the system's timbre evaluation results and expert evaluation results increased from 0.835 to 0.892 ($P<0.01$), proving that the new indicators can more comprehensively capture timbre characteristics and improve the professionalism and accuracy of quantitative evaluation.

TABLE 4. Core Technical Specifications of the Personalized Guidance System

System Modules	Core Functions	Technical Support	Performance Indicators
Vocal Data Collection	Real-time recording, audio upload, preprocessing	WebRTC, Audio denoising algorithm	Supported formats: MP3/WAV/FLAC; Max file size: 500MB
Quantitative Evaluation	Feature extraction, multi-skill parallel evaluation	Transformer-based model, Librosa	Evaluation accuracy: 90.1%; Response time: $\leq 0.8s$
Personalized Guidance	Shortcoming positioning, training plan generation	Collaborative filtering, Reinforcement learning	Plan adaptation rate: 88%; Dynamic update cycle: Weekly
Training Tracking & Feedback	Progress recording, real-time feedback	Data visualization, Edge computing	Progress curve update: Real-time; Feedback delay: $\leq 0.3s$

V. RESEARCH CONCLUSION

This article aims to address the core pain points of subjective evaluation, insufficient personalized guidance, and difficulty in quantifying techniques in traditional vocal teaching by integrating the structural logic of high-performance technical production into the pedagogical framework. By utilizing the interwoven data modeling found in advanced material engineering, the research establishes an objective evaluation system that parallels the mechanical precision of industrial craftsmanship. It conducts research on AI powered optimization of vocal teaching and constructs a quantitative evaluation and personalized guidance system for singing techniques based on voiceprint feature extraction. Through theoretical analysis and empirical testing, the following core conclusions are drawn:

We have developed a fully functional personalized guidance system: adopting a “front-end and back-end separation” architecture, we have designed and developed a personalized guidance system that includes core functions such as singing data collection, quantitative evaluation, personalized guidance, training tracking, and resource recommendation. The system has the characteristics of easy operation, rapid response, strong personalization, and stable reliability. The technical performance of the system has been verified through multi-dimensional testing.

Empirical tests have verified the effectiveness and adaptability of the system: Through a 16 week controlled experiment, it was found that the experimental group of learners using the system showed significant improvements in singing skills (21.3% vs 11.5%), learning efficiency (0.98 points/week vs 0.51 points/ week), learning satisfaction (8.7 points vs 6.2 points), and learning initiative compared to the control group; The system demonstrates exceptional adaptability across the specialized textures of bel canto, ethnic, and popular singing styles, mirroring the precision required to calibrate tension and structure density for beginner, intermediate, and advanced craftsmanship. By seamlessly integrating these vocal techniques with the foundational principles of material manipulation and material construction, the platform has earned high recognition from both learners and instructors for its versatile, multi-disciplinary approach.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

This study was approved by the Ethics Committee of Xi'an Shiyou University, and was performed in accordance with the principles of the Declaration of Helsinki. All eligible participants signed an informed consent form.

AUTHOR CONTRIBUTIONS

Yuzi Wei designed, collected and analyzed the data, and drafted the manuscript. Yuzi Wei conducted the study, critically revised the manuscript for important intellectual content, and gave final approval of the version to be published. Yuzi Wei participated fully in the work, take public responsibility for appropriate portions of the content, and

agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

CONFLICTS OF INTEREST

The author declares no conflict of interest.

FUNDING

This research received no external funding.

ACKNOWLEDGEMENTS

Not applicable.

REFERENCES

- [1] C. Peng, “A Personalized Recommendation Method for Vocal Music Teaching Resources Based on Network Topology,” *International Journal of High Speed Electronics and Systems*, vol. 35, no. 03, 2025, doi: 10.1142/S012915642540467X.
- [2] F. Bin, “Exploration of the Integration of Singing Techniques and Artistic Expression in Vocal Teaching,” *Adult and Higher Education*, vol. 7, no. 1, 2025, doi: 10.23977/ADUHE.2025.070107.
- [3] Z. Zhang, “The Teaching Method of STEAM Education-based Audio-visual Aesthetics in College Vocal Music Teaching,” *Frontiers in Educational Research*, vol. 8, no. 2, 2025, doi: 10.25236/FER.2025.080216.
- [4] W. Wang, “Analysis of the Transformation Path of Digitally Empowered Vocal Music Teaching in Colleges and Universities,” *Exploration of Educational Management*, vol. 3, no. 2, 2025, doi: 10.12417/3029-2328.25.02.008.
- [5] Z. Zhina, “Research on the application and teaching practice of ‘thirteen Zhe’ in Chinese art songs,” *The Art & Design Research*, vol. 6, no. 1, 2025, doi: 10.37420/J.ADR.2025.006.
- [6] Y. Cheng, “The Significance of Chinese Original Vocal Music for the Development of Chinese Ethnic Vocal Music,” *Education Research and Innovation*, vol. 1, no. 2, 2025, doi: 10.62639/SSPERI02.20250102.
- [7] L. Qin, “Research on Vocal Teaching under the Perspective of Multicultural Nationalities,” *Education Research and Innovation*, vol. 1, no. 2, 2025, doi: 10.62639/SSPERI05.20250102.
- [8] Z. Hu, “Research on Interactive Content Development of Vocal Music Teaching in College Music Courses in the New Media Era,” *Applied Mathematics and Nonlinear Sciences*, vol. 10, no. 1, 2025, doi: 10.2478/AMNS-2025-0029.
- [9] M. Wang and H. Wang, “Research on Diversified Innovative Paths of Vocal Music Teaching in College Music Education Promoted by Information Technology,” *Applied Mathematics and Nonlinear Sciences*, vol. 10, no. 1, 2025, doi: 10.2478/AMNS-2025-0638.
- [10] X. Huang, “The Optimization of Strategies for Precision Teaching Reform of Vocal Music Education in Colleges and Universities in the Framework of Information Technology,” *Applied Mathematics and Nonlinear Sciences*, vol. 10, no. 1, 2025, doi: 10.2478/AMNS-2025-0679.
- [11] Y. Chen, “Analysis of the Reform of Vocal Music Teaching by Using Network Platform in the New Media Era,” *Lecture Notes in Electrical Engineering*, pp. 747-754, 2022, doi: 10.1007/978-981-19-4775-9_96.
- [12] M. Moreno and E. Woodruff, “Exploring the Effects of Background Music on Real-Time Emotional Expressions, Performance, and Confusion Mediation in Middle School Students,” *Technology, Knowledge and Learning*, vol. 28, no. 1, pp. 143-163, 2023, doi: 10.1007/s10758-021-09556-9.
- [13] Z. Sun, “The Reform of Vocal Music Education in Universities Under the Concept of Innovation,” *Contemporary Education and Teaching Research*, vol. 5, no. 12, 2024, doi: 10.61360/BoniCETR242017251203.
- [14] Y. Zhu, “Discussion on the Application of Innovative Teaching Methods in the Vocal Course of Middle and High Preschool Vocational Education Majors,” *Educational Innovation Research*, vol. 2, no. 2, pp. 92-97, 2024.
- [15] J. Zheng, Y. Zhang, and S. Zhang, “Audio-visual aesthetic teaching methods in college students’ vocal music teaching by deep learning,” *Scientific Reports*, vol. 14, no. 1, 2024, doi: 10.1038/s41598-024-80640-7.
- [16] F. Bin, “Innovation and Practice of Vocal Teaching Mode in Digital Environment,” *Curriculum and Teaching Methodology*, vol. 7, no. 8, 2024, doi: 10.23977/CURTM.2024.070828.

- [17] M. Hu, "Research on the Value and Path of Family Art Education-Take Vocal Music Education as an Example," *Journal of Research in Vocational Education*, vol. 6, no. 10, pp. 43-47, 2024, doi: 10.53469/JRVE.2024.6(10).08.
- [18] X. Wenyi and W. Decai, "Research on the employment dilemma and path of college vocal music graduates," *Adult and Higher Education*, vol. 6, no. 7, 2024, doi: 10.23977/aduhe.2024.060709.
- [19] Q. Du, "A Systematic Approach to Innovative Strategies for Vocal Instruction in Higher Education: Enhancing Student Performance," *Pacific International Journal*, vol. 7, no. 5, 2024, doi: 10.55014/PIJ.V7I5.696.
- [20] R. Schutz, *Welsh Vocal Music: A Guide to Lyric Diction and Repertoire*. London, U.K.: Taylor & Francis, 2024, doi: 10.4324/9781003347347.