SIN-CHN SCIENTIFIC PRESS PTE.LTD.

Molecular & Cellular Biomechanics 2025, 22(1), 1003.
https://doi.org/10.62617/mcb1003

Article

Oral biomechanical feature recognition method and system for vocal pronunciation resonance

Shu Liu

School of music, Shenyang Normal University, Shenyang 110032, China; 15998301875@163.com

CITATION

Liu S. Oral biomechanical feature recognition method and system for vocal pronunciation resonance. Molecular & Cellular Biomechanics. 2025; 22(1): 1003. https://doi.org/10.62617/mcb1003

ARTICLE INFO

Received: 4 December 2024
Accepted: 12 December 2024
Available online: 14 January 2025

Abstract: This paper discusses a vocal pronunciation resonance recognition method based on oral biomechanical features, aiming to collect the movement data inside the oral cavity through high-precision sensors, and combine support vector machines and convolutional neural networks to achieve precise recognition and optimization of pronunciation resonance features. First, by using high-precision sensors such as tongue position sensors and soft palate movement sensors, key biomechanical data such as tongue position, oral morphology and soft palate movement of singers during pronunciation are obtained in real-time. Then, feature extraction and analysis are performed to extract biomechanical features such as movement amplitude, frequency, and speed of oral organs, and a resonance feature recognition model is constructed using support vector machines and convolutional neural networks to recognize and optimize the resonance area during pronunciation. Finally, this paper designs and develops an automated recognition system that can provide real-time feedback on the singer's pronunciation data, and provide personalized training suggestions based on the recognition results to optimize the pronunciation resonance effect. The experimental results show that the resonance recognition method based on biomechanical data can significantly improve the accuracy and personalization of pronunciation, and help improve the efficiency and scientificity of vocal training. The system has good stability, with a response time of less than 500 ms and a CPU (Central Processing Unit) usage rate of no more than 70%. This method provides effective technical support for the digitalization and personalized optimization of vocal pronunciation, and has high practical value and promotion potential.

Keywords: vocal training; oral biomechanics; phonation resonance; machine learning; personalized optimization

1. Introduction

Vocal pronunciation resonance is an important part of vocal training, which directly affects the quality, volume and timbre of the sound. The effect of resonance not only depends on the vibration of the vocal cords, but is also closely related to the coordination of the pronunciation organs such as the mouth, tongue, and soft palate [1,2]. Traditional vocal training mainly relies on the subjective perception of the singer and the experience guidance of the teacher. However, this method has certain limitations [3,4]. Due to individual differences in the oral cavity and pronunciation organs, traditional methods often find it difficult to provide personalized optimization solutions, and cannot precisely quantify the resonance characteristics during the pronunciation process [5,6]. In addition, traditional research has rarely conducted quantitative analysis of oral biomechanical characteristics, and has not yet fully utilized modern scientific and technological means to achieve scientific and systematic analysis and optimization [7,8]. Therefore, how to achieve scientific

recognition and optimization of vocal pronunciation resonance through modern technical means has become a major challenge in current vocal research and practice.

In recent years, many scholars have begun to pay attention to the role of oral biomechanical characteristics in the process of vocal pronunciation [9,10]. Some studies have attempted to analyze the movement patterns of organs such as the mouth, tongue, and soft palate during vocal pronunciation through physical modeling and experimental observation, but these studies are often limited to theoretical analysis or a small amount of experimental data, lacking large-scale dynamic monitoring and real-time feedback [11,12]. In addition, some studies have used acoustic features or physiological signals to analyze pronunciation effects, but have not yet explored in depth how to optimize pronunciation resonance through precise biomechanical feature recognition [13,14]. Although some studies use sensors to monitor oral movements, due to technical limitations, these methods often cannot obtain sufficiently precise data in real-time during dynamic pronunciation, resulting in low recognition accuracy [15,16]. Therefore, how to combine biomechanical data with feature recognition technology to solve these problems in traditional research has become an urgent issue to be solved.

To solve the above problems, vocal pronunciation analysis methods based on modern technology have gradually gained attention in recent years [17,18]. For example, using oral sensors, tongue position detectors and three-dimensional (3D) imaging technology, researchers have begun to try to quantitatively monitor oral movements during vocal pronunciation, and combined with machine learning and data analysis technology, proposed recognition methods for pronunciation resonance characteristics [19,20]. These methods achieve high-precision analysis of oral movements by collecting biomechanical data during the pronunciation process [21,22]. However, these methods still face some challenges in practical applications, such as improving data collection accuracy, real-time issues and adaptability to individual differences [23,24]. Although these technologies provide strong support for vocal training, they still need to be further improved [25,26]. This paper proposes a recognition method based on oral biomechanical characteristics, combined with machine learning and high-precision sensing technology, aiming to overcome the shortcomings of existing research and achieve precise analysis and optimization of vocal pronunciation resonance [27,28].

The purpose of this study is to propose a new vocal pronunciation resonance recognition method based on oral biomechanical characteristics and construct a corresponding recognition system. By combining sensor technology, machine learning and data analysis methods, this paper first collects oral biomechanical data during vocal pronunciation, and precisely analyzes the movement and changes of organs such as the mouth, tongue and soft palate during pronunciation through feature extraction and recognition. Secondly, this paper uses feature recognition methods to effectively recognize and optimize the resonance characteristics in vocal pronunciation, thereby providing quantitative data support and personalized optimization solutions for vocal training. Through this method, this paper can not only provide new technical means for vocal training, but also provide a new perspective and theoretical basis for the study of vocal pronunciation resonance.

2. Oral biomechanical data acquisition

2.1. Sensor selection and layout

This study uses a variety of high-precision sensors to ensure that the subtle changes in oral movement can be fully and accurately captured. The specific sensors include:

Tongue position sensor: Based on the principle of electromagnetic induction, the tongue position sensor consists of multiple sensor elements and is installed in different parts of the tongue, such as the root, back and tip of the tongue. These sensors can precisely measure the position changes of the tongue in 3D space and provide real-time spatial coordinate information. Soft palate movement sensor: Using a sensing technology based on capacitance change, the soft palate sensor captures the vertical movement (up/down) and horizontal movement (forward and backward displacement) of the soft palate during pronunciation in real-time through an electrode array arranged in the soft palate area [29,30]. This sensor can reflect the details of the soft palate movement and provide data support for analyzing the impact of the soft palate on resonance. Oral opening and closing sensor: Based on optical sensing or displacement sensing technology, the oral opening and closing sensor is used to monitor the opening and closing degree of the upper and lower jaws in real-time. It can precisely measure the size and angle of the oral opening during singing, and then provide key data for analyzing the volume change of the resonance cavity [31,32]. The arrangement and installation position of the sensor are precisely designed to ensure that all important movement areas in the mouth are covered, from tongue movement to soft palate adjustment, and then to changes in oral opening and closing, to fully capture the singer's biomechanical data, as shown in **Figure 1**.

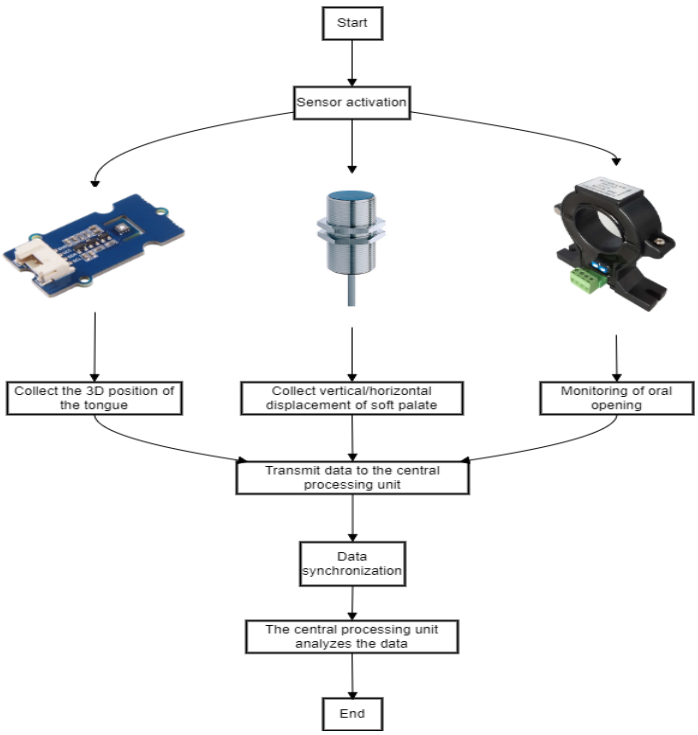

Figure 1. Sensor application.

2.2. Data collection process

Data is collected from various groups of singers (different genders, age groups, accents, etc.), and a standardized collection process is used to ensure data diversity and balance, thereby improving the model's generalization ability. When the singer is performing pronunciation training, all sensors record movement parameters such as tongue, soft palate and oral opening and closing in real-time, and transmit these data to the central processing unit through the wireless communication module.

Before collecting data, the system performs self-test and calibration on each sensor to ensure the accuracy and stability of the collected data. The tongue position sensor determines the standard coordinate system of the tongue by calibrating the initial position, and the soft palate sensor and oral opening and closing sensor are initially set according to the static posture of the oral cavity. When the singer begins to pronounce, the sensor continuously collects the movement data of the tongue, soft palate, and oral opening and closing. To further ensure the accuracy and reliability of data acquisition, considering the possible impact of the external environment (such as temperature and humidity) on the sensor's accuracy, the system conducts multiple experiments to simulate different environmental conditions (such as high temperature, low humidity, high humidity, etc.) and observe the changes in the output data of the sensors under different environments. During the experiment, the system's stability and data error under these conditions are recorded and analyzed to ensure that the system can maintain high accuracy in actual use.

The tongue position sensor provides the displacement information of the tongue on the X, Y, and Z axes; the soft palate sensor records the lifting angle and horizontal displacement of the soft palate; the oral opening and closing sensor monitors the opening and closing changes of the upper and lower jaws [33,34]. The data of all sensors are sent to the central data processing unit in real-time through the wireless transmission system. Regarding data acquisition and synchronization, the system adopts real-time data transmission technology based on wireless communication modules to ensure efficient and stable transmission of sensor data. Data synchronization is achieved at the software level through timestamp technology to ensure the timing consistency of tongue position, soft palate movement, and oral opening and closing data [35,36]. The system records a timestamp for each sensor data. During the transmission process, the data is arranged and synchronized in chronological order to ensure that the data of all sensors can precisely correspond to every moment of pronunciation.

In the process of multi-sensor data fusion, weighted averaging, standardization, and feature selection are used to optimize the data fusion effect. First, through the weighted averaging method, different weights are given to the data of each sensor according to the importance and reliability of the data of different sensors, so as to enhance the influence of key feature data. Secondly, to eliminate the differences in dimensions of different sensor data, the data is standardized so that the data of each sensor is within the same scale range to avoid excessive influence of some data with larger dimensions on the model. Finally, after data fusion, the most representative features are extracted to optimize the recognition effect of resonance features.

In **Table 1**, through high-precision sensors and wireless communication technology, the system can precisely and stably record movement data such as tongue, soft palate and oral opening and closing. These data provide a scientific basis for subsequent resonance feature recognition, pronunciation optimization and personalized training, and provide strong technical support for the digitization and scientificization of vocal training.

Table 1. Collected data.

Timestamp	Tongue Position (X, mm)	Tongue Position (Y, mm)	Tongue Position (Z, mm)	Soft Palate Angle (°)	Soft Palate Displacement (mm)	Oral Opening and Closing (mm)
0.00s	1.5	2.1	0.3	5	3.2	10.5
0.10s	1.55	2.15	0.32	5.1	3.25	10.7
0.20s	1.6	2.2	0.35	5.2	3.3	10.9
0.30s	1.65	2.25	0.37	5.4	3.35	11.1
0.40s	1.7	2.3	0.4	5.5	3.5	11.3
0.50s	1.75	2.35	0.42	5.7	3.55	11.5
0.60s	1.8	2.4	0.45	5.8	3.6	11.7
0.70s	1.85	2.45	0.47	6	3.7	11.9
0.80s	1.9	2.5	0.5	6.1	3.8	12.1
0.90s	1.95	2.55	0.53	6.2	3.9	12.3

2.3. Real-time monitoring and processing of data

To ensure the integrity and precision of the data, all collected movement data are monitored and preliminarily processed in real-time. The following are the processing steps:

All raw data are first denoised by filters to eliminate irrelevant signals generated by sensor errors or environmental interference. A low-pass filter is used to remove high-frequency noise, and Kalman filtering technology is used to smooth irregular fluctuations.

The influence of different window sizes on data denoising is evaluated through comparative experiments. The size of the sliding window is optimized according to the experimental results, and the optimal parameters are selected to improve the denoising effect.

To eliminate data deviations caused by individual differences, all sensor data are standardized, including unifying the values of tongue position, soft palate angle and oral opening and closing to the standard range, so as to ensure that the data between different singers are comparable.

While collecting data, the system feeds back the singer's oral movement status to the training platform in real-time to help singers understand the changes in tongue position, soft palate and oral opening and closing during pronunciation. The feedback system can dynamically adjust the singer's posture based on real-time data and provide personalized training suggestions.

2.4. Data accuracy verification and calibration

To ensure that the biomechanical data collected by the sensor is highly accurate and reliable, multiple verification steps are applied in the data collection process:

Through multi-point cross-validation, the consistency of the collection results of the tongue position sensor, soft palate sensor and oral opening and closing sensor is ensured. If a sensor data is abnormal, the system automatically calibrates or marks it as abnormal data. Before the system is officially put into use, the research team conducts multiple calibration experiments on all sensors. By comparing with the standard biomechanical model, the measurement accuracy and reliability of the sensor are verified [37,38]. The calibration results show that the tongue position sensor and soft palate movement sensor meet the high-precision requirements in terms of spatial resolution and time response, and can provide real movement data. The sensor is equipped with an adaptive correction function, which can adjust the sensitivity of data collection in real-time according to the singer's movement feedback to adapt to changes in different oral morphology and movement amplitude.

2.5. Data storage and subsequent analysis

All collected biomechanical data not only need to be fed back to the training platform in real-time, but also need to be stored and analyzed for a long time. Data storage uses a distributed database system to ensure the security and traceability of large-scale data collection. The stored data is further analyzed to recognize and optimize key movement patterns in the pronunciation process.

In the subsequent analysis, the singer's movement data is compared with the resonance characteristics in the sound signal to recognize the impact of different oral movement states on the vocal pronunciation resonance. Through big data analysis and machine learning algorithms, more accurate personalized training guidance can be provided to singers.

Regarding user data collection and privacy issues, data storage and processing use high-standard encryption technology, and all data is anonymized to ensure the security of user identity information. To further protect user privacy, the system uses a series of encryption algorithms, such as AES (Advanced Encryption Standard) to encrypt stored data, and conducts regular security audits to ensure the compliance and security of data storage.

3. Oral movement feature extraction

3.1. Movement amplitude extraction

First, the movement amplitude is a key indicator to measure the displacement of oral organs during pronunciation. For the spatial data collected by the tongue position sensor, soft palate movement sensor, and oral opening and closing sensor, the following steps are used to extract the movement amplitude:

According to the 3D coordinate data provided by the tongue position sensor, the displacement of the tongue along the X, Y, and Z axes during pronunciation is calculated. Specifically, the movement amplitude of the tongue position is obtained by calculating the difference between the maximum displacement and the minimum

displacement of the tongue in the pronunciation cycle. To ensure the accuracy of the data, the sliding window method is used to smooth the tongue displacement data and remove noise interference. The movement amplitude of the soft palate is extracted by recording its displacement in the vertical and horizontal directions. The maximum displacement range of the soft palate during pronunciation is calculated based on the angle change data of the soft palate movement sensor. The peak detection algorithm is used to analyze the soft palate movement, recognize its main movement cycle, and calculate the movement amplitude. The maximum range of change of the oral opening and closing is calculated based on the upper and lower jaw displacement data collected by the oral opening and closing sensor. The data is divided into multiple pronunciation cycles using the segmentation method, and the change amplitude of the opening degree is calculated in each cycle.

The red curve in the upper sub-figure of **Figure 2** represents the original tongue X-axis position data, which contains sensor noise. The blue curve represents the data after sliding average denoising. The blue curve is smoother than the red curve, and the noise fluctuation is significantly suppressed. The X-axis displacement of the tongue can better reflect the actual movement trend. The red curve in the middle sub-figure represents the original data of the tongue position in the Y-axis direction. The curve fluctuates greatly, indicating that there is obvious noise in the data. The blue curve shows the tongue position data after sliding average denoising. It is smoother than the red curve and can clearly reflect the change trend of the tongue in the Y-axis direction. The red curve in the lower sub-figure represents the original data of the tongue position in the Z-axis direction. Similar to the previous two, there are certain random fluctuations in the data. The blue curve is processed by sliding average denoising, showing an obvious smoothing effect, removing the influence of noise, and highlighting the actual movement trend of the tongue position in the Z-axis.

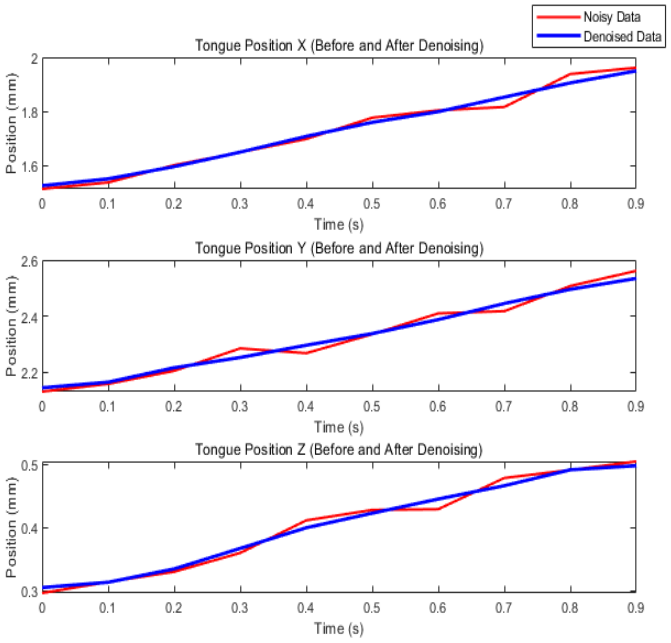

Figure 2. Data denoising.

3.2. Movement speed extraction

The movement speed reflects the dynamic changes of the oral organs during the pronunciation process. The movement speed of the organs is obtained by differential Processing of the time series data. The specific extraction steps are as follows: according to the displacement data of the tongue position sensor, the first-order difference algorithm is used to calculate the speed of the tongue position changing over time. By differentiating the displacement data at consecutive moments, the instantaneous speed of the tongue at each time point is obtained. To eliminate the influence of noise on the speed calculation, the weighted sliding average method is used to smooth the speed data. The movement speed of the soft palate is obtained by calculating the speed of the soft palate angle changing over time. Using the angle data provided by the soft palate sensor, a differential operation is performed to calculate its rate of change in each pronunciation cycle. The characteristics of the speed change of the soft palate during pronunciation are obtained by the peak extraction method. The speed of change of the oral opening and closing degree is obtained by calculating the rate of change of the upper and lower jaw displacements. The speed of the oral opening and closing at each time point is calculated using the differential method, and the sudden speed fluctuations are removed by smoothing filtering technology.

3.3. Movement frequency extraction

The movement frequency reflects the periodic characteristics of the movement of the oral organs, which directly affects the stability and resonance characteristics of the pronunciation. The specific steps for extracting the frequency are as follows: The movement frequency of the tongue is extracted by Fourier transforming the displacement data. In the pronunciation cycle, the tongue displacement data is analyzed in the frequency domain using the fast Fourier transform to recognize the main frequency components. This frequency is closely related to the periodicity of tongue movement and can reveal the relationship between tongue position and resonance cavity. The same frequency domain analysis method is used for the frequency of soft palate movement. The angle data of the soft palate is Fourier transformed to extract the main frequency components of the soft palate movement. By analyzing the peaks in the spectrum, the frequency of the soft palate movement can be determined and compared with the resonance characteristics to find the correlation between the two. By frequency domain analysis of the oral opening and closing data, the frequency of the opening and closing changes is extracted. Fourier transform applied to the data of each pronunciation cycle can reveal the opening and closing frequency and its matching degree with the resonance frequency. The frequency characteristics of oral opening and closing can reveal the coordination between the size of the resonance cavity and the vibration of the sound wave.

4. Biomechanical modeling and simulation

4.1. Physical modeling of oral pronunciation organs

Based on the collected biomechanical data (including tongue position, soft palate movement, oral opening and closing, etc.), a three-dimensional geometric model of the oral organs is first established, which includes the main structures such as the tongue, soft palate, upper palate, lower jaw, oral cavity and vocal cords.

In **Figure 3**, the specific modeling steps are as follows: Based on the collected oral organ displacement data, a geometric model of the oral organ is established. The morphology of the tongue, soft palate and oral cavity is represented in detail to ensure that the model can precisely reflect the geometric morphology and movement changes of the actual organs. Mesh division is performed on the 3D model of the oral organ, and the finite element method is used to discretize the entire oral area. According to the complexity of the model, a finer mesh division accuracy is used to ensure the accuracy in acoustic calculations and airflow dynamics simulations. High-density meshes are used for key parts such as the tongue and soft palate to obtain higher calculation accuracy in subsequent simulations. According to different biomechanical properties, appropriate physical properties are assigned to each part of the model. For example, the soft tissue properties of the soft palate and tongue use material parameters such as elastic modulus and density; the hard tissue parts of the oral cavity and the upper and lower jaws use rigid material properties. To simulate the vibration of the vocal cords more realistically, the vocal cord part uses a linear spring model to reflect the tension and vibration characteristics of the vocal cords.

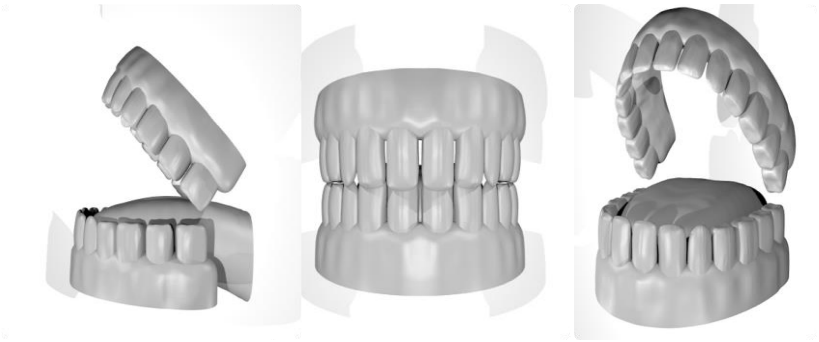

Figure 3. Oral organ model.

4.2. Coupling simulation of airflow and vocal cord vibration

After building the oral biomechanical model, the next goal is to simulate the interaction between vocal cord vibration and airflow in the oral cavity, which has a decisive influence on the resonance characteristics of pronunciation. To achieve this, the fluid-solid coupling simulation method is used, and the specific process is as follows:

In the oral cavity, the state of airflow is described by the fluid dynamics equation. The computational fluid dynamics model based on the Navier-Stokes equation is used to simulate the flow state of air in the oral cavity. By setting the inlet pressure, temperature and flow rate of the airflow, the vibration, airflow fluctuation and air pressure change caused by the airflow passing through the vocal cords are simulated. To accurately simulate the vibration characteristics of the vocal cords, the vibration equation based on the finite element method is used. The vocal cords are regarded as elastic films, and their vibration is caused by the tension of the

vocal cords and the action of the airflow. By solving the dynamic equations of vocal cord vibration, the displacement, velocity, acceleration and other parameters of the vocal cords are obtained, and the vibration mode of the vocal cords is further calculated based on these parameters. The airflow model is coupled with the vocal cord vibration model to form a dynamic fluid-solid coupling system. By gradually iteratively solving the coupling equations of airflow and vocal cord vibration, the interaction relationship between airflow and vocal cord vibration is obtained. The change of airflow causes the vibration of the vocal cords. Conversely, the vibration of the vocal cords also affects the flow pattern of the airflow, thus forming a complex feedback mechanism.

4.3. Finite element analysis modeling and solution

In the process of biomechanical modeling, the finite element analysis method is used to precisely model the biomechanical characteristics of the oral cavity. The specific steps are as follows:

Based on the elastic model of the oral organs, it is first assumed that the movement of the oral organs is a small deformation. According to the Newton-Euler equation, the interaction between the airflow and the organs is described. By establishing equations including airflow dynamics, elastic mechanics and contact mechanics, a complete finite element simulation model is derived. In FEM (Finite Element Method), the oral organs and the airflow area are discretized into a finite number of grid cells. To ensure the simulation accuracy, the degree of meshing depends on the simulation requirements. For example, the vocal cords and the surrounding areas use finer grids to simulate the tiny vibrations of the vocal cords and the detailed changes of the airflow. The entire system is iteratively solved by combining structural mechanics, fluid mechanics and acoustic equations through a numerical solver. The implicit time integration method is used to solve the biomechanical and airflow equations to obtain the solutions at each instant during the pronunciation process. Through simulation analysis, the vocal cord vibration frequency, airflow velocity, and the coupling effect between the airflow and the vocal cords are obtained.

4.4. Optimization of resonance area

The formation and change of the resonance area under different pronunciation conditions are further analyzed by finite element analysis method. According to the vibration characteristics of airflow and vocal cords, the influence of the resonance cavity shape and airflow distribution on the pronunciation resonance effect is studied. In the simulation process, the frequency response of airflow and vocal cord vibration is focused on, and the frequency distribution of air vibration in the resonance cavity is analyzed. By adjusting the position, shape and vibration frequency of the organs in the model, the influence of different oral movements on the resonance frequency is studied. According to the simulation results, the distribution of resonance frequency and airflow intensity is extracted to recognize the strongest resonance area in the oral cavity. Through these data analysis, specific optimization solutions can be provided for personalized training. By adjusting

factors such as tongue position, soft palate angle and oral opening and closing, the resonance effect under different training conditions is simulated to provide a scientific basis for vocal training.

5. System design and implementation

5.1. Resonance feature recognition model

To accurately identify the pronunciation resonance characteristics of singers and optimize their effects, this paper designs a resonance feature recognition model based on machine learning. The construction of this model is divided into the following steps:

First, by comparing the experiments of SVM, CNN, decision tree, random forest, and other algorithms, the accuracy, training time, and overfitting degree of each model are evaluated, and the best algorithm is selected for optimization.

Biomechanical feature data under different pronunciation states are annotated by experts, including tongue position, soft palate movement, and oral opening degree. These data provide training samples for machine learning models. This paper adopts supervised learning methods, among which support vector machine and convolutional neural network are the two main models. SVM (Support Vector Machine) is used to process high-dimensional feature data, which is particularly suitable for classification tasks with small samples. In this system, SVM is used to recognize the resonance features under different pronunciation states and distinguish different resonance patterns. CNN (Convolutional Neural Network) is good at processing data with spatial structure, especially suitable for the analysis of image and time series data. By converting the time series data of tongue position, soft palate movement, and oral opening and closing degree into feature maps, CNN can recognize complex pronunciation resonance patterns and optimize the model.

Regarding algorithm optimization, the optimization of SVM mainly focuses on feature selection and kernel function selection. By using feature selection algorithms (such as principal component analysis PCA) to reduce the redundancy of high-dimensional data, the model is easier to train and can maintain high accuracy. At the same time, the cross-validation method is used to select the optimal kernel function type (such as RBF kernel or linear kernel) to improve the accuracy of classification.

To optimize the model's performance, data enhancement techniques (such as time shift, time scaling, etc.) are used to increase the diversity of training samples, and methods such as Dropout regularization and batch normalization are used to avoid overfitting. To further improve the model's prediction accuracy, the hyperparameters of CNN are adjusted, including the number of convolution layers, the size of the convolution kernel, the design of the pooling layer, etc., and the grid search method is used to optimize the hyperparameter settings.

5.2. User interface design

The user interface design of the system is simple and intuitive, aiming to improve the user experience of singers. The main functions of UI (User Interface) design include:

During pronunciation training, users can view the movement status of tongue position, soft palate movement and oral opening and closing in real-time. The system displays various biomechanical data in the form of charts or animations to help users understand the pronunciation status. Based on the resonance feature recognition model, the system provides personalized training suggestions based on the singer's real-time data feedback. For example, if the user's tongue position is too low or the soft palate movement is insufficient, the system recommends that the user adjust the tongue position or soft palate movement to optimize the pronunciation resonance effect. The interface is shown in **Figure 4**.

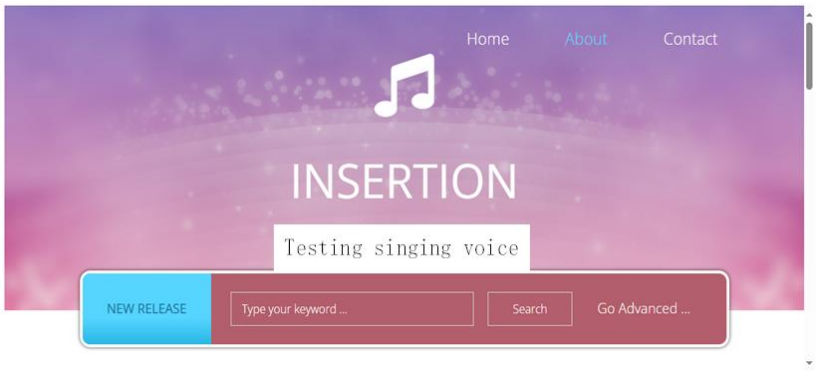

Figure 4. System content display.

The system provides the singer's training history, including daily training time, pronunciation resonance improvement, etc. Users can view their progress to adjust their training strategies.

Regarding the user interface design, the tongue position data is displayed through three-dimensional graphics; the soft palate movement is represented by a dynamic curve graph; the oral opening and closing degree is displayed with a real-time opening degree bar graph. The system also provides a real-time feedback interface to show users the current pronunciation status and optimization suggestions. For example, when incorrect tongue position or insufficient soft palate movement is detected, the system prompts the user through color changes or animations to indicate the part that needs to be adjusted, and provides specific training guidance. The interface also includes real-time voice prompts to help users make instant adjustments during training.

5.3. System deployment and integration

The system deployment phase involves integrating hardware, data processing, machine learning modules, and user interfaces and deploying them to the end user. In this phase, the following steps are performed:

The sensor is seamlessly connected to the central processing unit to ensure that data can be collected and transmitted to the processing unit in real-time. The collaborative work of hardware and software is a prerequisite for the smooth operation of the system. To improve data storage and processing capabilities, the system moves part of the computing and data storage work to the cloud. Through the cloud computing platform, the system can process a large amount of user data and

provide more personalized and efficient services. The system feeds back the analysis results to the user in real-time through the wireless transmission module, and provides personalized optimization suggestions based on the recognized pronunciation resonance features. Through the visual feedback interface, users can understand their pronunciation status in time and make adjustments.

After the system is deployed, regular version updates are carried out, and user feedback is collected. Fault detection and performance monitoring are implemented, and system health checks are performed through automated tools to ensure long-term stable operation.

6. Evaluation

6.1. Model accuracy

The feature recognition model accuracy measures the ability to recognize pronunciation resonance related features through machine learning models (such as SVM, CNN). Through cross-validation, the accuracy of the recognition results is calculated (the number of accurately classified samples/total number of samples).

The horizontal axis of **Figure 5** represents the number of training iterations, from 1 to 50. Each data point corresponds to a certain iteration of model training. The number of iterations is the number of times the model parameters are updated during the training process. The vertical axis represents the model's accuracy on the test dataset after each iteration. The value ranges from 0 to 1, representing the correct ratio of the model prediction. The accuracy starts from less than 0.60 and gradually approaches 0.95 during the training process. This means that as the model is trained and optimized, it can gradually learn more effective features, resulting in an increasing accuracy of the prediction results. Compared with traditional vocal training methods (such as recording and manual feedback), this system has better accuracy.

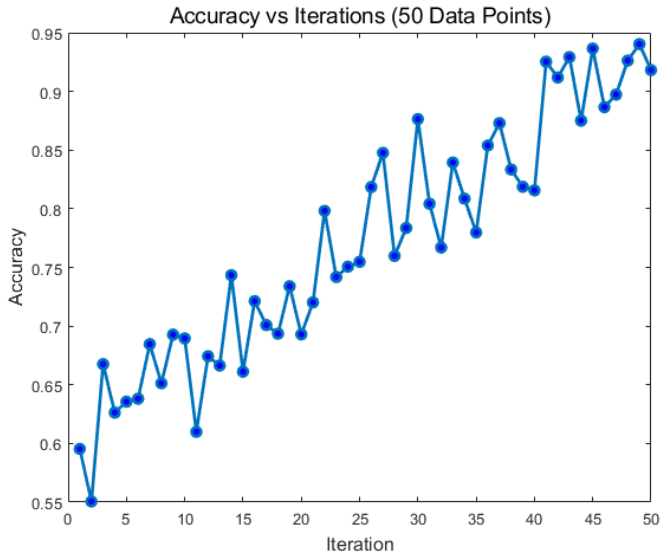

Figure 5. Model accuracy.

6.2. Response time

The real-time performance of the system is evaluated by measuring the time difference from data input to feedback output. Stress tests are performed under different hardware configurations, and response time and performance data are recorded to evaluate the system's real-time performance in various environments, ensuring that the system response time is less than 500 ms and that users can get timely feedback during the pronunciation process. Computational efficiency is improved through algorithm optimization, data preprocessing, and hardware acceleration (such as GPU computing), ensuring that the system responds quickly when processing large-scale datasets, as shown in **Table 2**.

Table 2. Real-time performance.

Timestamp (s)	Input data time (ms)	Feedback output time (ms)	Response time (ms)
0.00s	0	150	150
0.10s	100	240	140
0.20s	200	310	110
0.30s	300	420	120
0.40s	400	530	130
0.50s	500	640	140
0.60s	600	750	150
0.70s	700	860	160
0.80s	800	970	170
0.90s	900	1080	180

Table 2 records the time difference from data input to feedback output in multiple pronunciation cycles in order to measure the response performance of the system. Timestamp represents the timestamp of the singer's pronunciation. The input data time represents the time when the system receives the data (such as the tongue position, soft palate, oral opening and closing information transmitted by the sensor). The feedback output time represents the time when the system calculates and generates feedback (such as pronunciation status, resonance optimization suggestions, etc.). The response time of the system gradually remains within a certain range, and the response time is controlled within 200 ms in each time period. According to the evaluation criteria, the response time should be less than 500 ms, so this system fully meets the requirements of real-time feedback.

6.3. User satisfaction

User satisfaction reflects the friendliness, ease of use and effectiveness of feedback during the use of the system. The singers' evaluation of the system is collected through questionnaires and user interviews, and the Likert 5-point scale is used for scoring, as shown in **Table 3**.

Table 3. User satisfaction.

Question	Average Score	Score Distribution (1–5)
1. Is the system interface clear and easy to understand?	4.28	1:0, 2:2, 3:5, 4:20, 5:23
2. Is the system feedback timely?	3.96	1:1, 2:3, 3:8, 4:20, 5:15
3. Is the system's advice effective?	4.296	1:0, 2:2, 3:5, 4:22, 5:25
4. Is the system operation intuitive?	4.442	1:0, 2:1, 3:4, 4:18, 5:29
5. Does personalized feedback help adjust vocalization?	4.218	1:1, 2:2, 3:6, 4:21, 5:25
6. Does the system effectively optimize vocal resonance?	4.309	1:0, 2:2, 3:5, 4:22, 5:26
7. Would you continue using the system?	4.577	1:0, 2:0,3:2, 4:18, 5:32

Table 3 shows the feedback from users in various aspects. The average score for question 1 is 4.28, and the score distribution is 1:0, 2:2, 3:5, 4:20, 5:23. Most users (20 people give 4 points and 23 people give 5 points) are satisfied with the clarity of the interface, and almost no users give low scores (only 2 people give 2 points and 5 people give 3 points). This shows that the system interface is well designed and meets the needs of most users. The average score for question 3 is 4.296, and the score distribution is 1:0, 2:2, 3:5, 4:22, 5:25. The vast majority of users (22 people give 4 points and 25 people give 5 points) believe that the suggestions are effective and the system can help improve pronunciation, but 7 users (2 people give 2 points and 5 people give 3 points) are dissatisfied, which may be because some users feel that the suggestions are not personalized or impractical. The average score of question 6 is 4.309, and the score distribution is 1:0, 2:2, 3:5, 4:22, 5:26. Most users are satisfied with the system's optimized pronunciation resonance effect (22 people give 4 points and 26 people give 5 points), but 7 users (2 people give 2 points and 5 people give 3 points) state that the effect is average, which may be that some users do not clearly feel the optimization effect in actual application. The various functions of the system are highly evaluated by most users. Users are generally satisfied with the system's interface design, timely feedback, personalized suggestions, and intuitive operation.

The flexibility of the feedback mechanism has been further improved, especially in how to provide personalized dynamic feedback based on the user's different pronunciation levels, training progress, and resonance improvement. The depth of personalized feedback has been enhanced to provide more customized feedback content for different users based on the user's pronunciation progress, feedback history, and system analysis results.

6.4. User engagement

User engagement refers to the frequency and continuity of singers using the system. The attractiveness of the system is evaluated by recording the user's usage frequency, duration, and other data through logs.

The horizontal axis of the upper subgraph of **Figure 6** represents the month, and each month corresponds to a data point. The vertical axis represents the number of users, including the number of monthly active users (MAU) and the number of daily

active users (DAU). As time goes by, MAU is on an upward trend, indicating that more users are interacting with the system on a monthly basis. This may be because the popularity of the system is increasing or the frequency of use by users has become more stable. DAU is a smaller number relative to MAU, indicating that although users are active within each month, not every user participates in the system every day. Although DAU also shows an upward trend, it is generally lower than MAU, indicating that although the number of monthly active users of the system is gradually increasing, the daily activity of the system still has room for improvement. If the DAU/MAU ratio is low, it may mean that users only use the system occasionally in some months, rather than continuously every day. The horizontal axis of the sub-graph below **Figure 6** represents the month. The vertical axis represents the DAU/MAU ratio, that is, the ratio of the number of daily active users to the number of monthly active users, in percentage. Both the number of monthly active users and the number of daily active users show that the system continues to grow in the mid-term stage, and the number of active users has increased significantly, indicating that the popularity of the system is becoming more and more widespread. However, there is a significant decline in the last two months, and relevant investigations are needed.

Regarding the evaluation of long-term usage effects, it is planned to conduct a long-term experiment lasting several months in subsequent work to record the pronunciation improvement effects of users after using the system, and evaluate the system's long-term effectiveness through multiple experiments.

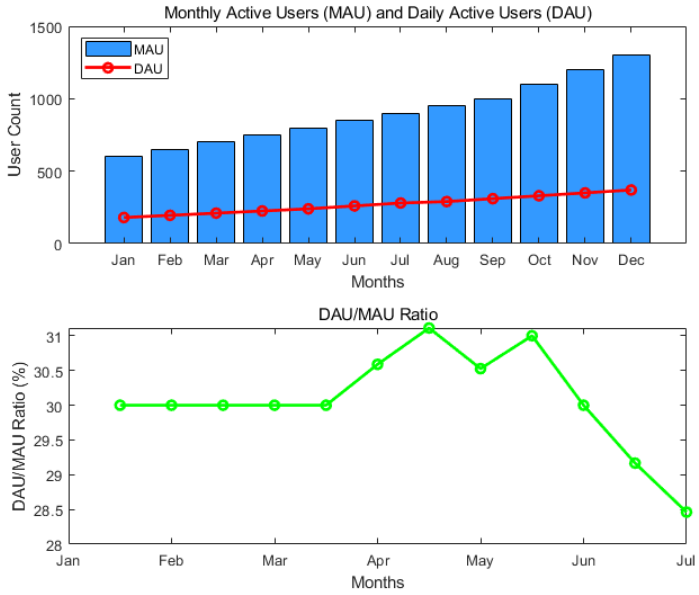

Figure 6. User engagement.

6.5. System stability

System stability refers to whether the system crashes, errors, or performance degradation occurs during long-term operation. Long-term system pressure testings are performed and the system's performance under high load, such as memory consumption, usage rate, etc., is recorded. By recording memory consumption, CPU

usage, and response time, the reliability and stability of the system under different loads are evaluated.

In **Figure 7**, the X-axis represents the system operating time (unit: s), and the blue curve represents the CPU (Central Processing Unit) usage of the system during the pressure test. CPU usage is a key indicator to measure the system's computing load, which reflects the intensity of the system's processing tasks at each time point. The red curve represents the system's memory usage. Memory usage reflects the RAM (Random Access Memory) resources consumed by the system during operation, which directly affects the system's processing power and operating efficiency. **Figure 7** shows that the CPU usage and memory usage remain within a reasonable range (50%–70% of CPU usage and 500 MB–600 MB of memory usage), and there is no significant fluctuation or surge, indicating that the system is stable during the pressure test, without crashing, freezing or performance degradation.

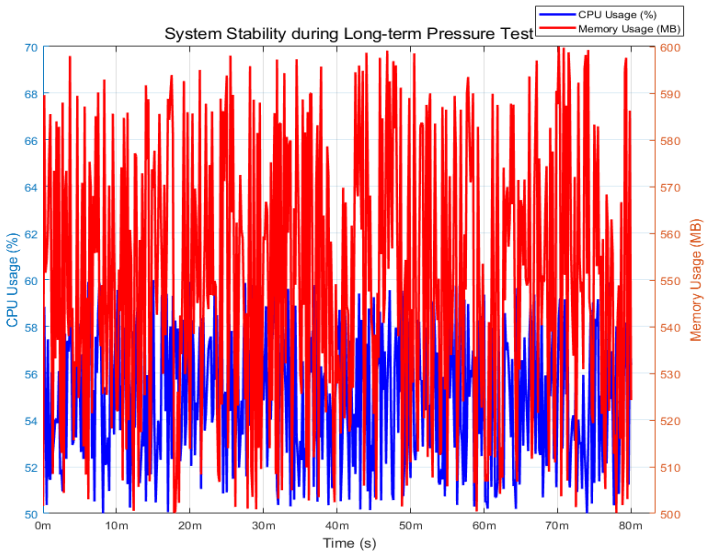

Figure 7. System stability.

7. Conclusions

This paper discusses a vocal pronunciation resonance recognition method based on oral biomechanical characteristics. By collecting oral movement data through high-precision sensors and combining machine learning algorithms such as support vector machines and convolutional neural networks, an automated pronunciation resonance recognition system is successfully developed. The system can precisely recognize oral movement patterns associated with good pronunciation resonance, provide personalized optimization suggestions for vocal training, and significantly improve the accuracy and personalized effect of pronunciation.

Through experimental verification, the method in this paper has achieved good results in oral movement feature recognition and resonance optimization, which can effectively help vocal trainers optimize the resonance area during pronunciation and improve training efficiency. However, the current method still faces problems such as large individual differences and limited training dataset size. In the future, the adaptability and accuracy of the system can be further improved by expanding the

dataset, applying more biomechanical variables, and improving the deep learning model, which can promote the development of vocal training, speech therapy, and digital music education.

Ethical approval: Not applicable.

Conflict of interest: The author declares no conflict of interest.

References

1. Jeanneteau M, Hanna N, Almeida A, et al. Using visual feedback to tune the second vocal tract resonance for singing in the high soprano range. Logopedics Phoniatrics Vocology, 2022, 47(1): 25-34.
2. Maulana I, Lestiono R, Wiraatmaja T, et al. Arriving at pronunciation accuracy in singing English songs: Hijaiyah consonants as the mediation. Satwika: Kajian Ilmu Budaya Dan Perubahan Sosial, 2021, 5(2): 303-316.
3. Titze I. Training the Vocal Instrument. Journal of Singing, 2019, 76(1): 43-46.
4. Xie X. On the integration of national singing and bel canto in vocal music teaching in normal universities. Transactions on Comparative Education, 2021, 3(3): 18-22.
5. Jang K, You K B, Park H. A Study on Correcting Korean Pronunciation Error of Foreign Learners by Using Supporting Vector Machine Algorithm. International Journal of Advanced Culture Technology, 2020, 8(3): 316-324.
6. Graf S, Schwiebacher J, Richter L, et al. Adjustment of vocal tract shape via biofeedback: influence on vowels. Journal of Voice, 2020, 34(3): 335-345.
7. Li M, Khysru K, Shi H, et al. A 3D Geometry Model of Vocal Tract Based on Smart Internet of Things. Comput. Syst. Sci. Eng., 2023, 46(1): 783-798.
8. Jin L. Research on the construction of English exams for healthcare professionals in the health sector. Journal of Commercial Biotechnology, 2022, 27(5): 175-187.
9. Duan G, Zeng J, Ji H. THE IMPORTANCE OF PSYCHOLOGICAL ANALYSIS TO VOCAL MUSIC SINGING TEACHING. Psychiatria Danubina, 2021, 33(suppl 7): 315-317.
10. Xie X. On the Integration of Bel Canto and Popular Singing in Vocal Music. International Core Journal of Engineering, 2021, 7(3): 1-4.
11. Baolin Z, Yodwised C. Constructing the Children Chorus Guidebook for Teaching Student at Jinhua Primary School in Zhengzhou, Henan Province. Asia Pacific Journal of Religions and Cultures, 2024, 8(1): 63-77.
12. Yao Y. The Study of the Effects of Yunnan Yuxi Dialect on Received Pronunciation. Theory and Practice in Language Studies, 2020, 10(6): 664-671.
13. Zhou Q, Liu S, Dai C, et al. A study of the resonance peak characteristics of Mongolian long-key ode. Advances in Education, Humanities and Social Science Research, 2023, 6(1): 291-291.
14. Deng S, Phokha P, Chiangthong N. The Creation of Vocal Performance Course Case Study Guangxi Arts University. Sciences of Conservation and Archaeology, 2024, 36(4): 139-149.
15. Kelly R. Latin Pronunciations for Singers: A Comprehensive Guide to the Classical, Italian, German, English, French, and Franco-Flemish Pronunciations of Latin. The Choral Journal, 2019, 59(11): 88-89.
16. Pocan O. Vocal Composition in Creating the Commedia dell'Arte Characters. Studia Universitatis Babes-Bolyai-Dramatica, 2023, 68(1): 175-189.
17. Gong N. Discussion on the "Localization" Singing Technique Transformation of Opera in the Background of Chinese Culture. Review of Educational Theory, 2019, 2(1): 16-20.
18. Duan L. The application of modern virtual reality technology in the teaching of vocal music. Curriculum and Teaching Methodology, 2023, 6(19): 77-81.
19. Zhang Y. Research and application of spoken English evaluation system based on biological voiceprint feature recognition. International Journal of Wireless and Mobile Computing, 2022, 22(2): 194-203.
20. Wu P, Wang R, Lin H, et al. Automatic depression recognition by intelligent speech signal processing: A systematic survey. CAAI Transactions on Intelligence Technology, 2023, 8(3): 701-711.
21. Wan M. Designing an online vocal learning based on ZigBee-enabled wireless platform. Wireless Networks, 2024, 30(1): 179-192.

22. Gao S. The Construction Of Students' Aesthetic Ability In Opera Singing Teaching Based On Vocal Aesthetics. Educational Administration: Theory and Practice, 2024, 30(4): 1723-1728.
23. Chen W L, Ye Q, Zhang S C, et al. Aphasia rehabilitation based on mirror neuron theory: a randomized-block-design study of neuropsychology and functional magnetic resonance imaging. Neural regeneration research, 2019, 14(6): 1004-1012.
24. Bin F. The Application and Inheritance of Chinese Ancient Poetry and Art Songs in Vocal Teaching. Art and Performance Letters, 2024, 5(4): 1-6.
25. Făgărășan R C. The Diction in the Art of Singing in the English Language. Învățământ, Cercetare, Creație, 2023, 9(1): 106-113.
26. Zixuan L. Research on the current situation of Chinese national vocal music education from the perspective of educational psychology. Frontiers in Art Research, 2021, 3(6): 48-58.
27. Chiaramonte R, Di Luciano C, Chiaramonte I, et al. Multi-disciplinary clinical protocol for the diagnosis of bulbar amyotrophic lateral sclerosis. Acta Otorrinolaringologica (English Edition), 2019, 70(1): 25-31.
28. Surahman A. An analysis of voice spectrum characteristics to the male voices recording using praat software. IJFL (International Journal of Forensic Linguistic), 2021, 2(2): 69-74.
29. Fagarașan R C. Pronunciation in the Art of Singing in English and French During the Middle Ages. Învățământ, Cercetare, Creație, 2023, 9(1): 97-105.
30. Janaswamy R, Vasudev S K. Vocal Training and Practice Methods: A Glimpse on the South Indian Carnatic Music. International Journal of Humanities and Social Sciences, 2020, 14(12): 1281-1285.
31. Moran W. Discussion on the Characteristics and Skills of Tenor Singing in Different Musical Styles. Art and Performance Letters, 2024, 5(3): 97-102.
32. Chen S, Gan R. Analysis of the Acoustic Characteristics of the Stops in Northern Yi Dialects. Journal of Sociology and Ethnology, 2024, 6(3): 19-28.
33. Junjie L, Meesorn P. Overview of the Bel Canto and Development. Journal of Modern Learning Development, 2024, 9(8): 769-777.
34. Suhery D, Idami Z, Wati S. The Phonological Interference of Acehnese in Pronouncing Indonesian Language. Journal of Languages and Language Teaching, 2024, 12(1): 120-135.
35. Wu H X, Li Y, Ching B H H, et al. You are how you speak: The roles of vocal pitch and semantic cues in shaping social perceptions. Perception, 2023, 52(1): 40-55.
36. Duyen T M T. Exploring Phonetic Differences and Cross-Linguistic Influences: A Comparative Study of English and Mandarin Chinese Pronunciation Patterns. Open Journal of Applied Sciences, 2024, 14(7): 1807-1822.
37. Sun X. The Training Value of Handel's Vocal Works in Vocal Pedagogy. Research and Advances in Education, 2023, 2(3): 19-22.
38. Hasanah N, Rahmadhani P R, Lubis Y. Inconsistency of Some Consonants in English. Jurnal Pendidikan Rafflesia, 2023, 1(2): 31-34.