SIN-CHN
SCIENTIFIC PRESS PTE.LTD.

Molecular & Cellular Biomechanics 2025, 22(2), 937.
https://doi.org/10.62617/mcb937

Article

Impact of biomechanical properties of tongue muscles on accuracy of English vowel pronunciation

Ping Zhang[1], **Xiaoguang Chen**[2,*]

[1] School of Education and Foreign Languages, Wuhan Donghu University, Wuhan 430212, China
[2] College of Applied Technology, Dalian Ocean University, Dalian 116300, China
*** Corresponding author:** Xiaoguang Chen, chenxiaoguang@dlou.edu.cn

CITATION

Zhang P, Chen X. Impact of biomechanical properties of tongue muscles on accuracy of English vowel pronunciation. Molecular & Cellular Biomechanics. 2025; 22(2): 937.
https://doi.org/10.62617/mcb937

ARTICLE INFO

Received: 28 November 2024
Accepted: 11 December 2024
Available online: 21 January 2025

Abstract: Pronunciation is a complex physiological process. Traditional research usually uses static pronunciation tests and fails to observe the dynamic changes of tongue muscles during pronunciation. This paper aims to comprehensively analyze the structure and function of tongue muscles and their role in English vowel pronunciation from the perspective of tongue muscle biomechanics, and provide a systematic framework for understanding. This paper designs multiple pronunciation tasks to evaluate participants' pronunciation accuracy and dynamic changes of tongue muscles. Through multi-modal technology, dynamic images and electromyographic signals of the tongue are synchronously acquired to analyze the precise relationship between tongue movement and muscle activity in the pronunciation of English vowels. A tongue biomechanical model is constructed based on finite element analysis and Hill model to precisely simulate the mechanical response of tongue muscle activity and tongue position changes during pronunciation. The experimental results show that there is a significant negative correlation between electromyographic activity and pronunciation quality. The closer the correlation coefficient is to −1, the higher the consistency. The tongue is positioned higher and forward during pronunciation, making it easier to control, so that the pronunciation can be more accurate with less deviation. The greater the movement and flexibility of the tongue, the better it is able to form clear vowel pronunciations. In short, the tongue muscles achieve precise control of tongue position through the coordinated action of internal and external muscles during vowel pronunciation, which is beneficial to improving pronunciation accuracy.

Keywords: English vowel pronunciation; biomechanical properties of tongue muscles; pronunciation task; finite element analysis; data acquisition

1. Introduction

In language communication, the precision and accuracy of pronunciation are key factors that directly affect the effectiveness and clarity of interpersonal communication. When learning foreign languages such as English, the accurate pronunciation of vowels is particularly important because it is related to the listeners' understanding of the message. However, traditional research methods mainly rely on static evaluation and cannot effectively capture the dynamic changes of tongue muscles during pronunciation, which limits the comprehensive understanding of the biomechanical characteristics of tongue muscles and their impact on pronunciation accuracy.

To address these challenges, this paper applies advanced biomechanical research methods to focus on analyzing how the dynamic performance of tongue muscles affects the pronunciation precision of English vowels. By combining dynamic imaging analysis technology with electromyogram (EMG), the activity of tongue muscles is monitored in real time to capture even the slightest pronunciation changes, thereby improving the detail of the data. This multifaceted research method not only provides more precise and reliable biomechanical data, but also helps to clarify the complex relationship between tongue muscle function and pronunciation results, providing a broader perspective for understanding the physiological mechanism of pronunciation control.

This paper aims to comprehensively analyze the structure and function of tongue muscles and their role in the pronunciation of English vowels, and to establish a biomechanical model. Through dynamic image analysis and EMG monitoring, the dynamic changes of tongue muscles during pronunciation are tracked, and the tongue muscle activities of different participants are deeply analyzed to reveal the specific impact of their biomechanical characteristics on pronunciation accuracy. This provides a theoretical basis and practical guidance for language learning and speech therapy.

2. Related work

Pronunciation accuracy has an important impact on language learners' communication and comprehension abilities [1,2]. An increasing number of scholars have conducted research from multiple dimensions to improve pronunciation accuracy [3,4]. Pakpahan [5] emphasized the significance of learning English phonology in improving pronunciation accuracy, but did not discuss specific teaching strategies in detail. Garita Sanchez [6] investigated students' and teachers' views on English vowel pronunciation issues and, although providing valuable insights, lacked quantitative data to support his conclusions. Accurate pronunciation is the key to learning English, because incorrect pronunciation can lead to communication barriers. Therefore, Kobilova [7] analyzed the importance of pronunciation and discussed English pronunciation and communication, but did not delve into the specific impact of pronunciation problems on communication effectiveness. Floare Bora [8] studied the effectiveness of a blended drama approach in improving oral accuracy. However, there were limitations in sample selection and experimental design that might affect the generalizability of the results. Sardegna [9] proposed a strategy-based model for teaching English pronunciation. Although it provided empirical support, it lacked in-depth analysis of actual cases of implementing the model. Overall, although these studies provide diverse perspectives on English pronunciation teaching, they still have certain shortcomings in methodology, sample selection, and empirical support.

The relationship between tongue muscles and speech pronunciation has received attention. Fogarty [10] explored the contraction, fatigue, and fiber type characteristics of tongue muscles. By experimentally measuring the contraction force and fatigue tolerance of tongue muscles under different stimulation conditions, it was found that tongue muscles have distinct contraction characteristics and fatigue patterns. These findings provided a physiological basis for understanding the role of tongue muscles in speech and swallowing. Goto [11] used the tongue muscles as a model to analyze the dynamics of head muscle regeneration and observed the regeneration process of the tongue muscle after injury to reveal its regeneration mechanism. Szelenyi [12] stimulated the tongue muscles in different parts and found that the reaction time and intensity were significantly different, indicating the complexity of the tongue muscles in neural reflexes. Zhu Mengxian [13] reviewed the application of ultrasound imaging technology in analyzing the movement patterns of the tongue during pronunciation, and pointed out that this technology can record the movement trajectory and morphological changes of the tongue in real time, providing a new perspective for the study of pronunciation mechanisms. Jia [14] focused on the teaching challenges of pronouncing the /a/ sound for Chinese Thai language learners. The analysis showed that systematic pronunciation training significantly improved learners' pronunciation accuracy, providing practical guidance for foreign language teaching. Taken together, these studies together highlight the importance of tongue muscles in terms of physiological function and regeneration potential and provide important contributions to the biomechanical properties of tongue muscles. This paper combines biomechanical analysis with a dynamic experimental method to explore the impact of tongue muscles on the pronunciation of English vowels.

3. Biomechanical impact of tongue muscles

Tongue muscles are divided into two categories: Intrinsic and extrinsic [15]. Intrinsic tongue muscles are mainly responsible for the shape of the tongue, while extrinsic tongue muscles control the position and movement of the tongue. Intrinsic tongue muscles include vertical muscles and transverse muscles. Vertical muscles control the extension and contraction (length and height) of the tongue, while transverse muscles adjust the width of the tongue.

The extrinsic tongue muscles include the tongue surface muscles, sublingual muscles, and supralingual muscles. Supralingual muscles and sublingual muscles move the tongue up and down respectively, while the tongue surface muscles control the movement of the tongue tip.

The change of tongue position during pronunciation is achieved by the coordinated action of multiple muscles, and the functional performance of tongue muscles when pronouncing different vowels varies significantly. When pronouncing front vowels such as /iː/, the tip of the tongue rises upward, and the tongue body stretches and rises slightly. At this time, the vertical and transverse muscles in the tongue muscles show a high degree of coordination, ensuring that the tongue body maintains the appropriate curvature to produce a clear sound quality.

When pronouncing the back vowel /uː/, the back of the tongue moves upward and backward, and the sublingual muscles and tongue surface muscles of the extrinsic tongue muscles play a leading role in maintaining the stability of the tongue position and avoiding unnecessary tongue movement.

EMG data demonstrate that the electrical activity of different tongue muscles show obvious timing differences at different stages of pronunciation. When pronouncing the vowel /ɑː/, the activity intensity of the extrinsic tongue muscles is high to maintain the low position of the tongue. When pronouncing the /eɪ/, the activity of the intrinsic tongue muscles appears more frequent to ensure the raising and lowering movement of the tongue.

The coordination of tongue muscles is crucial during pronouncing, especially during transitions. When transitioning from the vowel /iː/ to /æ/, the intrinsic and extrinsic tongue muscles must work in harmony to ensure a smooth transition of tongue position and avoid unclear transition pronunciations.

Non-native English learners have low coordination and accuracy in the pronunciation of certain vowels. The synchronization between the intrinsic and extrinsic tongue muscles is poor, especially when pronouncing /æ/ and /ɑː/. The movement path of the tongue is not clear, resulting in reduced pronunciation accuracy. This shows that the precise control of tongue muscles is closely related to language habits.

The EMG signal analysis formula is [16,17]:

$$E = \int_0^T A_m(t)dt \tag{1}$$

E represents the strength of the EMG signal within a certain period of time. $A_m(t)$ represents the strength of the EMG activity at time t. T represents the time interval of interest.

Combining EMG signals and dynamic imaging data, it is found that there is a significant relationship between the functional changes of tongue muscles and pronunciation accuracy. The tongue position change model formula is as follows:

$$\Delta P = k_1 \times \Delta X + k_2 \times \Delta Y \tag{2}$$

ΔP represents the change of tongue position during pronunciation. k_1 and k_2 are adjustment parameters, representing the impact of the intrinsic and extrinsic tongue muscles. ΔX and ΔY represent the displacement of the tongue tip and tongue body, respectively.

Excessive tension or relaxation of the tongue muscles may lead to unclear pronunciation: One is stiffness and unnaturalness, and the other is fuzzy pronunciation. Especially when certain muscle groups fail to adjust or work together in time during pronunciation, the phonemes of pronunciation may deviate. The accuracy of pronunciation depends not only on the accurate control of tongue position but also on the coordination of muscle activity.

3.1. Design of pronunciation tasks

Through the university’s linguistics department and language learning center, a recruitment notice is published to invite native English speakers and non-native English learners (from other language backgrounds) to sign up for the study. The age requirement is 18–35 years old, with no oral structure or function disorders. The invited non-English learners must have a certain level of English speaking ability (intermediate level or above).

Common vowel phonemes are selected in the vowel selection and classification task, covering different oral openings, tongue positions and tongue shape changes. These vowels include front vowels, back vowels, rounded vowels, and unrounded vowels [18,19]. Task 1 is the clear pronunciation of a single vowel. The pronunciation task of each vowel is repeated three times at a fixed time interval (1 min) to reduce the influence of memory and pronunciation fatigue. Task 2 is short sentence pronunciation, which contains natural conversational sentences with a variety of vowels. Before each pronunciation task, participants first do a brief warm-up exercise to familiarize themselves with the task requirements and equipment and ensure the natural activity of the tongue muscles. There is a 30 s interval between each task to give participants sufficient rest time to prevent tongue muscle fatigue.

Participants first pronounce the words according to their actual language habits without external intervention. Then they pronounced them individually in the recording studio, clearly and slowly pronouncing each vowel in turn. Each vowel is pronounced for about 1 s to ensure that the tongue muscle activity time is clear enough. The order of the tasks is randomized to avoid order effects.

After completing the above two tasks, each participant is required to perform a challenging task, which requires the participants to switch vowels freely in fast pronunciation and transitional syllables. Create a set of phrases containing common consonant sounds (“black cat”, “fast train”) and ask participants to pronounce them quickly within a limited time. Gradually increase the difficulty, from simple consonant sounds to complex three consonant sounds. During the process of connecting consonants, participants are required to focus on maintaining the fluency and clarity of their voices. Provide a series of sentences and ask participants to pronounce them under different emotions (happiness, anger, sadness). Observe the impact of emotions on pronunciation speed, stress, and pitch, as well as changes in tongue muscle activity.

A unified pronunciation assessment standard is set. The pronunciation of each vowel is compared with the standard pronunciation at the phoneme level to assess its similarity. The standard pronunciation is recorded by a native English speaker, and the pitch, duration, and frequency characteristics of each vowel are calculated using acoustic analysis software. After the pronunciation tasks are completed, the participants’ pronunciation accuracy is analyzed based on the difference between the audio signal and the standard pronunciation, and compared with the tongue muscle activity data.

3.2. Data acquisition

The data acquisition process is shown in **Figure 1**.

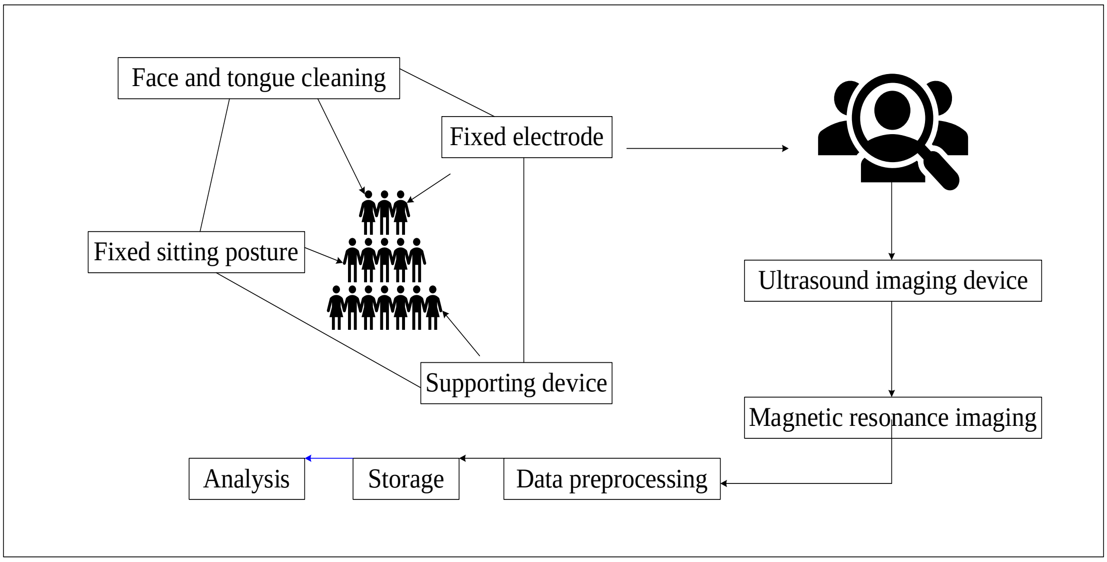

Figure 1. Data acquisition process.

Figure 1 shows that participants clean their faces and tongues before data acquisition. Medical sandpaper is used to lightly exfoliate to ensure that the electrodes are firmly attached and reduce signal noise. Dual-channel surface electrodes are used, with key locations of the intrinsic and extrinsic tongue muscles as target points, and electrodes are fixed on the upper and lower sides of the tongue respectively. The placement of electrodes is based on anatomical landmarks to ensure coverage of the main active areas of the tongue, while avoiding attachment to facial or jaw muscles outside the tongue to maximize signal specificity.

To complete pronunciation tasks in different complex language environments, participants are required to maintain a fixed sitting posture and gently secure their lower jaw with a support device to avoid unnecessary head or jaw movements interfering with tongue movement.

Based on high-precision dynamic image analysis and EMG [20,21] recording, combined with multi-modal technology, real-time dynamic information of tongue muscles during the pronunciation of English vowels is obtained. Ensure precise alignment of data through timestamp and standardized statement calibration. With the support of environmental control and digital filters, noise interference is reduced, and real-time adjustment is made through adaptive filtering to improve data quality. Combining multi sample analysis and machine learning algorithms to enhance data stability, and utilizing deep learning to improve the accuracy of signal pattern recognition.

The camera (300 fps) is set in front of the participants, using a side view to capture the dynamic movement of the tongue. Ultrasound imaging and magnetic resonance imaging (MRI) are used to record the tongue shape and movement trajectory in real time. The surface EMG device used is a high-sensitivity acquisition system to precisely capture the electrical activity signals of the tongue muscles. The sampling frequency of the device is set to 1 kHz to ensure that the temporal and spatial resolutions meet the analysis requirements of the dynamic changes of the tongue muscles.

During the acquisition process, real-time filtering technology is used to reduce environmental electromagnetic interference, and a bandpass filter [22,23] is selected to process the EMG signal to remove baseline drift and low-frequency noise. The formula is as follows:

$$EMG_{filtered}(t) = EMG(t) * H(f), H(f) = \begin{cases} 1, & f_{low} \leq f \leq f_{high} \\ 0 & otherwise \end{cases} \quad (3)$$

$H(f)$ refers to the filter transfer function, and there are $f_{low} = 10HZ$ and $f_{high} = 500HZ.$

All acquired raw data is stored in a high-precision format, and dynamic images are saved in DICOM (digital imaging and communications in medicine) format [24]. EMG signals are stored as raw EMG files. All data is encrypted and backed up to multiple servers to ensure the security and integrity of subsequent analysis.

The contraction and relaxation patterns of each muscle during pronunciation are calculated using time domain and frequency domain analysis methods [25,26].

The Pearson correlation coefficient is used to calculate the correlation between tongue muscle contraction and movement trajectory. The formula is as follows [27,28]:

$$r = \frac{\sum_{i=1}^{N}(EMG_{RMS}, i - \overline{EMG}_{RMS})^2 (v_i - \bar{v})}{\sqrt{\sum_{i=1}^{N}(EMG_{RMS}, i - \overline{EMG}_{RMS})^2 \sum_{i=1}^{N}(v_i - \bar{v})^2}} \quad (4)$$

$\overline{EMG}_{RMS}$ and $\bar{v}$ are the mean values of EMG signal intensity and velocity, respectively.

3.2. Establishment of biomechanical model

Biomechanics technology has the ability to evaluate and monitor tongue muscle function. In speech disorder research, it is used to analyze in detail the movements of the oral cavity, tongue, and vocal cords. By utilizing high-precision motion capture and electromyography techniques, subtle changes in the articulatory organs are displayed. In the study of articulation disorders, biomechanical techniques help evaluate the functional abnormalities of the tongue and oral muscles during individual pronunciation, provide accurate data support, and assist in developing personalized treatment plans. To more precisely analyze the mechanical changes of tongue muscles during pronunciation, a biomechanical model of tongue muscles with multi-factor interactions is constructed, as shown in **Figure 2**. The model is based on the dynamic characteristics of muscle activity and simulates the mechanical performance of tongue muscles under different pronunciation conditions.

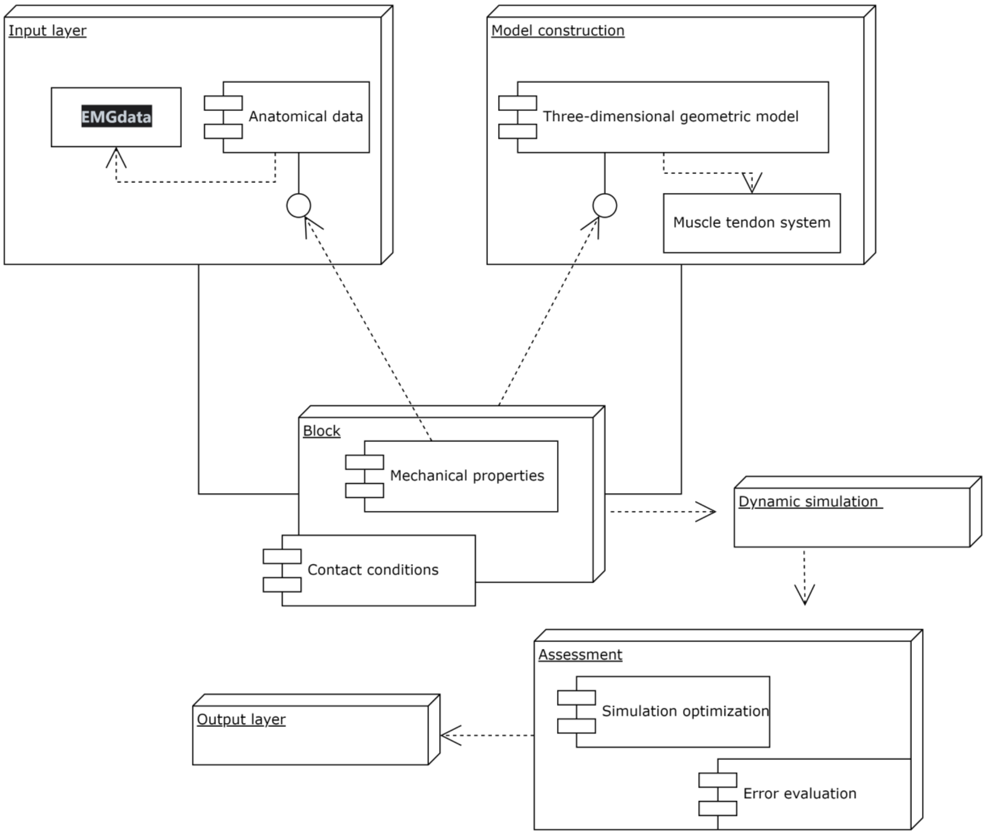

Figure 2. Architecture of tongue muscle biomechanical model.

Figure 2 shows the architecture of the tongue muscle biomechanical model, which mainly includes the input layer, model construction, mechanical properties, contact conditions, dynamic simulation, assessment, and output layer. Based on anatomical data, the finite element analysis (FEA) method is used to establish a three-dimensional geometric model of the tongue [29,30]. The model takes into account the actual physiological characteristics of each muscle group of the tongue, and realizes the individual differences of the tongue through calibration and dynamic adjustment.

The Hill model of the muscle-tendon system is used to describe the process of muscle contraction and tension generation. By inputting EMG data, the model can reflect the intensity of muscle activity in real time and simulate the effect of muscle contraction on tongue position. The calculation of muscle contraction force is [31,32]:

$$F = a \cdot (L_{opt} - L) + b \cdot \frac{dL}{dt} \tag{5}$$

F represents the contraction force generated by the muscle; a is a constant related to the muscle contraction characteristics; L_{opt} is the optimal length of the muscle; b is a constant related to the contraction speed; $\frac{dL}{dt}$ is the rate of change of muscle length.

To capture the force output of the tongue muscles during pronunciation, the model takes into account the force-length characteristics, force-velocity characteristics and synergy between muscles. The force output of each muscle is combined with its contraction velocity, length change and interaction force to calculate the mechanical response of each part of the tongue.

According to the physiological requirements of different pronunciations, the contact conditions between the tongue and the inside of the oral cavity (the contact force between the tongue and the hard palate and soft palate) are set. In the simulation, the mechanical response of the tongue depends on the interaction between these contact forces and the force exerted by the tongue muscles.

Through time-stepping dynamic simulation, the contraction and relaxation of tongue muscles during pronunciation and their changes in tongue position and shape are simulated. The change in tongue position at each stage affects the activity pattern of tongue muscles, and the mechanical response of the muscles in turn affects the precise positioning of the tongue.

In different pronunciation tasks, simulation optimization technologies (genetic algorithm [33,34] and particle swarm optimization algorithm [35,36]) are used to adjust the various parameters in the model and optimize the activity pattern of the tongue muscles to ensure that the simulated tongue position and shape can restore the actual pronunciation situation to the greatest extent. The simulation optimization model is defined as:

$$\overset{\min}{\underset{\theta}{}}\sum_{j=1}^{N}(O_j - S_j(\theta))^2 \tag{6}$$

θ is the model parameter that needs to be optimized; O_j is the actual pronunciation effect of the j-th one; $S_j(\theta)$ is the output of the model with parameter θ.

The error between the model output and the actual pronunciation effect is evaluated by combining the EMG, dynamic images, and acoustic analysis data of each pronunciation task. The model parameters are continuously adjusted through the error back propagation algorithm to achieve higher-precision simulation [37-38].

The error between the model output and the actual pronunciation effect is evaluated by combining the EMG, dynamic images, and acoustic analysis data of each pronunciation task. The model parameters are continuously adjusted through the error back propagation algorithm to achieve higher-precision simulation [37,38].

4. Impact assessment and analysis

60 participants between 18 and 35 years old are taken as an example, as shown in **Table 1**.

Table 1. Details of participants.

Group number	Language background	Age	Gender	Number
N1-1	N1	18–35	Male	15
N1-2			Female	15
N2-1	N2		Male	15
N2-2			Female	15

In **Table 1**, N1 refers to native English speakers, and N2 refers to non-native English learners. Three common vowel pronunciations are selected for testing, and the contraction of each group of participants is recorded, and the average value is taken, as shown in **Table 2**.

Table 2. Results of tongue muscle activity pattern analysis.

Language background	Group number	Vowel pronunciation	S1 (μV)	S2 (μV)	S3 (ms)
N1	N1-1	/a/	170	150	120
		/e/	160	140	110
		/i/	180	155	125
	N1-2	/a/	175	145	115
		/e/	165	135	105
		/i/	185	160	130
N2	N2-1	/a/	130	120	150
		/e/	125	115	140
		/i/	135	125	145
	N2-2	/a/	128	118	155
		/e/	122	112	145
		/i/	132	120	150

In **Table 2**, S1 and S2 refer to the contraction amplitudes of the intrinsic and extrinsic tongue muscles, respectively, and S3 refers to the contraction duration. From the data, it can be seen that there are significant differences in the activity patterns of the intrinsic and extrinsic tongue muscles when participants of N1 and N2 procounce different vowels. The contraction amplitude of the intrinsic and extrinsic tongue muscles of N1-1 and N1-2 is greater than that of N2-1 and N2-2. In addition, the contraction duration data also shows that participants in the N1 group takes shorter time, further emphasizing the tongue flexibility of native English speakers when making pronunciations.

Add a detailed comparison between the model simulation results and actual experimental data to verify the accuracy of the model's predictive ability, as shown in **Table 3** below. Among them, S1-M, S2-M, and S3-M respectively refer to the simulation results of the model.

Table 3. Comparison between model simulation results and actual experimental data.

Language background	Group number	Vowel pronunciation	S1 (μV)	S1-M (μV)	S2 (μV)	S2-M (μV)	S3 (ms)	S3-M (ms)
N1	N1-1	/a/	170	168	150	150	120	120
		/e/	160	158	140	142	110	110
		/i/	180	180	155	154	125	126
	N1-2	/a/	175	174	145	146	115	115
		/e/	165	164	135	135	105	106
		/i/	185	185	160	160	130	129
N2	N2-1	/a/	130	131	120	119	150	150
		/e/	125	125	115	114	140	138
		/i/	135	135	125	126	145	144
	N2-2	/a/	128	128	118	120	155	155
		/e/	122	123	112	110	145	146
		/i/	132	133	120	121	150	149

Table 3 provides a detailed comparison between experimental data and model simulation results, demonstrating the predictive ability and accuracy of the model. The predicted results of the model are not significantly different from the experimental data.

The software is used to analyze the highest and lowest points of the participants' tongues when pronouncing vowels, and to calculate the range of tongue position, flexibility, and peak values of the internal and external tongue muscles, as shown in **Figure 3**.

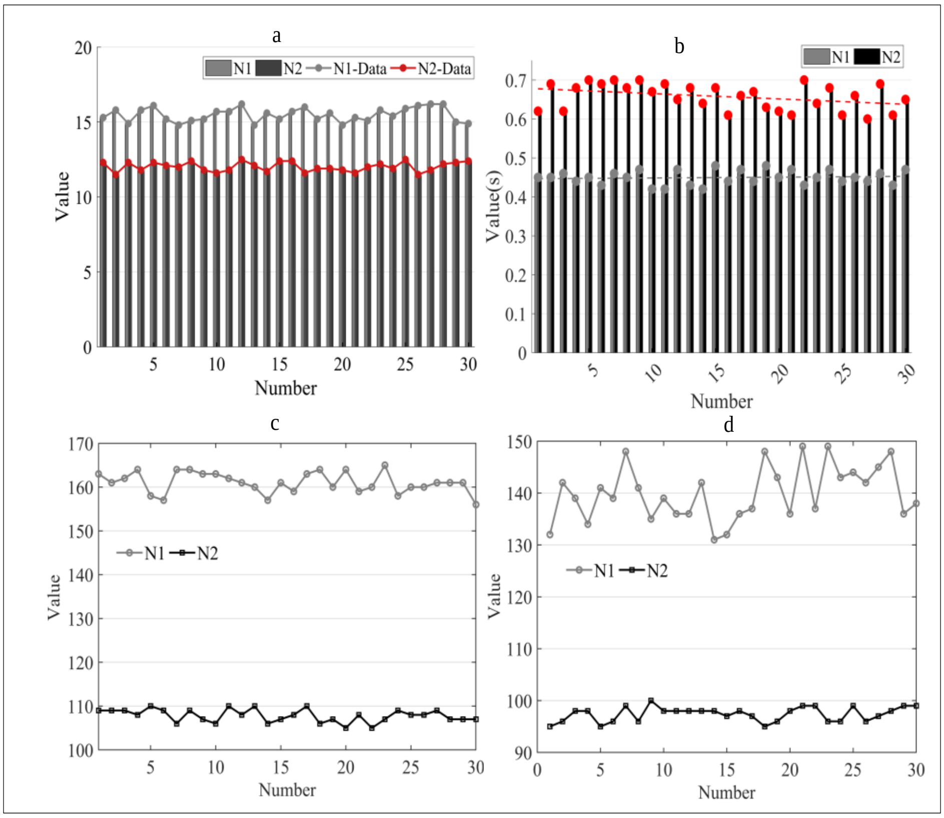

Figure 3. Tongue pronunciation details of English native speakers and non-native English learners. (**a**) tongue position range of English native speakers and non-native English learners (unit: mm); (**b**) flexibility of English native speakers and non-native English learners (unit: s); (**c**) peak value of intrinsic tongue muscles of English native speakers and non-native English learners (unit: µV); (**d**) peak value of extrinsic tongue muscles of English native speakers and non-native English learners (unit: µV).

Figure 3 shows the tongue pronunciation details of native English speakers and non-native English learners. Among them, in **Figure 3a**, the tongue position range of the participants of N1 when pronouncing vowels is between 14.8 mm and 16.2 mm, while the range of the participants of N2 is smaller, between 11.5 mm and 12.5 mm. In **Figure 3b**, the flexibility of the participants of N1 in pronouncing vowels, that is, the time of tongue movement, is between 0.42 s and 0.48 s, while the time of the participants of N2 is longer, between 0.6 s and 0.7 s. In **Figure 3c**, the peak value of the intrinsic tongue muscles of the participants of N1 when pronouncing vowels is between 156 µV and 165 µV, while the peak value of the intrinsic tongue muscles of the participants of N2 is generally smaller, between 105 µV and 110 µV. In **Figure 3d**, the peak value of the extrinsic tongue muscle of the participants of N1 is between 131 µV and 149 µV, and the peak value of the extrinsic tongue muscle of the participants of N2 is generally smaller, between 95 µV and 100 µV.

The pronunciation accuracy of these participants is scored on a scale of 0 to 1. The pronunciation deviation and the correlation with electromyographic activity are also recorded, as shown in **Figure 4** for details.

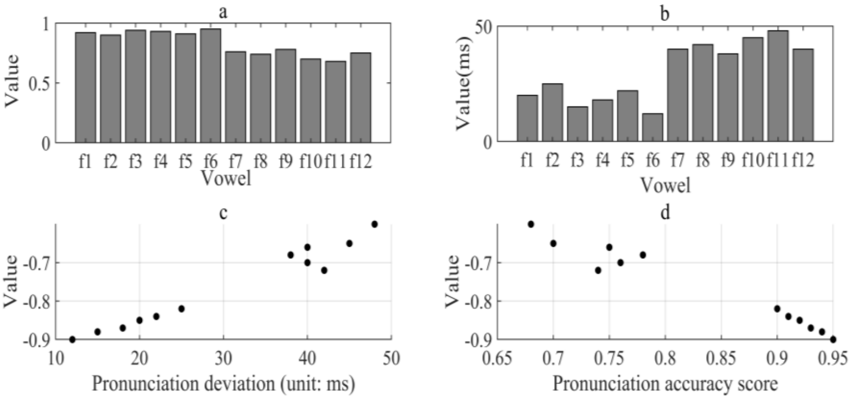

Figure 4. Analysis of the relationship between tongue muscle activity and pronunciation accuracy. (**a**) pronunciation accuracy score; (**b**) pronunciation deviation (unit: ms); (**c**) correlation between pronunciation deviation and electromyographic activity; (**d**) correlation between pronunciation accuracy score and electromyographic activity.

In **Figure 4**, the overall performance of f1 to f6 is outstanding. According to **Figure 4a**,**b**, it can be found that the pronunciation accuracy and pronunciation deviation of English native speakers are better than those of the remaining non-native English learners. According to the data in **Figure 4c**,**d**, it is found that the correlation of electromyographic activity of native English learners is close to −1. There is a significant negative correlation between electromyographic activity and pronunciation quality. The closer the correlation coefficient is to −1, the higher the consistency between the two. In addition, for all participants, the performance in pronouncing /i/ is generally better than that of the other two vowels. This is related to the pronunciation characteristics of /i/, because /i/ is a high front vowel, and the tongue position is high and forward when pronouncing it, which is easier to control, resulting in more accurate pronunciation and less deviation.

The specific references of f1 to f12 are shown in **Table 4**.

Table 4. Specific references of f1 to f12.

Language background	Group number	Vowel pronunciation	Label
N1	N1-1	/a/	f1
		/e/	f2
		/i/	f3
	N1-2	/a/	f4
		/e/	f5
		/i/	f6
N2	N2-1	/a/	f7
		/e/	f8
		/i/	f9
	N2-2	/a/	f10
		/e/	f11
		/i/	f12

Table 4 lists the specific references of f1 to f12 in detail. In summary, the muscle activity of participants with more accurate pronunciation and less deviation is more closely related to the standard pronunciation pattern when they pronounce. There is a certain positive correlation between the contraction amplitude of the intrinsic tongue muscles and the accuracy of pronunciation. The greater the contraction amplitude of the intrinsic tongue muscles, the stronger the movement ability and flexibility of the tongue, thus being able to better form clear vowel pronunciation.

The contraction and relaxation periods of the tongue muscles of the experimental subjects are recorded and analyzed, as shown in **Figure 5**.

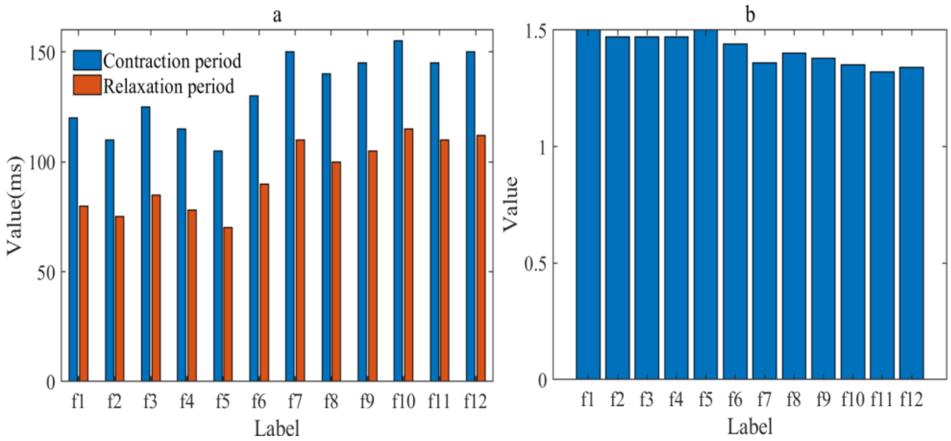

Figure 5. Analysis of tongue muscle contraction and relaxation periods. (a) Contraction and relaxation period of tongue muscle; (b) Contraction and relaxation ratio of tongue muscle.

In **Figure 5**, the contraction period of f1 when pronouncing the /a/ sound is 120 ms, and the relaxation period is 80 ms, with a contraction-relaxation ratio of 1.5, indicating that its contraction time is relatively long when pronouncing the sound, which is suitable for producing clear vowels. The contraction period of f4 when pronouncing /a/ is 115 ms, and the relaxation period is 78 ms, with a contraction-relaxation ratio of 1.47, showing a similar pronunciation pattern. The contraction period of f7 when pronouncing /a/ is 150 ms, and the relaxation period is 110 ms, with a contraction-relaxation ratio dropping to 1.36, indicating that it takes a longer time to complete the pronunciation, which may affect the fluency of pronunciation. In general, with the extension of the contraction period and the increase of the relaxation period, the contraction-to-relaxation ratio of the participants gradually decreases, affecting the accuracy and fluency of their pronunciation. Therefore, in the training of tongue muscles, finding the appropriate contraction-to-relaxation ratio is the key to improving the quality of pronunciation.

The cross-correlation analysis method is used to evaluate the synchronization between the activity signals of the intrinsic and extrinsic tongue muscles. The quantitative comparison of the time lag and synchronization between muscle activities further reveals the relationship between the efficiency of tongue muscle synergy and the clarity and accuracy of pronunciation. The details are shown in **Table 5**.

Table 5. Results of cross-correlation analysis.

Language background	Group number	Vowel pronunciation	Label	r1	r2	*p*
N1	N1-1	/a/	f1	0.92	5	< 0.01
		/e/	f2	0.90	4	< 0.01
		/i/	f3	0.95	3	< 0.01
	N1-2	/a/	f4	0.88	6	< 0.01
		/e/	f5	0.85	5	< 0.01
		/i/	f6	0.91	4	< 0.01
N2	N2-1	/a/	f7	0.70	8	< 0.05
		/e/	f8	0.72	7	< 0.05
		/i/	f9	0.68	6	< 0.05
	N2-2	/a/	f10	0.65	9	< 0.05
		/e/	f11	0.62	10	< 0.05
		/i/	f12	0.66	8	< 0.05

In **Table 5**, r1 refers to the synchronization between intrinsic and extrinsic tongue muscles; r2 refers to the time lag (ms); *p* refers to the significant correlation. **Table 5** shows the results of the cross-correlation analysis between the activity signals of the intrinsic and extrinsic tongue muscles when the participants pronounce different vowels. The synchronization of f1 when pronouncing /a/ is 0.92; the time lag is 5 ms; the correlation significance is less than 0.01, indicating that the synergy between the intrinsic and extrinsic tongue muscles is very high. In comparison, f4 has a synchronization of 0.88 and a time lag of 6 ms, which, although slightly lower in synchronization, still show a significant correlation. The results of N2 are relatively poor, with a long time lag, both showing that the efficiency of tongue muscle synergy is low, which affects the accuracy and clarity of the pronunciation. Overall, higher synchronization and smaller time lag are positively correlated with pronunciation accuracy and clarity, reflecting that good synergy between intrinsic and extrinsic tongue muscles is a key factor in achieving high-quality pronunciation.

Table 6. Comparison between different models.

Evaluation indicators	Biomechanical model of tongue muscles	Nonlinear elastic model	Machine learning based biological models
Prediction accuracy (%)	95	85	92
Computational complexity (Unit: GFLOPS)	20	50	15
Data requirements (Unit: GB)	1	2	10
Model universality (Score: 1–5)	4.5	3	4.5
Real time application capability (delay: ms)	50	200	40
Ability to capture pronunciation details (Score:1–5)	4.5	3.5	4.5
Development and maintenance complexity (Score: 1–5)	3	4	4.5
Robustness (Score: 1–5)	4.5	3.5	4
Efficiency of resource constrained device operation (Score: 1–5)	4.5	2.5	4

In order to further highlight the advantages of our model, we compared it with nonlinear elastic models and machine learning based biological models, as shown in **Table 6** above.

In **Table 6**, the biomechanical model of the tongue muscle achieved a prediction accuracy of 95%, significantly better than the 85% of the nonlinear elastic model, and higher than the 92% of the machine learning based biological model, indicating its high accuracy in modeling the relationship between tongue muscle activity and pronunciation. The data from other aspects overall demonstrate the outstanding performance of the tongue muscle biomechanical model, which is particularly suitable for language research and speech therapy fields that require high precision and real-time feedback.

A real-time pronunciation training tool based on tongue muscle activity feedback can target non-native speakers' pronunciation practice and individuals with specific pronunciation disorders. Accurately capturing and analyzing the movement of the tongue through a biomechanical model of the tongue muscles, providing real-time visual and auditory feedback to help users identify and correct pronunciation errors. Design a pronunciation practice course specifically designed for non-native speakers. In this course, learners can view real-time images of their oral and tongue movements through software. The software compares standard pronunciation patterns and identifies deviations and areas for improvement. By providing real-time feedback, learning can quickly understand and imitate the correct tongue position and oral posture.

5. Conclusion

This paper designs a pronunciation task experiment and combines multi-modal technology to synchronously acquire tongue dynamic images and electromyographic signals to comprehensively analyze the impact of the biomechanical properties of tongue muscles on the accuracy of English vowel pronunciation. A reasonable biomechanical model of the tongue is established for precise simulation, highlighting the importance of the synergistic effect of the internal and external muscles of the tongue for precise control of the tongue position. The innovation of this paper lies in establishing a reasonable biomechanical model of the tongue to accurately simulate the synergistic effect of the internal and external muscles of the tongue, emphasizing the importance of precise control of the tongue position. The research results indicate that native English speakers outperform non-native speakers in terms of pronunciation accuracy and coordination of tongue muscle activity. In order to overcome the shortcomings of existing research, especially the limitations in sample selection, future research plans to expand the sample size to include participants with more language backgrounds, in order to explore cross linguistic differences in tongue muscle activity. In addition, to address the issue of insufficient dynamic data, research will increase the number of experiments and extend observation time to obtain more comprehensive dynamic data. Regarding the difficulties in integrating multimodal technologies, efforts will be made to optimize data synchronization and calibration processes, as well as develop more advanced algorithms to enhance the integration capability of multimodal data and ensure its accuracy and reliability.

Author contributions: Writing—original draft preparation, PZ; writing—review and editing, XC. All authors have read and agreed to the published version of the manuscript.

Funding: Hubei Province Education and Science Planning Project: Research on the Multimodal Narration of “Tell Chinese Stories Well” in College English Teaching. (Project No.: 2024GB507).

Ethical approval: Not applicable.

Conflict of interest: The authors declare no conflict of interest.

References

1 Sugiarto R, Prihantoro P, Edy S. THE IMPACT OF SHADOWING TECHNIQUE ON TERTIARY STUDENTS’ENGLISH PRONUNCIATION. Linguists: Journal Of Linguistics and Language Teaching. 2020; 6(1): 114-125.

2 Pourhosein Gilakjani A, Rahimy R. Using computer-assisted pronunciation teaching (CAPT) in English pronunciation instruction: A study on the impact and the Teacher’s role. Education and information technologies. 2020; 25(2): 1129-1159.

3 Tsunemoto A, McDonough K. Exploring Japanese EFL learners’ attitudes toward English pronunciation and its relationship to perceived accentedness. Language and speech. 2021; 64(1): 24-34.

4 Djurayeva YA. Enhancing English pronunciation in learning process. Academic research in educational sciences, 2021; 2(2): 302-306.

5 Pakpahan M. The Significance of English Phonology Learning Towards the Improvement of Pronunciation Accuracy of English Vowel Sounds: A Review. BLESS. 2023; 3(1): 9-19.

6 Garita Sanchez M R, Gonzalez Lutz M I, Solis Perez N. English vowel sounds: Pronunciation issues and student and faculty perceptions. Actualidades Investigativas en Educación. 2019; 19(3): 33-67.

7 Kobilova NR. Importance of pronunciation in English language communication. Academic research in educational sciences. 2022; 3(6): 592-597.

8 Floare Bora S. Taking literature off page! The effectiveness of a blended drama approach for enhancing L2 oral accuracy, pronunciation and complexity. Language Teaching Research. 2024; 28(5): 1869-1892.https://doi.org/10.1177/13621688211043490

9 Sardegna VG. Evidence in favor of a strategy-based model for English pronunciation instruction. Language Teaching. 2022; 55(3): 363-378.https://doi.org/10.1017/S0261444821000380

10 Fogarty M J, Sieck G C. Tongue muscle contractile, fatigue, and fiber type properties in rats. Journal of Applied Physiology. 2021; 131(3): 1043-1055.

11 Goto A, Kokabu S, Dusadeemeelap C, et al. Tongue muscle for the analysis of head muscle regeneration dynamics. Journal of Dental Research. 2022; 101(8): 962-971.https://doi.org/10.1177/00220345221075966

12 Szelenyi A, Fava E. Long latency responses in tongue muscle elicited by various stimulation sites in anesthetized humans–New insights into tongue-related brainstem reflexes. Brain Stimulation. 2022; 15(3): 566-575.

13 Zhu Mengxian, Jiang Chenghui, Zhou Guangchao, Wang Binbing, Li Sheng, Shi Xinghui. Research progress on the analysis of tongue-movement patterns during articulation by ultrasound imaging. International Journal of Stomatology, 2022, 49(3): 356-361.

14 Jia M, Keasuwan C. A STUDY ON THE PRONUNCIATION TEACHING OF /A/ FOR CHINESE THAI LANGUAGE LEARNERS. Chinese Language and Culture Journal. 2021; 8(1): 197-206.https://so02.tci-thaijo.org/index.php/clcjn/article/view/246756

15 Wang Tingwei, Wu Junfa, Hu Ruiping, Zhang Yunfeng, Jiang Congyu, Fan Shunjuan, et al. Feasibility study of tongue pressure meter for evaluating tongue muscle strength and endurance. Chinese Journal of Physical Medicine and Rehabilitation. 2021; 43(9): 832-834.https://doi.org/10.3760/cma.j.issn.0254-1424.2021.09.015

16 Lu L, Mao J, Wang W, Ding G, Zhang Z. A study of personal recognition method based on EMG signal. IEEE Transactions on Biomedical Circuits and Systems. 2020; 14(4): 681-691.

17 Sharma A, Sharma I, Kumar A. Signal Acquisition and Time–Frequency Perspective of EMG Signal-based Systems and Applications[J]. IETE Technical Review. 2024; 41(4): 466-485

18 Kent RD, Rountrey C. What acoustic studies tell us about vowels in developing and disordered speech. American Journal of Speech-Language Pathology. 2020; 29(3): 1749-1778.

19 Oganian Y, Bhaya-Grossman I, Johnson K, Chang E F. Vowel and formant representation in the human auditory speech cortex. Neuron. 2023; 111(13): 2105-2118.

20 Gohel V, Mehendale N. Review on electromyography signal acquisition and processing. Biophysical reviews, 2020, 12(6): 1361-1367.

21 Pakniyat N, Soundirarajan M, Gohari S, Burvill C, Krejcar O, Namazi H. Decoding of facial muscle-brain relation by information-based analysis of electromyogram (EMG) and electroencephalogram (EEG) signals. Waves in Random and Complex Media, 2024, 34(4): 3599-3608.

22 Zolkov E, Weiss R, Cohen E. Analysis and design of N-path band-pass filters with negative base band resistance. IEEE Transactions on Circuits and Systems I: Regular Papers, 2020, 67(7): 2250-2262.

23 Garcia-Martinez H, Avila-Navarro E, Torregrosa-Penalva G, Delmonte N, Silvestri L, Marconi S, et al. Design and fabrication of a band-pass filter with EBG single-ridge waveguide using additive manufacturing techniques. IEEE Transactions on Microwave Theory and Techniques, 2020, 68(10): 4361-4368.

24 Clunie D A. DICOM format and protocol standardization—a core requirement for digital pathology success. Toxicologic Pathology, 2021, 49(4): 738-749.

25 Wei Y, Luo Q, Mantooth A. Comprehensive comparisons between frequency-domain analysis and time-domain analysis for LLC resonant converter[J]. IET Power Electronics, 2020, 13(9): 1735-1745.

26 Shi B, Zhao Z, Zhu Y, Wang X. Time-Domain and Frequency-Domain Analysis of SiC MOSFET Switching Transients Considering Transmission of Control, Drive, and Power Pulses. IEEE Journal of Emerging and Selected Topics in Power Electronics, 2021, 9(5): 6441-6452.

27 Kong L, Nian H. Fault detection and location method for mesh-type DC microgrid using pearson correlation coefficient. IEEE Transactions on Power Delivery, 2020, 36(3): 1428-1439.

28 Peng S, Han W, Jia G. Pearson correlation and transfer entropy in the Chinese stock market with time delay. Data Science and Management, 2022, 5(3): 117-123.

29 Uerlich R, Ambikakumari Sanalkumar K, Bokelmann T, Vietor T. Finite element analysis considering packaging efficiency of innovative battery pack designs. International Journal of Crashworthiness, 2020, 25(6): 664-679.

30 Mahantesh MM, Rao K V S R, Chandra A C P, Vijayakumar M N, Nandini B, Prasad C D, et al. Design and modeling using finite element analysis for the sitting posture of computer users based on ergonomic perspective[J]. International Journal on Interactive Design and Manufacturing (IJIDeM), 2024, 18(8): 5875-5891.

31 Biltz N K, Collins K H, Shen K C, Schwartz K, Harris C A, Meyer G A. Infiltration of intramuscular adipose tissue impairs skeletal muscle contraction[J]. The Journal of physiology, 2020, 598(13): 2669-2683.

32 Viljoen S, Hanekom T, Farina D. Effect of characteristics of dynamic muscle contraction on crosstalk in surface electromyography recordings[J]. SAIEE Africa Research Journal, 2021, 98(1): 18-29.

33 Jalali Z, Noorzai E, Heidari S. Design and optimization of form and facade of an office building using the genetic algorithm[J]. Science and Technology for the Built Environment, 2020, 26(2): 128-140.

34 Ahmmed T, Akhter I, Karim S M R, Ahamed F S. Genetic algorithm based PID parameter optimization. American Journal of Intelligent Systems, 2020, 10(1): 8-13.

35 Gad A G. Particle swarm optimization algorithm and its applications: a systematic review. Archives of computational methods in engineering, 2022, 29(5): 2531-2561.

36 Zeng N, Wang Z, Liu W, Zhang H, Hone K, Liu X. A dynamic neighborhood-based switching particle swarm optimization algorithm. IEEE Transactions on Cybernetics, 2020, 52(9): 9290-9301.

37 Qin Lijuan, Feng Naiqin. Sparse Data Feature Extraction of Sparse Data Based on Deep Learning Back Propagation[J]. Computer Simulation, 2022, 39(5): 333-336.

38 Hongwei GUO, Ce ZHU, Xu YANG, Lei LUO. Temporal dependent rate-distortion optimization based on distortion backward propagation. Journal on Communications, 2022, 43(12): 222-232.https://doi.org/10.11959/j.issn.1000-436x.2022230