Design of a Speaking Training System for English Speech Education using Speech Recognition Technology

Hengheng He
School of Foreign Languages, Hankou University, Wuhan, Hubei 430212, China

***Abstract*—A good English speaking training system can provide an aid to the learning of English. This paper briefly introduced the English speaking training system and described the speaking training scoring and pronunciation resonance peak display modules in the system. The speaking training scoring module scored pronunciation with the Long Short-Term Memory (LSTM). The pronunciation resonance peak display module extracted the resonance peak with Fourier transform and visualized it. Finally, the speaking scoring module, the pronunciation resonance peak display module, and the effect of the whole system in improving students' speaking pronunciation was tested. The results showed that the LSTM-based speaking scoring algorithm had highest scoring accuracy than pattern matching and the recurrent neural network (RNN) algorithm, and its accuracy was 95.21% when scoring the LibriSpeech dataset and 90.12% when scoring the local English dataset. The pronunciation resonance peak display module displayed the change of mouth shape before and after training, and the pronunciation after training was closer to the standard pronunciation. The P value in the comparison of the speaking level before and after training with the system was 0.001, i.e., the difference was significant, which indicated that the students' English speaking proficiency significantly improved.**

***Keywords*—English speech; long short-term memory; speaking training; speech recognition**

I. Introduction

In order to ensure that information is communicated as accurately as possible during national exchanges, a common language is needed. English is one of the common languages for international communication [1]. Educational institutions are also paying more and more attention to English teaching, especially spoken English. The biggest role of spoken English is to communicate, and it is difficult to communicate fluently just by reading and writing. The pronunciation quality of oral English will directly affect communication efficiency [2]. In the traditional English teaching mode, classroom teaching is the main focus, and students often learn what the teacher teaches them. The lack of time for oral training in the classroom and the teacher's irregular pronunciation will affect the learning effect of students' oral language [3]. Information technology is gradually integrated into the traditional teaching mode and changes the teacher-oriented structure of the traditional teaching mode. When using information technology to assist oral English training, teachers can still teach relevant speaking knowledge in the classroom, and students can train their speaking independently through the independent oral English learning system under information technology before or after class [4]. The oral English training system can provide the standard pronunciation of English, collect students' pronunciation, evaluate the pronunciation through the scoring module in the system, and help students adjust their pronunciation [5]. The oral English scoring module is the core of the oral training system; the higher the scoring accuracy, the higher the reference value for the oral training. This paper studied the speaking English training system that adopted the speech recognition technology in order to enhance the students' speaking English level. The LSTM was used to score the students' pronunciation in the speaking scoring module of the speaking training system. The resonance display module used the Fourier transform to extract and visualize the resonance peaks during pronunciation to assist in pronunciation training. The two modules of the speaking training system were tested. The effectiveness of the speaking training system in improving students' speaking skills was tested. The final results showed that the LSTM had higher scoring accuracy than the pattern matching and RNN algorithm, the pronunciation resonance peak display module effectively visualized the resonance peaks of the spoken pronunciation, and the students who used the speaking training system for learning had significantly improved English speaking skills. The research on the English speaking training system and the experimental results of this paper provide an effective reference for improving students' English speaking skills. The limitations of this paper are that only the scoring module and pronunciation visualization module were tested and there were few test subjects in the test of the performance of the speaking training system. The future research direction is to conduct in-depth research on the scoring module and pronunciation visualization module of the training system and to increase the scale of test subjects.

This paper is organized in the order of abstract, introduction, and literature review, introduction of English speaking training system, simulation experiments, discussion, conclusion, and references.

II. Literature Review

Hsieh et al. [6] used a technology acceptance model to investigate the effect of Line on speaking training in a non-native English teaching environment and found that Line had a more positive effect on English speaking instruction than traditional classroom instruction. Reitz et al. [7] embedded the English learning process into a generic 3D cooperative virtual

reality (VR) game and found through example analysis that the designed VR game effectively trained students' English communication skills. Cao et al. [8] proposed a lip movement judgment algorithm based on ultrasonic detection to aid spoken English pronunciation. They performed experiments and found that the system had a speech accuracy of 85%, i.e., it could improve the English speaking trainers to a certain extent.

III. English Speaking Training Evaluation System

A. The Basic Structure of English Speaking Training System

The basic architecture of the English speaking training system is shown in Fig. 1. The speech acquisition and playback module [9] is equivalent to the ear and mouth of the training system. The acquisition module is responsible for acquiring the audio signal of the spoken language, and the collected audio will be stored in the buffer for later processing. The playback module plays the stored English standard pronunciation to provide a reference for the students [10].

The primary function of the pronunciation resonance peak image display module is to graphically display the standard spoken audio and the student's spoken audio so that students can clearly see the difference between their pronunciation and the standard pronunciation to adjust their pronunciation [11].

The main function of the spoken pronunciation scoring module is to score the pronunciation of the spoken audio signals collected by the audio acquisition module, i.e., to rate the standard level of students' spoken English pronunciation and to make a quantitative analysis of the students' speaking training level. The scores can be used as a feedback incentive for the students' speaking training [12].

B. Speech Recognition-based Spoken Pronunciation Scoring Module

For the whole English speaking training system, the speaking pronunciation scoring module is the core, and its main function is to quantify the students' speaking pronunciation level to give a feedback incentive to the students' speaking training [13]. First, the spoken speech signal is collected using the speech acquisition module; then, features are extracted from the speech signal after preprocessing; finally, the speech is scored according to the features [14]. This paper used LSTM to score the speech collected by the speech acquisition module, and the basic flow is shown in Fig. 2.

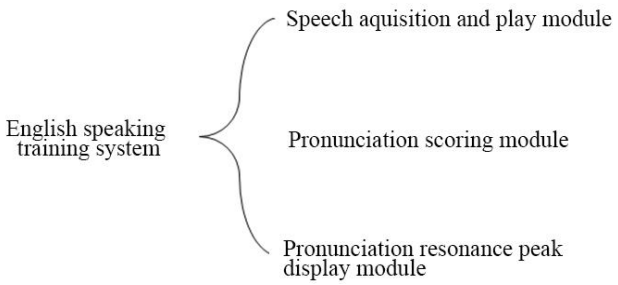

Fig. 1. The Basic Architecture Module of English Speaking Training System.

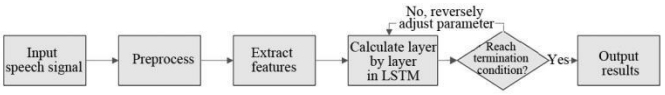

Fig. 2. The Workflow of the Pronunciation Scoring Module.

The flow of spoken pronunciation scoring based on LSTM speech recognition is shown in Fig. 2.

1) The speech signal is collected.

2) Pre-processing such as filtering, windowing, and framing [15] is performed.

3) Features are extracted from the pre-processed pronunciation signal using the Mel-frequency cepstral coefficient [16] to extract audio features.

4) The Mel-frequency cepstral coefficient features of every audio frame are input into the LSTM in order for forward calculation [17]:

$$\begin{cases} i_t = g(\omega_i \cdot [h_{t-1}, x_t] + b_i) \\ \tilde{C}_t = \tanh(\omega_C \cdot [h_{t-1}, x_t] + b_C) \\ C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \\ f_t = g(\omega_f \cdot [h_{t-1}, x_t] + b_f) \\ o_t = g(\omega_o \cdot [h_{t-1}, x_t] + b_o) \\ h_t = o_t \cdot \tanh(C_t) \end{cases} \quad (1)$$

where x_t represents the current input of the cell, h_{t-1} is the last hidden state, C_{t-1} is the last cell state, i_t is the weight that determines the new information to be remembered, $\tilde{C}_t$ is the cell state of the new information added [18], C_t refers to the current cell state after the new information is added, ω_i and ω_t are weights, b_i and b_C are biases [19], f_t is the weight of the information not to be forgotten in C_{t-1}, ω_f is the weight, b_f is the bias, o_t represents the weight that determines the final output information amount, and h_t refers to the final output or the next hidden state [20].

5) Whether the training of the algorithm reaches the termination condition is determined. If the termination condition is reached, the training is finished, and the parameters of the LSTM are fixed [21]. When applied to the actual pronunciation scoring, the extracted features of pronunciation are input into the LSTM in order to get the scoring results. If the termination condition is not reached, the parameters in the LSTM are adjusted reversely using the stochastic gradient descent method [22]. The termination conditions include: (1) the number of algorithm iterations reaches a preset number; (2) the error between the forward calculation result and the expected result converges to the preset threshold.

Fig. 3. Workflow of the Pronunciation Resonance Peak Image Display Module.

C. Pronunciation Resonance Peak Image Display Module

The pronunciation scoring module in the speaking training system quantifies students' pronunciation level, but this quantification only converts the level of pronunciation into a number, which does not reflect the students' pronunciation process visually [23]. The scoring module can only give a final target, and it is difficult to guide students to correct their pronunciation directly. Therefore, there is a need for a module that can visually assist in pronunciation correction. The pronunciation resonance peak image display module is a module that can visually display the pronunciation process of students.

The human vocal tract and the oral cavity together form a resonance cavity. After the sound wave signal of the vocal cord vibration is filtered by the resonance cavity, the energy will be redistributed in different frequencies. When the mouth shape changes, the resonance cavity will also change; thus, playing a different filtering effect to change the pronunciation. The connection between the mouth shape and the resonance peaks makes it possible to guide the mouth shape and correct the pronunciation based on the changes in the resonance peaks. The workflow of the resonance peak image display module in the speaking training system is shown in Fig. 3.

1) The speech signal is input and pre-processed by filtering, windowing, and framing [24].

2) The spectrum of a single-frame speech signal is obtained by using Fast Fourier Transform (FFT).

3) The maximum resonance peak of a single-frame speech signal is calculated.

4) An image is drawn to reflect the change in students' pronunciation. Time is the horizontal coordinate axis, and the smallest unit of the horizontal coordinate axis is a frame. Frequency is the vertical coordinate axis. The line graphs are plotted in the coordinate chart in the chronological order of the speech signals and the frequency of the maximum resonance peak of every speech signal frame. The difference between students' pronunciation and standard pronunciation can be seen visually when the line graphs of students' pronunciation and standard pronunciation are placed in the same coordinate plane [25].

The combination of pronunciation scores and resonance peak comparison charts can guide students to correct their pronunciation.

IV. Simulation Experiments

A. Experimental Data

The LibriSpeech dataset (openslr.org/12/) was used for simulation experiments, which includes 1000 h of audio data. It is a dataset of audiobooks, including texts and speeches. The sampling rate of the audio data in this dataset was 16 kHz. After subdivision and collation, every audio in the dataset was about 10 s long.

In addition to the above English speech dataset collected from the public dataset, this paper also collected English spoken pronunciation from students to construct a local English speech dataset in order to verify the actual effect of the speaking training system on students' speaking scores. Fifty students, including 25 males and 25 females, participated in the English pronunciation collection. Every student read aloud 15 randomly selected non-repeated sentences from the common oral English sentence database. The read-aloud speech was captured in a recording room using recording software with a sampling rate of 16 kHz. The sentences included:

1) How do you feel?
2) What's the weather like?
3) See you tomorrow.

B. Experimental Projects

1) Scoring module test of the speaking training system: In order to test the scoring performance of the proposed LSTM speech recognition-based speaking training system, the scoring performance of the pattern matching-based and Recurrent Neural Network (RNN) speech recognition-based speaking training system was tested in addition to the accuracy detection experiment.

The scoring algorithm of the pattern matching-based speaking training system used a dynamic time regularization algorithm to calculate the minimum matching distance of extracted features between oral pronunciation and standard pronunciation. The score was calculated based on the minimum matching distance.

The relevant parameters of the speaking training scoring algorithm based on RNN speech recognition are as follows. Thirteen input nodes, 100 hidden nodes, a sigmoid function, and one output node were used. The stochastic gradient descent was used for backward learning. The learning rate was 0.1.

The relevant parameters of the speaking training scoring algorithm based on LSTM speech recognition are as follows. The number of nodes in the LSTM input layer was set as 13. The number of node cells in the hidden layer was set as 100, and there was an input gate, forgetting gate, and output gate in every node cell. The activation function for the node cells was the sigmoid function. The number of nodes in the output layer was set as 1, and the softmax function was used. The stochastic gradient descent method is used to reverse the parameters in the node cells of the hidden layer, the learning rate was set as 0.1, and the maximum number of iterations was set as 1000.

In the process of testing the scoring accuracy of the above three scoring algorithms, the scores calculated by the algorithms were compared with the standard scores. The mean value of manual scoring by 20 experts was used as the standard score. The scoring accuracy of the algorithms is calculated as follows:

$$\begin{cases} R_i = 1 - \dfrac{|S_{i1} - S_{i2}|}{S_{i2}} \\ R = \dfrac{\sum_{i=1}^{n} R_i}{n} \end{cases} \quad (2)$$

where R_i is the score similarity of the algorithm for the i-th sample, R is the score accuracy of the algorithm, S_{i1} is the score given by the algorithm for the i-th sample, S_{i2} is the expert manual score for the i-th sample, and n is the number of test samples.

2) Testing of the pronunciation resonance peak image display module of the speaking training system: The standard pronunciation of the local English speech dataset was input in the pronunciation resonance peak image display module of the speaking system. Then, the pronunciation of the students' pronunciation of the local English speech dataset before receiving training by the speaking training system and the pronunciation after receiving training by the system were input into the module to output the pronunciation resonance peak comparison graph.

3) Testing of the improvement of students' speaking levels after training by the speaking training system: One hundred students were randomly selected from the School of Foreign Languages of Hankou University and were divided into two groups: a control group and an experimental group. The control group used the traditional teaching method for English speaking training, while the experimental group used the speaking training system in addition to the traditional teaching method for speaking training. The students in both groups were taught to speak for two weeks, and their speaking levels were scored by 20 experts before and after the teaching, using a hundred-mark system.

C. Test Results

To verify the scoring accuracy of the LSTM speech recognition-based speaking training scoring algorithm, the LibriSpeech dataset and the local English speech dataset were used for scoring accuracy testing, and it was also compared with the two speaking training scoring algorithms based on pattern matching and RNN speech recognition. The test results are shown in Fig. 4. For the LibriSpeech dataset, the accuracy of the pattern matching-based scoring algorithm was 79.69%, the accuracy of the RNN speech recognition-based scoring algorithm was 89.65%, and the accuracy of the LSTM speech recognition-based scoring algorithm was 95.21%. The corresponding accuracy for the local English speech dataset was 72.34%, 81.33 %, and 90.12%, respectively. It was seen from the comparison in Fig. 4 that the speaking training scoring algorithm based on LSTM speech recognition had the highest accuracy, followed by the speaking scoring algorithm based on RNN speech recognition, and the speaking training scoring algorithm based on pattern matching had the lowest accuracy when scoring the same speech dataset. In addition, the accuracy of the three speaking training scoring algorithms was higher when scoring the LibriSpeech dataset.

Due to the space limitation, only a resonance peak graph for comparing a student's pronunciation of an English sentence with the standard pronunciation is shown here, as shown in Fig. 5. Fig. 5 shows the changes in the resonance peak of the students' pronunciation over time. Because of the connection between the resonance peak and the mouth shape, the graph also qualitatively reflected the changes in the student's mouth shape over time. The resonance peak line graph comparison demonstrated that the resonance peak of the student's pronunciation was larger than that of the standard pronunciation, indicating that the student's tongue position was low and the mouth was opened too wide during pronunciation. Therefore, students should raise the tongue position and reduce the mouth shape. After the training, the resonance peak broken line of the student's pronunciation almost coincided with that of the standard pronunciation, indicating that the mouth shape training was effective.

In order to verify the effect of the LSTM speech recognition-based speaking training system in improving pronunciation, 100 students were randomly selected from the Foreign Language Institute of Hankou University and then divided into two groups. The control group was taught traditional speaking, and the experimental group was trained with the speaking training system in addition to traditional speaking teaching. The speaking level of students in both groups was scored by 20 experts before and after receiving the teaching, and the results are shown in Table I.

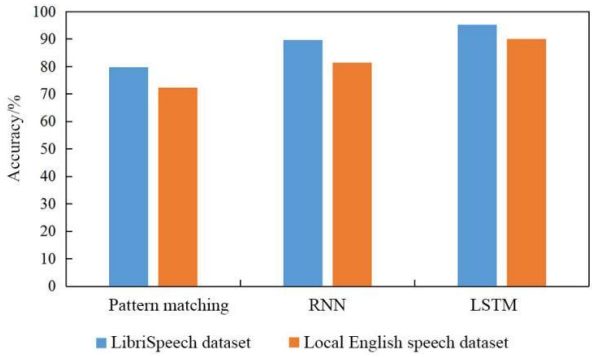

Fig. 4. Scoring Accuracy of Three Speaking Training Scoring Algorithms for Two Speech Datasets.

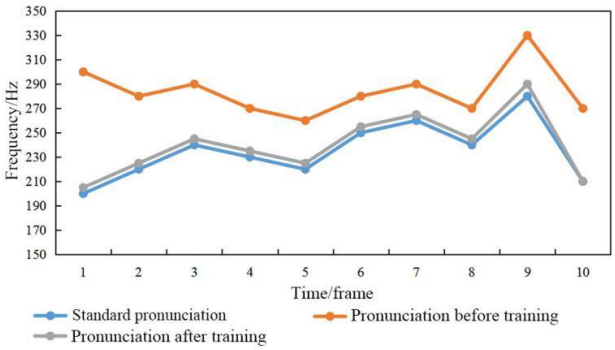

Fig. 5. Comparison of the Resonance Peak between the Student's Pronunciation before and after Training and the Standard Pronunciation.

TABLE I. MEAN SCORES OF PRONUNCIATION IN THE CONTROL AND EXPERIMENTAL GROUPS BEFORE AND AFTER ORAL INSTRUCTION

	Control group	Experimental group	P value
Pre-teaching	75.1 ± 5.2	75.3 ± 6.1	0.126
After teaching	78.6 ± 5.7	92.1 ± 1.5	0.001
P value	0.144	0.001	

It was seen from Table I that there was no significant difference between the speaking level of the control group and the experimental group before conducting the speaking instruction. After the teaching, there was a significant difference between the speaking level of the control and experimental groups, and the speaking level of the experimental group was significantly higher than that of the control group.

It was found from the comparison of the average speaking score of the same group before and after the teaching that the P value of the difference between the average speaking score of the control group before and after the teaching was 0.144, which was greater than 0.05, i.e., the speaking level of the control group was not significantly improved after the teaching; the P value of the difference between the average speaking score of the experimental group before and after the teaching was 0.001, which was less than 0.05, i.e., the speaking level of the experimental group was significantly improved after the teaching.

V. DISCUSSION

Students need to pronounce English successfully in addition to being able to read and write English correctly in the process of learning English. As English is a language for communication, the level of spoken English is somehow more important than the level of English reading and writing. However, in the process of learning spoken English, a standard pronunciation is needed as a reference. In the traditional teaching of spoken English, the teacher usually pronounces the words, the students follow the teacher's pronunciation, and the teacher corrects the students' pronunciation. In this teaching mode, the students' speaking training effect depends on the teacher's speaking level, and it is impossible for the teacher to teach one-on-one.

With the development of speech recognition technology, it has been gradually applied to various fields, including English speaking training. The English speaking training system built with speech recognition technology can evaluate students' pronunciation by the scoring module and visualize their pronunciation characteristics by the pronunciation resonance peak display module. With these two modules, students can evaluate their own speaking level according to the standard pronunciation given by the system and adjust their pronunciation according to the differences in pronunciation characteristics. The scoring module and pronunciation resonance peak display module were tested in the simulation experiment, and the effect of the speaking training system in improving the students' English speaking level has been studied above.

The results of the accuracy test of the scoring module showed that the accuracy when using the LSTM for pronunciation evaluation was higher than the pattern matching and the RNN algorithm. The reason is as follows. Although the dynamic time regularization method was used in the pattern matching-based speaking training scoring algorithm to reduce the difficulty of matching due to duration randomness, some information was lost in the process of stretching or compressing the audio, and the standard speech templates in the template library were relied on, so it had better scoring accuracy for the LibriSpeech dataset containing a larger amount of data. The RNN speech recognition-based speaking training scoring algorithm used the RNN to score speech. Compared with pattern matching, the RNN did not require standard templates for scoring but directly scored the pronunciation according to the pattern obtained during training, which was more efficient. Moreover, the activation function in the RNN effectively fit the nonlinear pattern between pronunciation features and scoring, so it was more accurate than pattern matching. Compared with the RNN speech recognition-based algorithm, the LSTM speech recognition-based speaking training scoring algorithm used an activation function that can effectively fit the nonlinear pattern, but the introduced forgetting gate unit avoided the gradient explosion when facing long data, so it had better accuracy when scoring long speech.

The test results of the pronunciation resonance peak display module showed that the module could visualize the resonance peaks of students' spoken pronunciation. The resonance peaks of pronunciation before and after training were compared with those of standard pronunciation, and the results showed that the resonance peaks of pronunciation after training were closer to those of standard pronunciation. Taking the results presented in Fig. 5 as an example, the resonance peaks of the students' pronunciation before training were larger than those of the standard pronunciation, indicating that the tongue position was low and the mouth opened too wide during the pronunciation process. Thus, the tongue position needed to be raised, and the mouth shape should be smaller during the training process. The resonance peaks after training also showed the effectiveness of adjusting the pronunciation.

The results of testing the effectiveness of the speaking training system showed that the experimental group that applied the speaking training system had a significant improvement in their pronunciation after teaching compared to the control group that adopted the traditional teaching mode. The reason is as follows. In the traditional teaching mode, students adjusted their own pronunciation according to the teacher's pronunciation; the pronunciation of the whole class would be affected if the teacher's pronunciation was wrong. Moreover, different students had different pronunciation habits, so it was difficult for the teacher to provide targeted tutoring to students on a one-to-one basis, and the common tutoring would make some students have difficulty in keeping up with the learning pace. When using the speaking training system, the students adjusted their pronunciation independently with the standard pronunciation as the target based on the scoring module and pronunciation visualization module, which was considered targeted training.

VI. Conclusion

This paper briefly introduced the English speaking training system and described the scoring algorithm in the speaking training scoring module of the system and the pronunciation resonance peak display module in the system. The pronunciation scoring module and the pronunciation resonance peak display module in the speaking training system were tested. In addition, the speaking level of the students that were trained traditionally and trained by the speaking training system was compared before and after the teaching. The results are as follows. (1) When scoring the same speech dataset, the LSTM-based speaking scoring algorithm was the most accurate, the RNN-based scoring algorithm was the second most accurate, and the pattern matching-based scoring algorithm was the lowest. (2) The three algorithms achieved higher accuracy when scoring the LibriSpeech dataset. (3) The pronunciation resonance peak display module effectively displayed the line graph of the resonance peak of the student's pronunciation over time and visually reflected the difference of the resonance peak between students' pronunciation and standard pronunciation. (4) The difference in the speaking level between the control group and the experimental group before receiving instruction was not significant; the speaking level of the control group improved insignificantly after receiving traditional speaking instruction, and the speaking level of the experimental group improved significantly and was significantly higher than that of the control group after being trained by the speaking training system.

References

[1] J. Lubek, "The Cost of Training: Oral and Maxillofacial Surgery at a Crossroad," Or. Surg. Or. Med. Or. Pa., vol. 127, pp. 465-467, March 2019.

[2] W. L. Martens, and R. Wang, "Applying adaptive recognition of the learner's vowel space to English pronunciation training of native speakers of Japanese," SHS Web Conf., vol. 102, pp. 1-8, January 2021.

[3] J. Cai, and Y. Liu, "Research on English pronunciation training based on intelligent speech recognition," Int. J. Speech Technol., vol. 21, pp. 633-640, September 2018.

[4] M. Suganuma, T. Yamamura, Y. Hoshino, and M. Yamada, "Proposal of the Way of English Pronunciation Training Evaluation by Lip Movement," J. Jpn. Pers. Comput. Appl. Technol. Soc., vol. 11, pp. 8-20, 2017.

[5] T. Yoshioka, S. Karita, and T. Nakatani, "Far-field speech recognition using CNN-DNN-HMM with convolution in time," IEEE Int. Conf. on Acoustics, Speech and Signal Processing, pp. 4360-4364, April 2015.

[6] J. Hsieh, Y. M. Huang, and W. Wu, "Technological acceptance of LINE in flipped EFL oral training," Comput. Hum. Behav., vol. 70, pp. 178-190, December 2016.

[7] L. Reitz, A. Sohny, and G. Lochmann, "VR-Based Gamification of Communication Training and Oral Examination in a Second Language," Int. J. Game-Based Learn., vol. 6, pp. 46-61, April 2016.

[8] Q. Cao, and H. Hao, "Optimization of Intelligent English Pronunciation Training System Based on Android Platform," Complexity, vol. 2021, pp. 1-11, March 2021.

[9] Y. Wang, F. Bao, H. Zhang, and G. Gao, "Research on Mongolian Speech Recognition Based on FSMN," Natl. CCF Conf. on Natural Language Processing and Chinese Computing, pp. 243-254, January 2017.

[10] C. Agarwal, and P. Chakraborty, "A review of tools and techniques for computer aided pronunciation training (CAPT) in English," Educ. Inf. Technol., vol. 24, pp. 3731-3743, November 2019.

[11] F. G. Jonas, "Podcast-based pronunciation training: Enhancing FL learners' perception and production of fossilised segmental features," ReCALL, vol. 31, pp. 150-169, April 2019.

[12] Y. Sun, X. Jiang, "The design and application of English pronunciation training software based on Android intelligent mobile phone platform," Rev. Fac. Ing., vol. 32, pp. 756-765, January 2017.

[13] M. J. Alam, V. Gupta, P. Kenny, and P. Dumouchel, "Speech recognition in reverberant and noisy environments employing multiple feature extractors and i-vector speaker adaptation," Eurasip J. Adv. Sig. Pr., vol. 2015, pp. 1-13, December 2015.

[14] F. G. Jonas, "Using apps for pronunciation training: An empirical evaluation of the English File Pronunciation app," Lang. Learn. Technol., vol. 24, pp. 62-85, February 2020.

[15] J. Szpyra-Kozowska, and S. Stasiak, "Verifying a holistic multimodal approach to pronunciation training of intermediate Polish learners of English," Lublin Stud. Mod. Lang. Lit., vol. 40, pp. 181-198, July 2016.

[16] L. Hsu, "A Longitudinal View of Look at the Effectiveness of Elicited Imitation with Computer Assisted Pronunciation Training (CAPT)," Int. Res. Educ., vol. 4, March 2016.

[17] N. Kleynhans, W. Hartman, D. V. Niekerk, C. J. van Heerden, R. Schwartz, S. Tsakalidis et al., "Code-switched English pronunciation modeling for Swahili spoken term detection," Proc. Comput. Sci., vol. 81, pp. 128-135, December 2016.

[18] L. Tian, D. F. Wong, L. S. Chao, P. Quaresma, F. Oliveira, S. Li et al., "UM-Corpus: A Large English-Chinese Parallel Corpus for Statistical Machine Translation," Proc. of the 9th International Conference on Language Resources and Evaluation (LREC'14), January 2014.

[19] N. Hammami, M. Bedda, and F. Nadir, "The second-order derivatives of MFCC for improving spoken Arabic digits recognition using Tree distributions approximation model and HMMs," Int. Conf. on Communications and Information Technology, pp. 1-5, June 2012.

[20] M. E. Celebi, H. A. Kingravi, and P. A. Vela, "A comparative study of efficient initialization methods for the k-means clustering algorithm," Expert Syst. Appl., vol. 40, pp. 200-210, January 2013.

[21] N. Leema, H. K. Nehemiah, and A. Kannan, "Neural Network Classifier Optimization using Differential Evolution with Global Information and Back Propagation Algorithm for Clinical Datasets," Appl. Soft Comput., vol. 49, pp. 834-844, August 2016.

[22] J. Suntornsawet, "Problematic Phonological Features of Foreign Accented English Pronunciation as Threats to International Intelligibility: Thai EIL Pronunciation Core," J. Eng. Int. Lang., vol. 2019, December 2019.

[23] J. Cao, H, Cui, H. Shi, and L. Jiao, "Big Data: A Parallel Particle Swarm Optimization-Back-Propagation Neural Network Algorithm Based on MapReduce," Plos One, vol. 11, pp. 1-17, June 2016.

[24] Z. Xu, J. Liu, X. Chen, Y. Wang, and Z. Zhao, "Continuous blood pressure estimation based on multiple parameters from eletrocardiogram and photoplethysmogram by Back-propagation neural network," Comput. Ind., vol. 89, pp. 50-59, August 2017.

[25] H. Zhou, Z. Deng, Y. Xia, and M. Fu, "A new sampling method in particle filter based on Pearson correlation coefficient," Neurocomputing, vol. 216, pp. 208-215, July 2016.