Eng Int Syst (2020) 3: 137–145

Engineering
Intelligent Systems

Evaluation of the Quality of Audio-Visual Aided Teaching of English

Mengyang Qin[*†]

Department of Tourism English, Zhengzhou Tourism College, Zhengzhou, Henan 450000, China

As one of the most widely used mainstream languages in the world, English is a key learning course for primary and secondary school students as well as college students. Using traditional classroom teaching, teachers often judge whether students' pronunciation is standard or not according to their own experience. The effect of their teaching largely depends on the level of the teachers' oral English, which is not an objective measure. Using voice visualization software, the pronunciation of spoken English can be displayed in a more intuitive form. The differences between standard pronunciation and the students pronunciation can be distinguished more quickly and accurately using this software, in order to correct the pronunciation. This paper briefly introduces Praat, audio-visual software used for teaching spoken English. 100 students in a college were then divided into two groups. One group was taught using traditional teaching methods, and the other group was taught using Praat. A questionnaire survey was then carried out. The results of the survey showed that the students who were taught by Praat had a marked improvement in stress and fluency. The results also showed that most people held a positive attitude towards learning English assisted by the Praat software. Therefore, English pronunciation and intonation teaching based on audio-visual software can improve the effectiveness of teaching and students' interest in learning.

Keywords: Praat, English intonation, teaching effectiveness, speech visualization

1. INTRODUCTION

With the development of globalization, exchanges between different countries are becoming more frequent, both economically and culturally. English is very important in these exchanges as it is one of the most common languages in the world (Bian et al., 2018). English is used for communication in many different fields; hence the listening and speaking abilities of English learners is extremely important. Phonetic learning is essential to improving listening and speech (Yamane, 2015). The quality of speech learning will directly affect the learners' confidence and literacy skills. The traditional teaching method is the teacher introducing language orally and the students listening and imitating the teacher, however this method relies completely on the teacher's level of English. Moreover, the influence of the teacher's native language may lead to mispronunciation, which cannot be identified by the students ears (Chun et al., 2015). In order to improve the teaching effectiveness of English pronunciation and intonation, it is necessary to introduce a tool that can objectively grade students' English pronunciation. Praat is software that can visualize audio recordings. With Praat software, students' English pronunciation can be visualized and compared with standard pronunciation in order to correct their pronunciation and improve the teaching effectiveness of English pronunciation and intonation.

To improve the quality of teaching English, Yang (2018) put forward a method based on Mel frequency cepstrum coefficient detection for feature extraction and identification of listening speech signals, processed the collected English speech signals based on the matching of the action attributes of speech organs, and extracted the information using an MLP detector. Using this method the ability of speech analysis and

[*]Email: yq62me@163.com
[†]Corresponding address: No. 81, North of Jingsan Road, Jinshui District, Zhengzhou, Henan 450000, China

recognition in speech visualization software was improved in the teaching of English; the expression ability of matching features was stronger, and there was a higher accuracy of speech recognition.

Samdudin (2017) put forward a visualization media technique to check and detect the pronunciation of language learners. This technique could convert the phonetic features of learners, such as pitch and rhythm, into animated texts and locate the erroneous part of the pronunciation by comparing it with correct samples. He found that participants gave an active response to the animated text visualization tools and were able to understand the phonetic features of visual text better from taking daily conversation as speech samples. Shinnakasu (2016) analyzed the pronunciation mode of patients with mandibular protrusion using a speech visualization system and evaluated the effectiveness of the system. He concluded that a speech visualization system could be used as a useful tool in the analysis of pronunciation modes of patients with mandibular protrusion. In this study, Praat software was introduced, 100 students from a university were then divided into group A and B and given an English pronunciation level test before and after teaching. Students in group A were taught using the traditional method, while students in group B were given the Praat visualization software assisted teaching method. Students in group B were surveyed after the test. The final results demonstrate that students who received Praat assisted teaching had learnt more effectively and were more interested in learning.

2. INTRODUCTION OF AUDIO-VISUAL SOFTWARE

Praat (Domínguez et al., 2016) was developed by Paul Boersma and David Weenink who work in the School of Humanities Institute of Phonetic Science, at the University of Amsterdam. The software can label and analyze speech signals and output corresponding sonagrams and other data such as pitch, duration and formant frequency (Maryn and Weenink, 2015). Due to the objectivity and intuition in speech signal analysis, Praat has many advantages in pronunciation teaching and error correction. Praat can be downloaded for free, which is an incredible advantage to using the software (Kumbhakarn and Sathe-Pathak, 2015). The operation of the software is simple. Students can utilize Praat for independent pronunciation learning after a short training session or an internet search. It can assist the teaching of teachers, objectively find out the problems of students, and give suggestions in order to improve teaching quality and efficiency.

Speech can be divided into four major elements, pitch, length, intensity and tone, which are also called tone quality phoneme elements (Razavi et al., 2016). Tone is shown as a segmental phoneme in its specific application in the teaching of English pronunciation. Segmental phonemes are divided into vowel phonemes and consonant phonemes (Tsujioka et al., 2016). While learning daily English speech these elements are prone to be mistaken, correcting these errors in the traditional method of oral instruction can be quite abstract, while pronunciation errors can be seen visually in the sonagram output created by Praat.

Although Praat plays a positive role in the teaching of English, it has several shortcomings. Firstly, professional knowledge is needed to understand the sonagram (Saito and Shintani, 2016). Secondly, it may not effectively analyze speech if the audio recording is of low quality. Thirdly, teachers should be able to help students analyze errors pointed out on the sonagram (Hamlaoui and Bengrait, 2016).

3. INSTANCE ANALYSIS

3.1 Research Subjects

One hundred student volunteers who major in English at university were taken as the research subjects. There were 50 males and 50 females. They were around 20 years of age, had clear phonation and had no factors which might affect phonation and recording quality in the experimental period. The English speaker used for contrast teaching was a 20-year-old British woman who had no factors which might affect phonation.

3.2 Related Materials of Teaching English

The English teaching materials used (Lovato et al., 2016) included 30 phrases, 25 sentences, 3 dialogues and 1 essay. The English speaker was asked to read the above English teaching materials, and the reading was recorded. Equipment used in the teaching included a Notebook computer and a Philips SHM1500K microphone. Recording software (22 kHz, 32-bit and double track) that comes with the computer was used. Praat was used for analysis of audio and duration. Data statistics were performed using Microsoft Excel.

3.3 Experimental Methods

3.3.1 Comparison of Front Vowels

As the pronunciation of males and females are different in the aspect of phonics and the English speaker used for contrast teaching was a female, only the female's pronunciation of front vowels was compared. Four words which included /i:/, /i/, /ei/ and /e/ respectively, i.e., peace, hid, fake and check, were selected from the above English teaching materials. The tested females and the English speaker were asked to read the four words and then read sentences which included the words, i.e., "I say ...". The pronunciation was input into the computer. The vowel pronunciation was then divided into several parts using Praat, and the average was taken. The frequency data of the first and second formants of the four vowels was extracted and statistically analyzed using Excel and SPSS.

Table 1 The content of English intonation test before teaching.

Sentence pattern	Content	Pronunciation characteristics of the English speaker
Declarative sentence	She's going to spend her holiday in China in October.	"Spend" and "China" were stressed. The first syllable of "going" and "holiday" were stressed. The second syllable of "October" was stressed. She completed reading the sentence in 2.8s. The tone fell during reading.
Imperative sentence	Sit straight down.	"Sit" and "straight" were stressed. The duration of reading was 0.9s. The tone became higher gradually.
Exclamatory sentence	What a wonderful day!	The first syllable of "wonderful" and day were stressed. The duration of reading was 1.2s. The tone fell gradually.

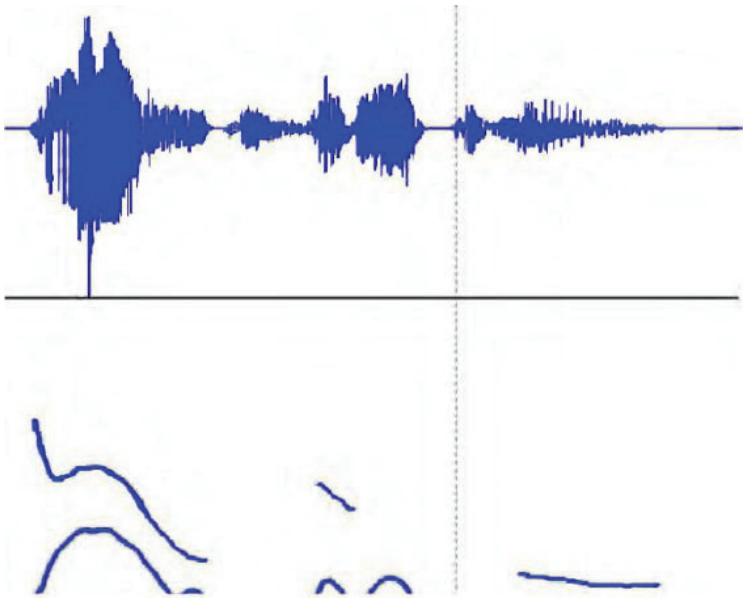

Figure 1 The spectrogram of blackboard

3.3.2 Comparison Between the Traditional Spoken English Teaching Mode and the Praat Software Assisted Teaching Mode

The volunteers were evenly divided into group A (traditional teaching group) and B (experimental teaching group). Students in the two groups received an English intonation test before teaching. The test content included reading three English sentences, as shown in Table 1.

Students in the two groups read the content shown in Table 1, and the reading was recorded. The recordings were analyzed and compared with the teaching audio using Praat. The data was recorded and input into Excel.

After testing, students in the two groups received one week of English lessons. The teaching content of the two groups was the same, with the only difference being the teaching method. Group A was a traditional teaching group; in group A, the students were taught in the traditional method, i.e., students listen to the teacher's explanation for rules of pronunciation in class and enter into a dialogue under the guidance of teachers. Group B was an experimental teaching group; in group B, the pronunciation of the students was recorded in addition to the traditional teaching method and then analyzed and compared with the teaching audio using Praat. The students' speech was visualized to help students correct their errors. The word "blackboard" was used as an example. The spectrogram of the correct pronunciation of "blackboard" is shown in Fig. 1. The first syllable of "blackboard" was stressed; therefore the intensity of the first syllable increased firstly and then decreased with the disappearance of stress. The pitch showed a downtrend. The spectrogram of the wrong pronunciation of "blackboard" where the intensity and stress was in the middle of the word is shown in Fig. 2. Though the characters of the word were the same, the meaning of the word changed. Similar mistakes are common when learning pronunciation and these mistakes can be difficult to judge objectively. Mistakes could clearly be seen after visualization.

After one week of lessons, students in the two groups again performed the English pronunciation test. The test content was the same as Table 2. The recordings were analyzed and compared with the teaching audios using Praat. The data was recorded and then input into Excel.

Finally the students in group B were surveyed. The questionnaire is shown in Table 2.

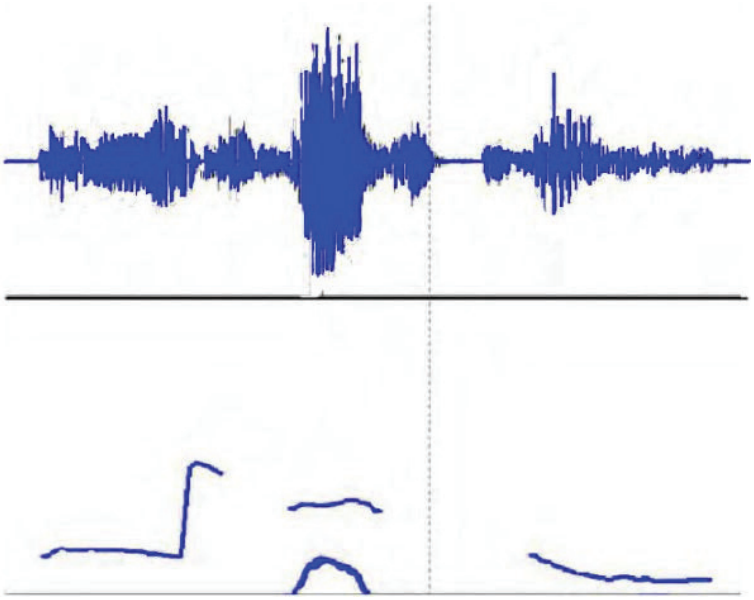

Figure 2 The spectrogram of black board

Table 2 The questionnaires of students in group B.

No.	Problem	Answer (Yes/No)
1	Do you think using Praat is a novel learning way in the learning of English pronunciation?	
2	Do you feel interesting or free in the learning process?	
3	Do you think your English proficiency improved?	
4	Are you able to communicate with English speakers fluently?	

4. RESULTS

4.1 Pronunciation Comparison of English Front Vowels

Generally, the first and second formants (F1 and F2) of the acoustic characteristics of vowels were in an obvious correlation with the place of articulation. The frequency of the first formant was inversely proportional to the tongue position of the vowel pronunciation, while the frequency of the second formant was directly proportional to the tongue position of the vowel pronunciation. Table 4 shows the average values of the first and second formants of the pronunciation of the female volunteers and the English speaker. Fig. 3 is the acoustic vowel chart drawn according to Table 4.

Table 3 and Fig. 1 demonstrate that the first and second formants of the pronunciation of /i:/ by the female volunteers and the English speaker only differed a small amount, and they were almost coincident in the vowel chart. Moreover, the pronunciation of /i/ by the female volunteers was incorrect. The F1 of the female volunteers was slightly lower than that of the English speaker, and the F2 of the former was significantly higher than that of the English speaker; therefore the tongue of the female volunteers was slightly higher than and ahead of that of the English speaker when making the /i/ sound. The F1 of the female volunteers was higher than that of the English speaker, and the F2 of the female volunteers was lower than that of the English speaker when making the /e/ sound; therefore the pronunciation of /e/ by the female volunteers and the English speaker differed greatly. The tongue of the volunteers was lower and further back than that of the English speaker. /ei/ is a diphthong, as opposed to the earlier mentioned monophthongs; therefore the initial and end formants should be recorded. Table 4 shows that the F2 of the female volunteers and the English speaker did not differ by much in the initial stage, and there was a greater difference in the F1; in the ending stage there was only a small difference between both F1 and F2. Therefore the tongues of the volunteers and the English speaker were at different positions in the initial stage, and the positions of tongues were similar in the ending stage.

4.2 Teaching Efficacy of Stress in the Pronunciation of English Sentences

Figs. 4, 5 and 6 demonstrate the comparison of the pronunciation level of stress in the declarative sentence, imperative sentence and exclamatory sentence between groups A and B, before and after teaching. As shown in the figures, there was very little difference in the pronunciation level of stress of group A and B before teaching, and moreover both groups performed poorly in the pronunciation of stress in

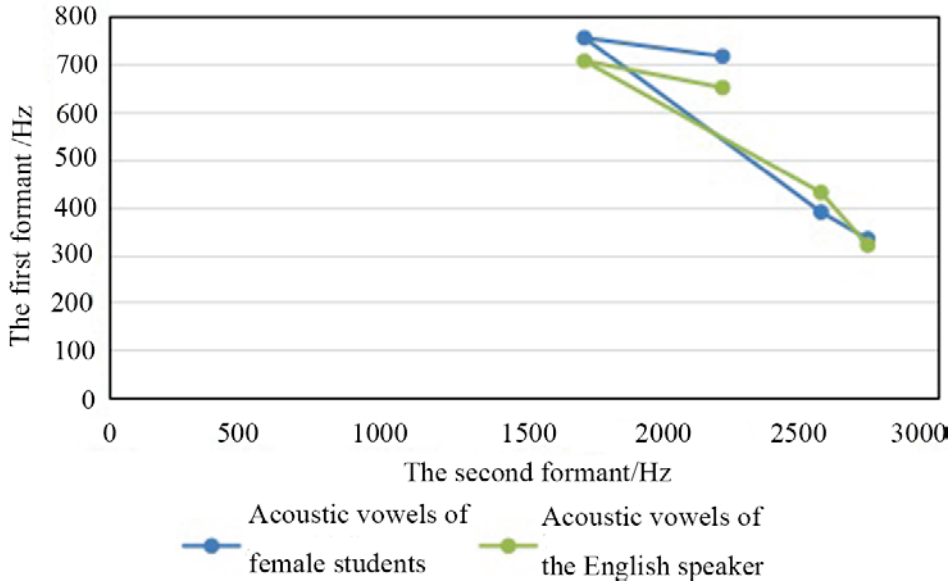

Figure 3 The acoustic vowel chart

Table 3 The average values of the first and second formants of front vowel pronunciation of the female students and English speaker.

Vowels	/i:/	/i/	/ei/(initial/end)	/e/
The first formant of pronunciation of the female students/Hz	333	389	746/491	716
The second formant of pronunciation of the female students/Hz	2748	2579	1721/2002	2222
The first formant of pronunciation of the English speaker/Hz	320	430	706/482	650
The second formant of pronunciation of the English speaker/Hz	2723	2301	1709/1911	2288

Table 4 The statistics of number of students who had different pronunciation duration in different sentence patterns in group A and B after teaching.

Sentence pattern	Group	Pronunciation duration longer than that before teaching (within 1s)/n	Pronunciation duration shorter than that before teaching (within 1s)/n
Declarative sentence	A	30	20
	B	10	40
Imperative sentence	A	25	25
	B	0	50
Exclamatory sentence	A	35	15
	B	5	45

the declarative sentence and performed at an average level in the pronunciation of stress in an imperative sentence and an exclamatory sentence. This may be due to the increased length of the declarative sentence when compared to the imperative sentence and exclamatory sentence; additionally the emotional expression of the imperative sentence and exclamatory sentence was direct (Niebuhr, 2017). Most of the students stressed the word "she" in the declarative sentence, and some students missed one or two words which needed to be stressed in the imperative sentence and the exclamatory sentence.

For example, most of the students made a mistake in the stress of the word "wonderful" in the exclamatory sentence. They stressed the second syllable, where the first syllable should be stressed. Such mistakes are difficult to identify in a traditional lesson. The pronunciation of the word "wonderful" after the application of Praat is shown in Figs. 7 and 8. The difference between the two pronunciations is clearly shown. The first syllable should be stressed, and there was only one stress; therefore the pitch of the word showed a downtrend, as shown in Fig. 7. Fig. 8 demonstrates the pitch of the wrong pronunciation of the word "wonderful" was not on the first syllable. With the assistance of Praat teachers can efficiently discover pronunciation errors of their students and help them to correct them. After one week of teaching there was a vast difference in the stress levels of the English pronunciation between group A and B. The performance of group A was inferior to group B. In the

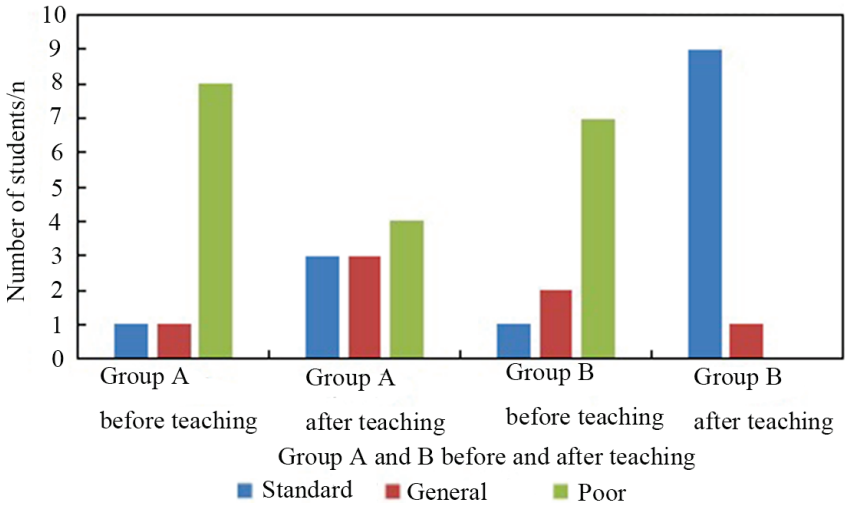

Figure 4 The pronunciation level of stress in the declarative sentence before and after teaching

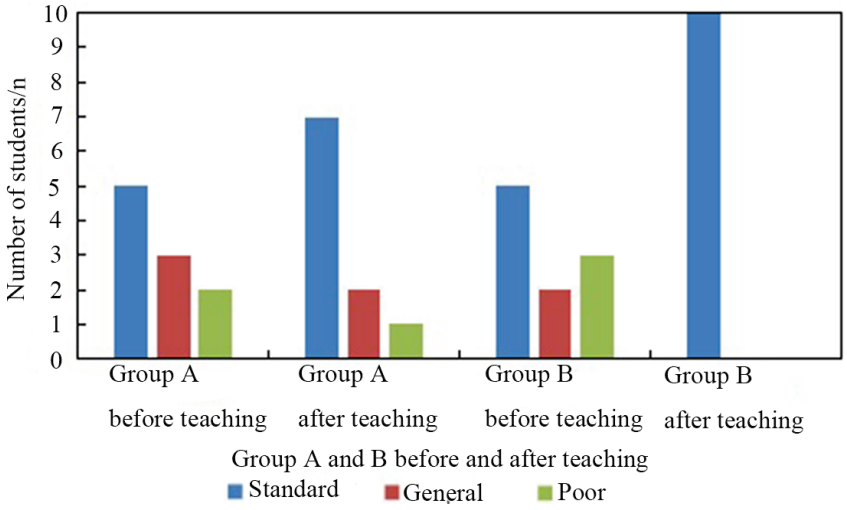

Figure 5 The pronunciation level of stress in the imperative sentence before and after teaching

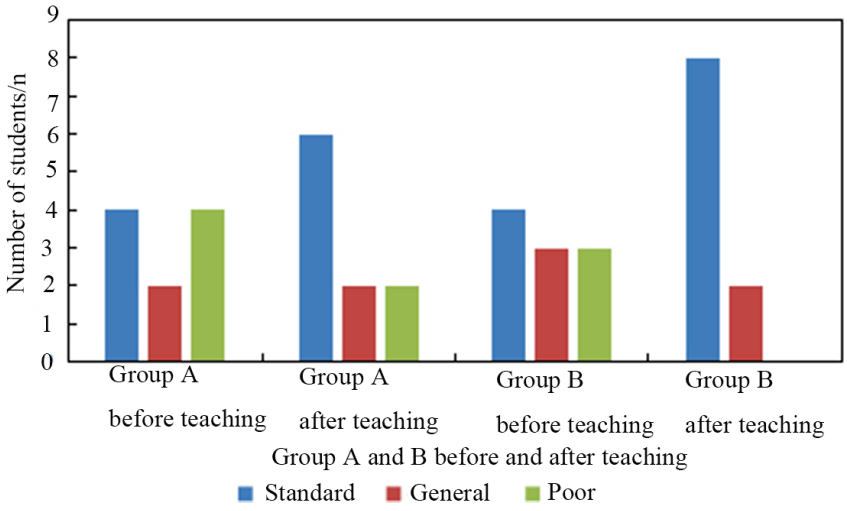

Figure 6 The pronunciation level of stress in the exclamatory sentence before and after teaching

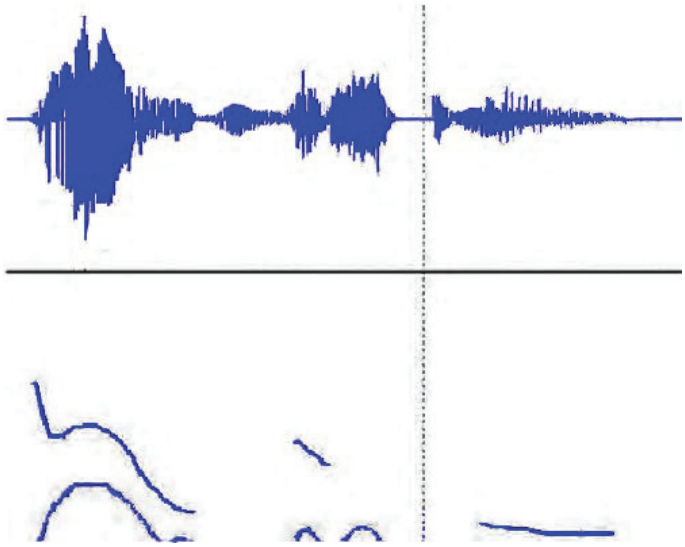

Figure 7 The spectrogram of correct pronunciation of wonderful

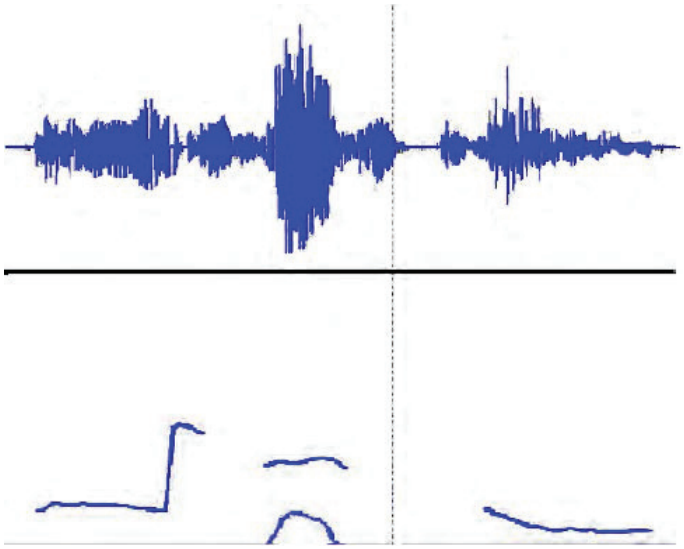

Figure 8 The spectrogram of wrong pronunciation of wonderful

pronunciation of the three sentence patterns, none of the students in group B performed poorly, with most students at a standard pronunciation level, and very few of the students at a moderate pronunciation level. In group A, the pronunciation level of the declarative sentence was balanced. The number of students with standard, moderate and poor pronunciation was similar. Many students had standard pronunciation in the imperative sentence and exclamatory sentence, but some students had poor pronunciation.

The level of pronunciation was compared between group A and B before and after teaching. As to the declarative sentence, the pronunciation of group A improved after the traditional teaching, the number of students with inaccurate stress decreased, but the teaching effectiveness was not as successful when compared with group B. After Praat assisted teaching, the number of students with standard pronunciation in group B increased sharply, the number of students with moderate pronunciation also decreased, and none of the students had poor pronunciation. For the imperative sentence, students in group A improved after the traditional teaching, and the number of students with inaccurate stress reduced, but all the students in group B performed at a standard level after Praat assisted teaching. For the exclamatory sentence, students in group A improved after the traditional teaching, and the number of students with standard stress increased; more students tended to have standard pronunciation after Praat assisted teaching, and none of the students had poor pronunciation.

In conclusion, Praat played an important role in improving the learning effectiveness of English sentences.

4.3 Teaching Efficacy of Pronunciation Fluency of English Sentences

As shown in Table 4, 40% of the students in group A had improved pronunciation fluency of the declarative sentence after the traditional teaching, 80% of the students in group B had improved pronunciation fluency of the declarative

Table 5 The number of students with correct and incorrect intonation when pronouncing different sentence patterns in group A and B before and after teaching.

Sentence pattern	Group	Correct intonation before teaching/n	Wrong intonation before teaching/n	Correct intonation after teaching/n	Wrong intonation after teaching/n
Declarative sentence	A	30	20	35	15
	B	25	25	30	20
Imperative sentence	A	35	15	40	10
	B	30	20	35	15
Exclamatory sentence	A	25	25	30	20
	B	25	25	35	15

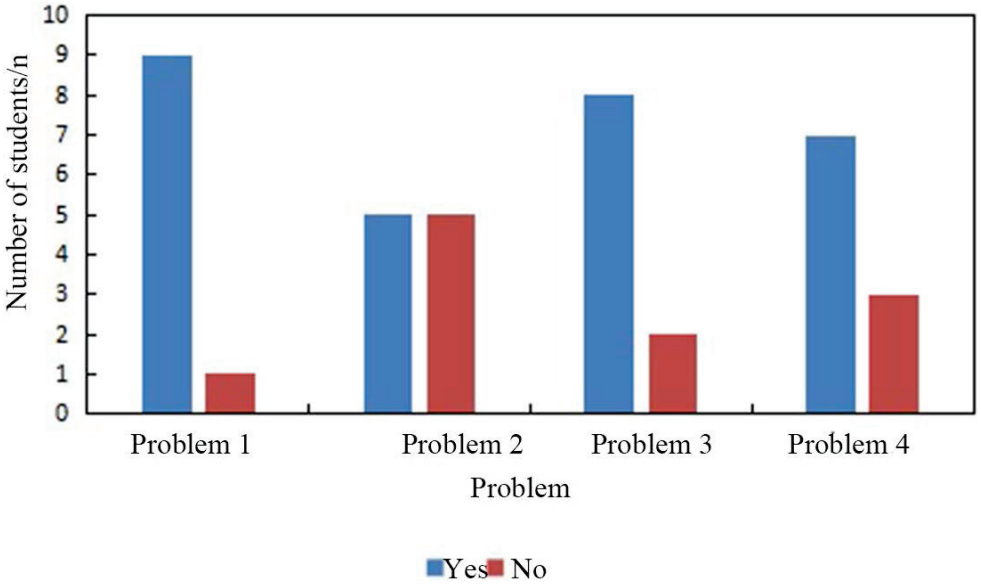

Figure 9 The questionnaire of group B after teaching

sentence of students after Praat assisted teaching. 50% of the students in group A had improved pronunciation fluency of the imperative sentence after the traditional teaching, all the students in group B had improved pronunciation fluency of the imperative sentence after Praat assisted teaching. 30% of the students in group A had improved pronunciation fluency of the exclamatory sentence after traditional teaching, and 90% of the students in group B had improved pronunciation fluency of the exclamatory sentence after Praat assisted teaching. As different people have different pronunciation habits, the time for an individual to read a sentence will differ slightly; the duration of speaking the provided sentences was used as a measure of fluency for reference in this study. In conclusion, Praat assisted teaching increased the fluency of the students' speech thereby reducing the duration of speaking the provided sentences.

4.4 Teaching Efficacy of Intonation of English Sentences

As shown in Table 5, in group A 20 students had the wrong intonation of the declarative sentence before teaching, and 15 students had the wrong intonation after teaching; the corresponding number of students in group B was 25 and 20. The improvement of intonation of the declarative sentence of group A and B was similar. For the imperative sentence, in group A 15 students had the wrong intonation before teaching, and 10 students had the wrong intonation after teaching; in group B 25 students had the wrong intonation before teaching, and 20 students had the wrong intonation after Praat assisted teaching; the improvement was very similar in both groups. For the exclamatory sentence, in group A 25 students had the wrong intonation before teaching, and 20 students had the wrong intonation after teaching; the corresponding number of students in group B was 25 and 15. Compared to the declarative sentence and imperative sentence, the improvement of intonation of the exclamatory sentence was more obvious. In conclusion, Praat assisted teaching had an insignificant effect in improving the learning of intonation of the declarative and imperative sentences, but had an increased effect in improving the learning of the exclamatory sentence.

4.5 Questionnaire of Group B After Teaching

As shown in Fig. 9, the majority of students in group B considered that the application of Praat when learning English pronunciation was a novel learning method, and only one student disagreed with this as she thought that similar English learning aided software had emerged. When asked whether the learning process was interesting and less restrictive, there was a divergence; the assentors held that the visualization

of sound in such a way was quite interesting and novel, while the dissenters thought that such a teaching method was very similar to the traditional teaching method and rote memorization was still inevitable. Most of students thought that their English level had improved and that they could communicate with English speakers fluently.

5. CONCLUSION

Fifty university students were included in this study. Firstly, the students received a pronunciation test of four front vowels, and their pronunciation was compared with a native English speaker. They were then divided into two groups; students in group A received lessons via the traditional teaching method, while students in group B received Praat assisted teaching. The English pronunciation of students in both groups was tested before teaching. The test content was three sentences, a declarative sentence, an imperative sentence and an exclamatory sentence. The pronunciation of the English speaker was taken as a test standard. The teaching content was the same in the two groups. After the end of one week of teaching, the two groups were tested again, and students in group B were surveyed using a questionnaire.

The following results were obtained. For the female students, the pronunciation of /i:/ and /ei/ were easy to pronounce, and the positions of tongue when pronouncing the two vowels were close; the pronunciation of /e/ and /i/ was difficult, the position of tongue was lower and further back than that of the English speaker when the students pronounced /e/, and the position of tongue of the students was higher and ahead of that of the English speaker when pronouncing /i/. The pronunciation level of stress of students in group A and B differed little before teaching, but students in both groups performed poorly in the pronunciation of stress in the declarative sentence and performed general in the pronunciation of stress in the imperative and exclamatory sentences. The pronunciation of stress of students in group A improved after teaching, but the improvement was not as good as that of group B. Hence it was concluded that Praat could improve the pronunciation level of stress. As to the duration of pronunciation, only 30% $\sim$ 40% of the students in group A improved while more than 80% of the students in group B improved. This indicated that Praat could effectively help reduce duration of English pronunciation and improve fluency. As to the accuracy of intonation, the improvement of group A and B differed little, especially for the declarative and imperative sentences. The questionnaire survey of group B revealed that most of the students held an affirmative attitude on Praat assisted English teaching, but there was a divergence on whether the teaching was novel and interesting.

REFERENCES

1. Bian, J., Li, S., Wang, Y., et al. (2018). A Survey of Tongue Modeling Methods in Speech Visualization. *International Conference on Robotics and Automation Engineering*, 431–435.
2. Chun, D.M., Jiang, Y., Meyr, J., et al. (2015). Acquisition of L2 Mandarin Chinese Tones with Learner-created Tone Visualizations. *Journal of Second Language Pronunciation*, 1(1), 86–114.
3. Domínguez, M., Latorre, I., Farrúsfarr'Farrús, M., et al. (2016). Praat on the Web: An Upgrade of Praat for Semi-Automatic Speech Annotation. COLING.
4. Hamlaoui, N., & Bengrait, N. (2016). Using Betteraccent Tutor and Praat for Learning English Intonation. *Social Science Electronic Publishing*.
5. Kumbhakarn, M., & Sathe-Pathak, B. (2015). Analysis of Emotional State of a Person and Its Effect on Speech Features Using PRAAT Software. *International Conference on Computing Communication Control and Automation. IEEE*, 763–767.
6. Lovato, A., De, C.W., Giacomelli, L., et al. (2016). Multi-Dimensional Voice Program (MDVP) vs Praat for Assessing Euphonic Subjects: A Preliminary Study on the Gender-discriminating Power of Acoustic Analysis Software. *Journal of Voice*, 30(6), 765.e1–765.e5.
7. Maryn, Y., & Weenink, D. (2015). Objective Dysphonia Measures in the Program Praat: Smoothed Cepstral Peak Prominence and Acoustic Voice Quality Index. *Journal of Voice*, 29(1), 35.
8. Niebuhr, O. (2017). Clear Speech - Mere Speech? How Segmental and Prosodic Speech Reduction Shape the Impression That Speakers Create on Listeners. *Interspeech*. 894–898.
9. Razavi, M., Rasipuram, R., & Magimai Doss, M. (2016). Acoustic Data-Driven Grapheme-To-Phoneme Conversion in the Probabilistic Lexical Modeling Framework. *Speech Communication*, 80(C), 1–21.
10. Saito, K., & Shintani, N. (2016). Do Native Speakers of North American and Singapore English Differentially Perceive Comprehensibility in Second Language Speech?. *Tesol Quarterly*, 50(2), 421–446.
11. Samsudin, N.S.B., & Mano, K. (2017). Comparison Between Random and Daily Speech Database in the Speech Visualization. *IEEE International Conference on Systems, Man, and Cybernetics. IEEE*, 3135–3140.
12. Shinnakasu, M., Hirahara, N., Sakata, T., et al. (2016). Evaluation of Usefulness of Speech Visualization System for Articulation Pattern Analysis in Patients with Mandibular Prognathism. *Nippon Koku Geka Gakkai Zasshi*, 62(1), 3–9.
13. Tsujioka, S., Sakti, S., Yoshino, K., et al. (2016). Unsupervised Joint Estimation of Grapheme-to-Phoneme Conversion Systems and Acoustic Model Adaptation for Non-Native Speech Recognition. *Interspeech*, 3091–3095.
14. Yamane, N. (2015). The Development of Visual CALL Materials for Learning L2 English Prosody. *ICT for Language Learning*.
15. Yang, F. (2018). Research on Speech Visualization Technology and Its Application in English Listening Teaching. International Conference on Intelligent Transportation, Big Data & Smart City.*IEEE Computer Society*, 355–358.