SLaTE 2019: 8th ISCA Workshop on Speech and Language Technology in Education
20-21 September 2019, Graz, Austria

voɪsTUTOR: Virtual Operator for Interactive Spoken English TUTORing

Chiranjeevi Yarra, Prasanta Kumar Ghosh
Electrical Engineering, Indian Institute of Science (IISc), Bangalore-560012, India
{chiranjeeviy,prasantg}@iisc.ac.in

Abstract

Second language (L2) learners of English could improve their pronunciation using automatic tutoring apps that provide detailed feedback. However, the availability of such apps are very few and, at most, the feedback from those apps are limited in the sense that they provide a minimal feedback for improving correct usage of phonemes. But, to achieve better pronunciation quality, the L2 learners are required to correctly place stress, intonation and pauses in their spoken utterances. In this demo, we present a self learning app called voɪsTUTOR that provide more detailed feedback on correct usage of phonemes as well as stress, intonation and pauses. In this app, the feedback is provided specific to each of these four categories by showing a score representing learner's quality as well as their mismatches with respect to expert's pronunciation. We believe that this app could be useful to the learners who do not have easy access to effective spoken English training.

1. Introduction

Second language (L2) learners of English are often influenced by the their nativity. However, these nativity influences could be minimized by using effective tutoring methods that train the correct aspects of using phonemes and placing stress, intonation and pauses. With the advancement of computer-aided language learning (CALL), mobile apps, that provide detailed automatic feedback on all these aspects, can act as an effective self-learning tools for the L2 learners [1]. However, the existing apps on pronunciation tutoring do not provide feedback in all these aspects and those are limited to only phoneme correctness with minimal feedback. For the benefit of L2 learners, we have developed voɪsTUTOR app which provides automatic feedback in the aspects, where the feedback is provided in real time. In this demo, we present the voɪsTUTOR app. To the best of our knowledge no similar apps are available.

2. Proposed design

The proposed app has two major parts – front-end (user interface) and back-end (web-server). The front-end is available at the learner's location and the back-end is situated at our location. Both the front-end and the back-end communicate via Internet. The learner can access the app using Android based smartphones or tablets. When a learner logs-in, it displays a screen showing four parts that cater the lessons into following four aspects to train the learner – 1) Phoneme correctness, 2) stress placement, 3) intonation usage and 4) pause placement in the sentences. The learner can improve their pronunciation skills by using feedback designed specific to each aspect. On click of each one, the lessons belonging to the respective category are displayed as shown in Figure 1. We discuss details of the lesson design, feedback in the following sub-sections along with an illustration of interface belonging to the stress placement aspect.

2.1. Lesson design

In each category, there are multiple lessons and in each lesson, the stimuli are designed to train the learners in pedagogical manner using simple to complex exercises. For example, in phoneme category, there are eight lessons, the stimuli in lesson 1 to 8 contain minimal pairs that are focused to tutor the phoneme correctness while uttering words containing fricatives, stops, nasals, glides & laterals, consonant sequences, vowels, diphthongs and vowel sequences respectively. The four lessons in stress category are focused to tutor stress placement while uttering words, words under masking, weak forms and phrases. The four lessons in intonation category are focused to tutor the intonation usage of four different types of intonation respectively. Finally, the four lessons in sentence category are focused on improving correct pause placement when uttering simple, complex, compound and long sentences.

2.2. Feedback

For each stimuli in all categories, we display a score representing pronunciation quality of learner's utterance with respect to an expert's utterance. Further, we provide expert's audio for the stimuli so that the learner can listen to it and can correct their errors. In addition, the learner can choose to view a detailed feedback, which is provided in a category specific manner by highlighting the mismatches in the learner's and expert's pronunciation in terms of scores and text messages. We detail the category specific feedback in the following subsections.

2.2.1. Phoneme correctness

- Display phonemes uttered and indicate mismatches in those phonemes in terms of phoneme insertions, deletions and substitutions.
- Display a score indicating matching quality.
- Show a video containing the correct articulatory movements embedded with expert's audio.

2.2.2. Stress placement

- Display syllable uttered and indicate stress markings on each syllable.
- Show the mismatches in the stress markings.
- Display the following values for each syllable – syllable duration and the highest and average loudness within a syllable. This is because these values significantly influence the syllable stress [2, 3].

2.2.3. Intonation usage

- Display syllable uttered and indicate pitch pattern in each syllable by stylizing the pitch within the syllable.
- Show the mismatches in the patterns for each syllable.

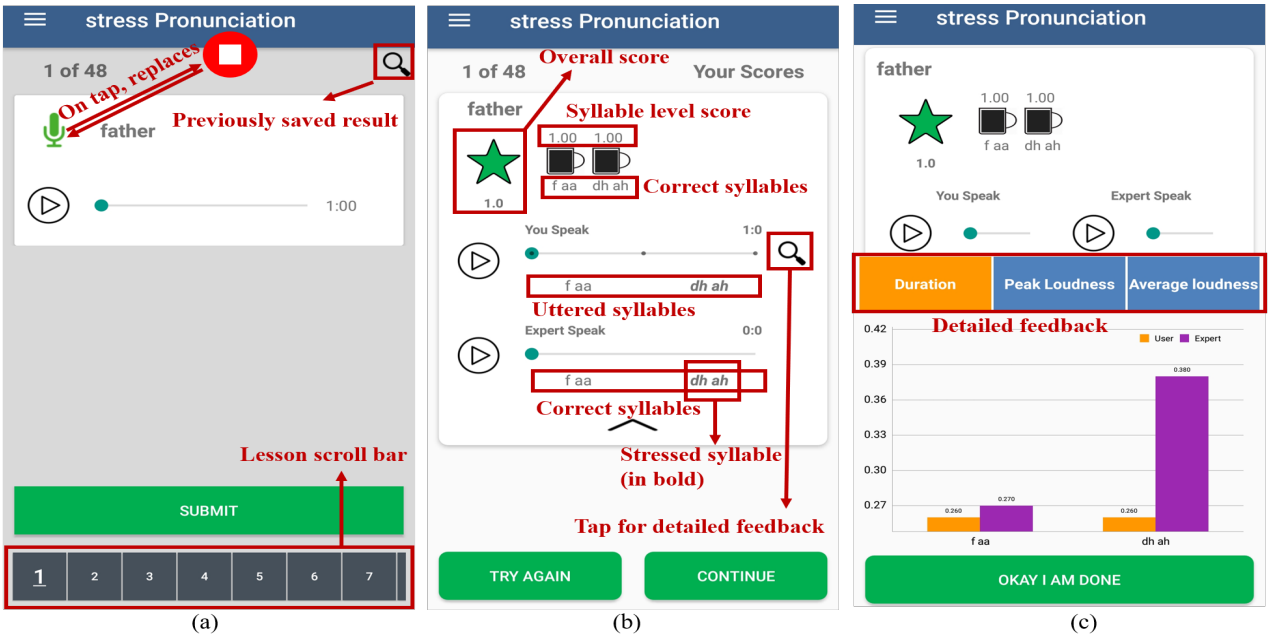
2.2.4. Pause placement

- Display syllable uttered and indicate the pauses made.
- Show the erroneous pause locations, when those are made within a word.
- Display scores for each word in a spoken utterance.

2.3. An exemplary user-interface

Figure 1a shows an exemplary user-interface (UI) that appears on the selection of a stimuli in a lesson belonging to stress category. In the UI, the horizontal scroll bar at the bottom allows

We thank the Department of Science & Technology, Government of India and the Pratiksha Trust for their support.



(a) (b) (c)

Figure 1: *An exemplary user-interface of the stress placement category in* voisTUTOR

the learner to traverse across all stimuli. The microphone symbol adjacent to each word stimuli should be tapped for recording the learner's voice. After recording, a play button appears below it, which can be used to listen to the learner's recording. With this, a learner can verify his/her recorded voice before submitting it for analysis. The learner can also continue recording until he/she has a satisfactory recording. Once the recording for each stimuli is completed, the learner can tap on the "SUBMIT" button to send the latest recording for analysis.

Figure 1b shows an exemplary UI that appears on tapping the "SUBMIT" button. In the UI, the feedback is provided with the help of mugs where the syllables below are the ones in the expert's utterance. The numbers above the mugs are the scores representing the matching quality between stress markings in learner's and expert's pronunciation for the syllable. The partially filled star to the left of the mugs represents the overall pronunciation quality score. Further, there are two play buttons, for listening to the learner's and expert's recordings respectively. Below the play buttons, syllables and the stressed syllable (in bold) in the respective learner's and expert's pronunciation are indicated.

Figure 1c shows an exemplary UI that appears after tapping the magnifying glass. In this UI, additional feedback is provided along with the feedback shown in Figure 1b. The additional feedback includes the three parameter values for each syllable in the expert's and learner's utterances. Furthermore, errors made by the learner in uttering syllable gets displayed in the form of text messages by tapping on each mug.

3. Demonstration

In order to demonstrate voisTUTOR, the uttered phonemes are obtained using forced-alignment process using Kaldi speech recognition tool kit [4] and a lexicon containing the phoneme pronunciations for each word. From these phonemes, syllables are obtained using P2TK syllabifier [5]. The front end is implemented using Android SDK and the back end server is set-up using LAMP (Linux, Apache, MySQL, PHP) stack on Ubuntu 14.04 LTS operating system. At the back end, the score computation is implemented using Python programming language. At the back-end, the stress markings and pauses are obtained by following the work proposed by Yarra et al. [3] and Ananthakrishnan et al. [6] respectively. The stimuli are taken from the material used for spoken English training [7]. We obtain the expert's audio by recording the stimuli from a voice-over artist, proficient in British English spoken communication.

4. Conclusion

We present an app, named voisTUTOR, for improving pronunciation skills of L2 learners. We design the front end with Android SDK and back end codes with Python programming language. The app provides the feedback that helps for correct pronunciation of phonemes and placement of stress, intonation and pauses. Further investigations are required to measure the effectiveness of the proposed tool as well as analyze sufficiency of the feedback parameters in the self-learning process.

5. References

[1] A. Kumar, A. Tewari, G. Shroff, D. Chittamuru, M. Kam, and J. Canny, "An exploratory study of unsupervised mobile learning in rural India," *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 743–752, 2010.

[2] L. Ferrer, H. Bratt, C. Richey, H. Franco, V. Abrash, and K. Precoda, "Classification of lexical stress using spectral and prosodic features for computer-assisted language learning systems," *Speech Communication*, vol. 69, pp. 31–45, 2015.

[3] C. Yarra, O. D. Deshmukh, and P. K. Ghosh, "Automatic detection of syllable stress using sonority based prominence features for pronunciation evaluation," *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 5845–5849, 2017.

[4] D. Povey, A. Ghoshal, G. Boulianne, L. Burget, O. Glembek, N. Goel, M. Hannemann, P. Motlicek, Y. Qian, P. Schwarz *et al.*, "The kaldi speech recognition toolkit," *IEEE workshop on automatic speech recognition and understanding (ASRU)*, 2011.

[5] J. Tauberer, "P2TK automated syllabifier," *Available at https://sourceforge.net/p/p2tk/code/HEAD/tree/python/syllabify/*, last accessed on 10-03-2018.

[6] S. Ananthakrishnan and S. S. Narayanan, "Automatic prosodic event detection using acoustic, lexical, and syntactic evidence," *IEEE transactions on audio, speech, and language processing*, vol. 16, no. 1, pp. 216–228, 2008.

[7] J. D. O'Connor, *Better English Pronunciation*. Cambridge University Press, 1980.