



Development of a pronunciation training system based on auditory-visual elements

Haruko Miyakoda

Department of English, Tsuda College, Tokyo, Japan

miyakoda@tsuda.ac.jp

Abstract

In this demonstration, we present the auditory-visual pronunciation system that we have developed. One of the key features of this system is that it employs easy-to-understand visuals of the speech organ that can be seen from different angles. In addition, the internal organs in movement can also be presented by changing the mode to transparent. Furthermore, unlike most systems that can present only the ideal model movements of the speech organs, our system allows users to freely adjust the tongue and jaw movements by controllers. This allows instructors, for example, to visually indicate and point out the deviant movement(s) of the learners so that the learners themselves can understand their present state (i.e. problems) with the help of visual information and feedback.

Index Terms: pronunciation system, speech organs, visual feedback

1. Introduction

It has been pointed out in the literature that one of the most neglected aspects in English language teaching is pronunciation [1]. Since difficulty in pronunciation can lead to communication breakdowns, it is most likely that the reason why pronunciation is marginalized in formal classroom teaching is not because instructors and teachers are making light of the pronunciation problems that L2 learners face. Rather, the reason why pronunciation teaching has not been taken up may be that although instructors are well aware of its importance, they may feel that time would be better spent by teaching grammar, for example, since learners, especially those who start to learn a new language after a certain age, will never be able to attain native-like pronunciation to begin with. However, major difficulties in pronunciation often results in the learners facing difficulty in daily life, including facing difficulty in finding employment [1]. Therefore, although there is no need to set the goal of instruction to attaining native-like pronunciation, it is very important to instruct and lead the learners so that they may attain reasonable intelligibility in speech [2] [3].

2. The visual element in pronunciation

The findings reported in the literature regarding effective language teaching/ learning methods seem to support the viewpoint that visual data plays an important role [4], [5]. In this section, we briefly review some of the research findings concerning the use of visual element in pronunciation.

Research in the field of phonology has long been dominated by a focus on the auditory aspect. However, in actual face-to-face communication, a significant source of information about the sounds a speaker produces comes from

visual cues such as lip movements [6, 7]. There are some studies that have reported that the information value of visual cues can be improved with training. In one study, hearing-impaired adults were trained in visual consonant recognition. After a total of 14 hours of training, the accuracy rate for the recognition of consonants showed dramatic improvement. For example, in the recognition of /r/, which was the most improved of all the consonants trained, the accuracy is reported to have increased from 36.1% to 88.6% after the training [8].

Although studies on the potential benefits of auditory-visual speech training for L2 learners has only recently started to gain focus, the importance of lip shapes as beneficial cues has long been recognized by language instructors in teaching English as a second language. For example, a study had already been published more than 40 years ago claiming that the degree of difficulty that lies in acquiring the phonemes of a foreign language may be due to the difference in visual cues [9]. In this study, the difficulty that Japanese learners of English face in making the distinction between /r/ and /l/ is taken up. While this difficulty is usually attributed to the fact that these two phonemes do not exist in Japanese, this study claims that the difficulty is due to the disadvantage of not being able to “read the lips of the speaker”, or to put it another way, these two phonemes lack visual cues. Since what takes place inside the mouth cannot be visualized in daily life, the only visual cue that we can rely on is usually the lips. However, if there were a way to present the state of the mouth in movement, then this might make it easier for the learners to link this information into actual physical action.

3. The system

Based on the findings of previous studies, we decided that our pronunciation system will be characterized by making good use of visual information, especially the movements that take place inside the mouth.

While there are numerous systems already in existence that give excellent detailed account of the movements of the speech organs (e.g. [10]), one of the major drawbacks is that although the learners can see how the *ideal* movements should be, they cannot receive visual feedback on what they are *doing wrong*. In other words, these systems, although are very useful in presenting the ideal state, they do not make it possible for learners to actually understand what their own problems are because there is no way the learners can visualize what takes place inside their mouths. Another problem with the existing systems is that in most cases, the articulatory diagrams can only be viewed from one side, i.e., the side view. This makes it extremely difficult for learners to link the visual information into actual physical action.

With the pronunciation system that we have developed, the users have access to the speech organs from two different angles: i.e., the side view (cf. Figure 1) and the front view (cf. Figure 2)

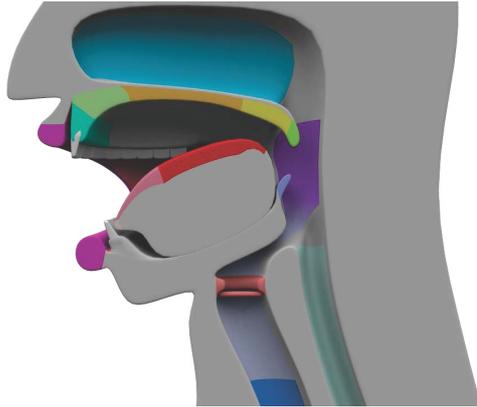


Figure 1: *The side view*

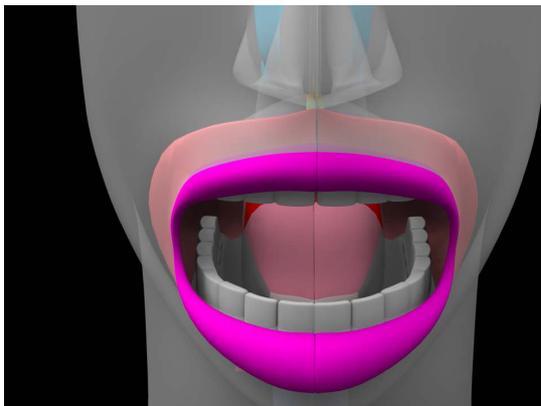


Figure 2: *The front view*

In addition to these two angles, there is the transparent mode as indicated in Figure 3.

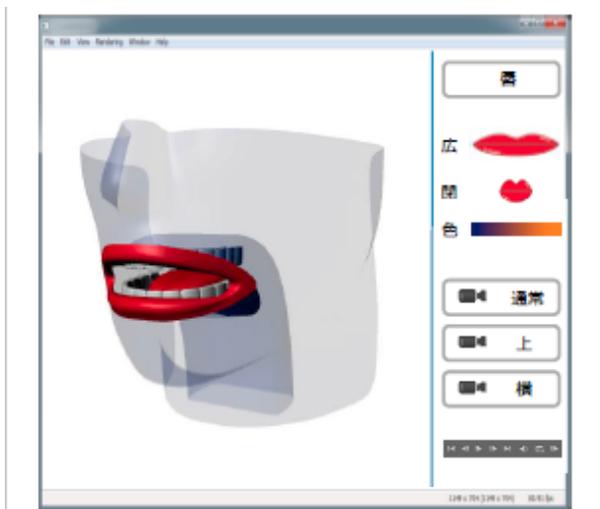


Figure 3: *The transparent mode with controllers*

The transparent mode allows users to take a look into the internal organs in actual movement. The user is able to rotate the angle at any degree that he/she likes while slowing the speed of animation, or taking pauses or freezing the frame in cases where they would like to replicate the movements that take place. We have also added on controllers that allow users (particularly instructors) to freely move and adjust the lip,

tongue, and jaw movements so that the deviant movements can easily be represented in visual form. (the controllers are indicated in the right column of Figure 3).

At present, we are compiling a list of English words which are claimed to be commonly difficult for Japanese to learn, but in addition to learners of foreign languages, this system can also be used to train speakers of pathological speech in their native languages. The transparent mode for the speech organs is expected to be especially useful for the hearing impaired population as well as others with severe aural and oral communication difficulties.

4. Conclusion

In this paper, we reported on the pronunciation system that we are developing. This practical and productive pronunciation software should be enjoyable to use for foreign language learners of all ages as well as speakers of pathological speech receiving training in their native languages.

5. Acknowledgements

The author would like to thank Fumiki Teratani, Tomohide Kano, Masahiro Tachibana and Akira Ishii for their contribution to this study. This study is funded by Grants in Aid for Scientific Research (C) (No. 23520594).

6. References

- [1] Gilakjani, A. P. and Ahmadi, M.R., "Why is pronunciation so difficult to learn?", *English Language Teaching*, 4(3) 74-83, 2011.
- [2] Scovel, T. *A time to speak. A psycholinguistic inquiry into the critical period for human speech.* Rowley, MA, Newbury House, 1988.
- [3] Celce-Murcia, M. Brinton, D.M. and Goodwin, J.M., *Teaching pronunciation: a reference for teachers of English to speakers of other languages.* Cambridge University Press, Cambridge, 1996.
- [4] Al-Seyghayar, K. "The effect of multimedia notation modes on L2 vocabulary acquisition: a comparative study", *Language Learning and Technology*, 202-232. 2001.
- [5] Yeh, Y. and Wang, C. "Effects of multimedia vocabulary annotations and learning styles on vocabulary learning", *CALICO Journal*, 21 (1), 131-144. 2003.
- [6] McGurk, H. and MacDonald, J. "Hearing lips and seeing voices", *Nature* 264, 746-748. 1976.
- [7] Hardison, D.M. "The visual element in phonological perception and learning", in M. C. Pennington [Ed], *Phonology in Context*, 135-158, Palgrave, 2007.
- [8] Walden, B. E., Prosek, R.A., Montgomery, A.A., Scherr, C.K., and Jones, C.J. "Effects of training on the visual recognition of consonants", *Journal of Speech and Hearing Research* 20, 130-145. 1977.
- [9] Goto, H. Auditory perception by normal Japanese adults of the sounds "l" and "r". *Neuropsychologia* 9, 317-323. 1971.
- [10] The university of Iowa, "The sounds of spoken language", Online: <http://www.uiowa.edu/~acadtech/phonetics/index.html#pluginCheck>, accessed on 2 Apr 2013.