Android version of *Visual Learning*

Hayato Watanabe[1], Ian Wilson[2] and Kyori Suzuki

CLR Phonetics Lab, University of Aizu, Japan

[1]*s1230063@u-aizu.ac.jp*, [2]*wilson@u-aizu.ac.jp*

Abstract

Upon entering Japanese university and communicating with international students and foreign professors, Japanese undergraduates typically have many experiences when their English or other second language cannot be understood by their interlocutors. In order to learn how to pronounce and communicate in a way that can easily be understood by others, it would probably be useful to have an application that could evaluate one's pronunciation and give feedback automatically, especially because pronunciation learning takes a lot of time.

We developed *Visual Learning 2* for iOS (*Visual Learning 2: Pronunciation app using ultrasound, video, and MRI*) in 2017 [1]. Now we have developed an initial version of this application for Android with more useful features such as lip-shape tracking and feedback.

1. Introduction

Software applications can be useful for improving the English pronunciation of learners. In most English-as-a-Foreign-Language (EFL) classes, teachers cannot spend much time on individual students to teach pronunciation, because the class sizes are too large. However, smartphone apps can help improve English pronunciation through student-centered practice.

Visual Learning for iOS, an English pronunciation app for second-language (L2) learners and phonetics students, was developed and has been downloadable from the App Store since November 2017. Our aim is to create an Android [2] version of this app with new features such as giving feedback on the user's articulation.

Our app links together audio, front and side video, midsagittal MRI images of English phonemes, and ultrasound movies of a native speaker of Canadian English reading a phonetically balanced text. Users can watch and shadow front and side video overlaid with an ultrasound tongue movie. They can play the video at three speeds and start the video from any word by tapping on it, with a choice of display in either English or the International Phonetic Alphabet (IPA). Users can record their own audio/video and play it back in sync with the model for comparison. The app can give feedback on users' lip movements. This feedback is useful for users to correct their pronunciation, because lip movement is an important factor that influences the pronunciation (acoustic signal produced) by learners [3].

2. Feature of the app

The app shows the video images with a list of all IPA phonemes in the Wolf passage [4] and corresponding midsagittal MRI images of a native speaker of Canadian English.

Another feature of *Visual Learning* is to display that same native speaker's 'Wolf' passage reading. The left side of Figure 1 shows front and side videos of the speaker. The side video is overlaid with an ultrasound movie of the tongue's surface (white line) moving in the mouth, and shows the palate (yellow line determined via MRI overlay). Two slow-motion speeds are available – ¾ speed and ½ speed. The user can record his/her face and voice, while simultaneously playing the model video and shadowing the top half of the display.

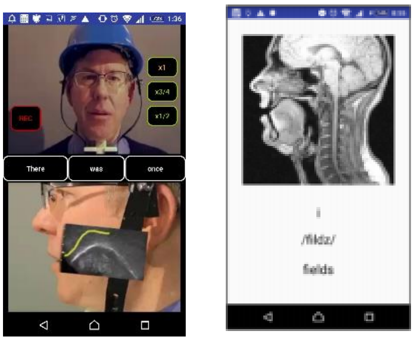


Figure 1: Screenshot of *Visual Learning*

Phrases can be displayed with subtitles in English or in IPA like the right side of Figure 1. Wearing earphones ensures that only the user's voice is recorded. After this, the user can load the recorded video from his/her smartphone video library. The

10.21437/ISAPh.2018-24

user can then play both videos (top and bottom) simultaneously for comparison [1].

As a new feature, the app calculates lip aperture and gives feedback to users. Using OpenCV (Open Source Computer Vision Library) [5], the app can recognize users' lips in a recorded video and compare lip roundness, horizontal lip aperture, and vertical lip aperture between minimal pairs of phonemes [6]. This information can be used to classify the users' pronunciation based on a phoneme-to-viseme map [7]. From this feedback, users can understand how to modify their articulation directly.

3. Method for feedback

Feedback is given by evaluating the lip shape of the user. First, lip features are detected. Then the lip roundness and aperture are calculated. Finally, the app guides users how to change their lip shape in order to correct or modify their pronunciation.

3.1. Face feature detection libraries

There are many kinds of libraries for mouth detection which can be used by programs for implementing mouth detection functions. Figures 2 and 3 show examples of libraries.

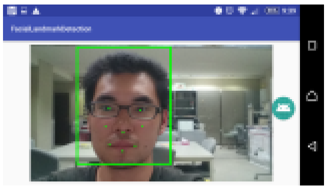


Figure 2: Face detection with Google FaceDetector API for Android

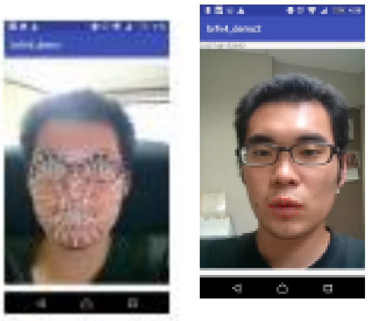
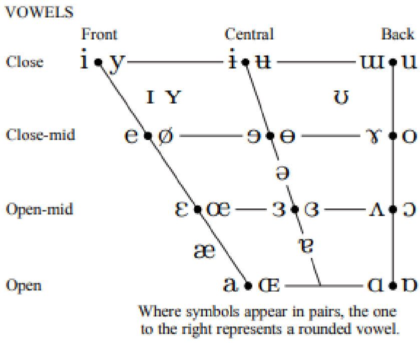


Figure 3: facial landmark and lip feature tracking with BRFv4

Some of the libraries are open source for free use. The detection quality is different from library to library. Among them, BRFv4 has a good performance of real time tracking of 68 facial landmarks. Positions of each landmark can be obtained. This library uses Haar Cascade frontal face detection as the object detection algorithm.

4. Lip circularity for vowel sound

Lip aperture and roundness are important features for making vowel sounds. The roundness of the lips affects the formants of the vowel, and these acoustic properties are what distinguishes each vowel.



Where symbols appear in pairs, the one to the right represents a rounded vowel.

Figure 4: The vowel chart used in the International Phonetic Alphabet (IPA).

Figure 4 shows the IPA vowel chart and it shows which vowel sounds are considered "round" (the symbol on the right of any pair of symbols). To define the roundness of the lips when vowel sounds are produced, we used the circularity of inner and outer contours of the lips, which was calculated using the formula below:

$$\text{CIRCULARITY} = 4 * \pi * \text{AREA} / \text{PERIMETER}^2 \quad (1)$$

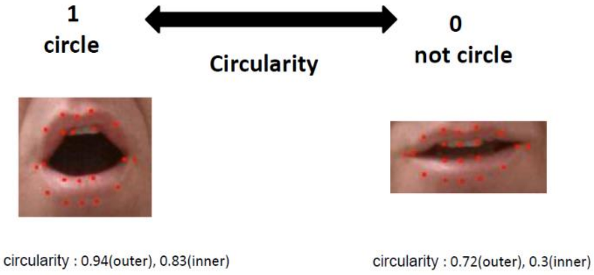

1
circle
Circularity
0
not circle
circularity : 0.94(outer), 0.83(inner)
circularity : 0.72(outer), 0.3(inner)

Figure 5: Example of the round and non-round lip shape with their inner and outer contour circularity.

5. Results

To measure the roundness for vowel sounds, three native speakers (two Canadian and one British) and three non-native speakers (one each of French, Vietnamese and Japanese) pronouncing English vowel sounds were recorded and their lip circularity was calculated.

Figure 6 shows the circularity of lip contours when "peel" and "pill" were pronounced. For native speakers, the circularity difference between two vowels is very small. For non-native speakers, the difference is more significant. Figure 7 shows the difference when “work” and “walk” are pronounced. For Canadian, French and Vietnamese, the vowel in "walk" is more rounded. For British and Japanese, the vowel in "work" is more rounded. Figure 8 shows the circularity calculation for the pronunciation of “bird” and “bored”. For British and Japanese, the inner contour has much smaller circularity when "bored" is pronounced compared to "bird". For the other speakers, the circularity for the two vowels are almost the same or "bored" is more rounded. Figure 9 is when “but” and “boot” are pronounced. For native and non-native speakers, the circularity of “but” is higher. Figure 10 shows the circularity calculations for when “pool” and “pull” are pronounced. For Canadians, the circularity of the vowels are almost same. For British, French and Japanese, the circularity of “pull” is higher. Figure 11 shows when “bag” and “beg” are pronounced. For natives and non-natives, the circularity of “bag” is higher than “beg”. The circularity calculations in Figure 12 are for when “lack” and “luck” are pronounced. For native and non-native speakers, the circularity of “lack” is higher or almost the same as “luck”. Figure 13 is when “hell” and “hill” are pronounced. For native and non-native speakers, the circularity of “hell” is higher than “hill”.

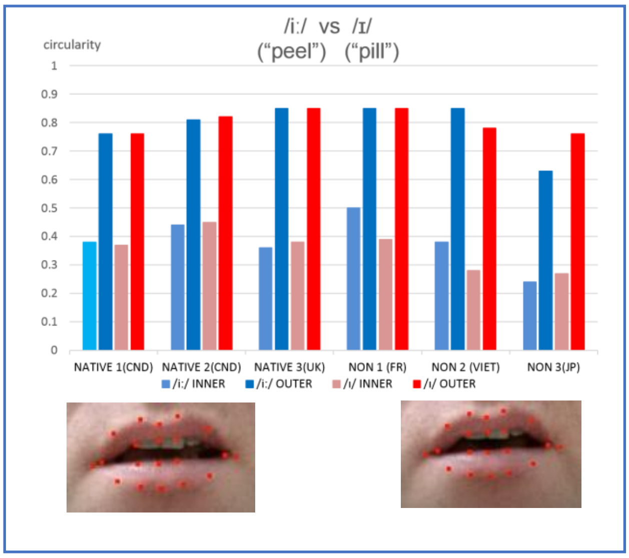


Figure 6: Graph showing circularity of inner and outer contour pronouncing two vowels in “peel” and “pill”

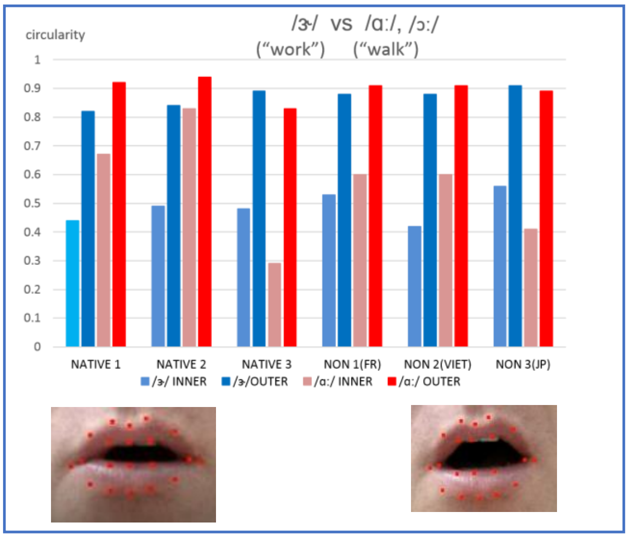


Figure 7: Graph showing circularity of inner and outer contour pronouncing two vowels in “work” and “walk”

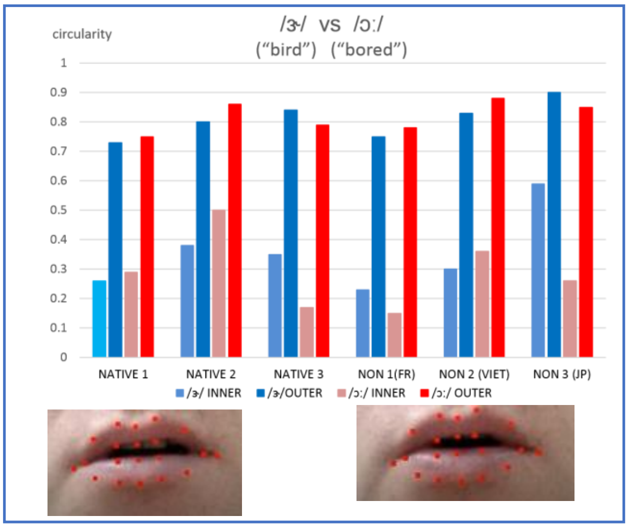


Figure 8: Graph showing circularity of inner and outer contour pronouncing two vowels in “bird” and “bored”

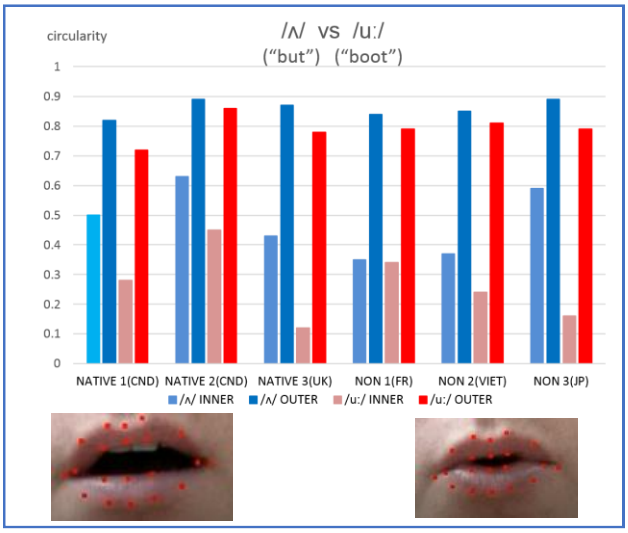


Figure 9: Graph showing circularity of inner and outer contour pronouncing two vowels in “but” and “boot”

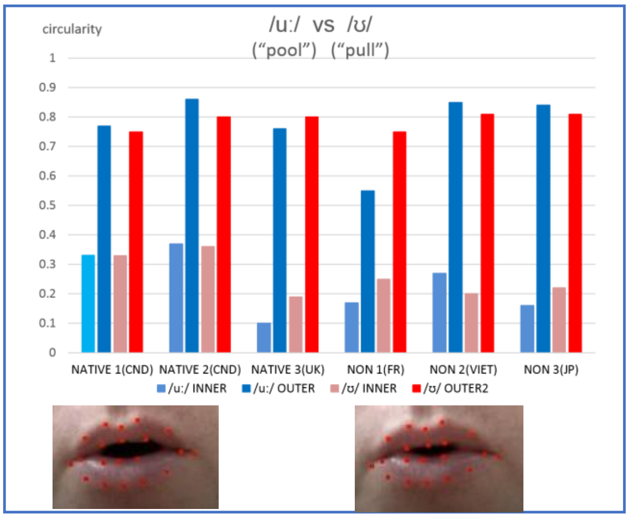


Figure 10: Graph showing circularity of inner and outer contour pronouncing two vowels in "pool" and "pull"

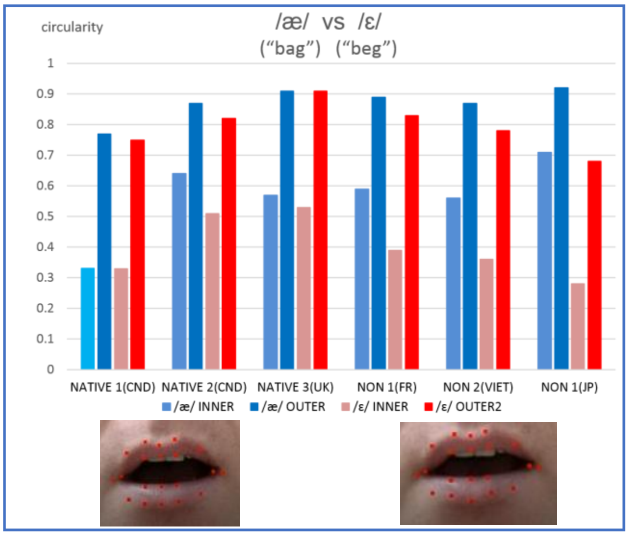


Figure 11: Graph showing circularity of inner and outer contour pronouncing two vowels in "bag" and "beg"

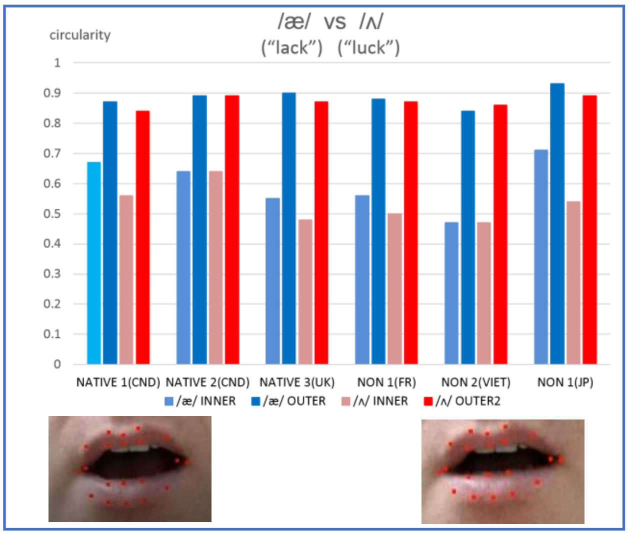


Figure 12: Graph showing circularity of inner and outer contour pronouncing two vowels in "lack" and "luck"

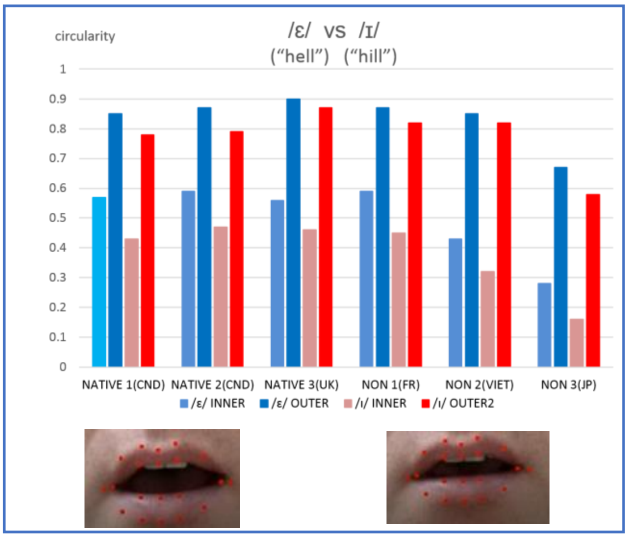


Figure 13: Graph showing circularity of inner and outer contour pronouncing two vowels in "hell" and "hill"

6. Discussion

Naturally, we expected round vowels (as defined in the IPA chart) to have a higher circularity value than unrounded vowels. Interestingly though, open vowels showed the result of higher circularity. So, for distinguishing the so-called "round vowels" from so-called "unround vowels", the movement (i.e., dynamic, not static data) of the lip corners would be more useful than circularity measurements. Circularity of the lip shape varies by individuals. For making the feedback feature of our app, it is important to think about this variety and how we can make useful feedback using visual information.

Another point to note was that lip tracking had less accuracy in some conditions like (a) the user recording in a very dark or a very bright place, (b) sudden movements of the target object (i.e., the user's lips), or (c) not including the whole of user's face on the display. Other than in these conditions, the tracking was fluent and accurate.

About converting the iOS version of Visual Learning into an Android version, we could mostly finish except for the feedback function. After implementing recording and feedback functions, we hope to publish the Android version of the app soon so that more people can use our app. For making feedback, we can focus on some specific phonemes like vowel sounds as a first step. We are going to think about an efficient way to give feedback using lip tracking functions and other visual information.

7. Acknowledgements

We would like to thank Jeremy Perkins, Allan Nicholas, and Ari Saito for their support.

8. References

[1] K. Suzuki, I. Wilson, and H. Watanabe, “Visual Learning 2: pronunciation app using ultrasound, video, and MRI,” Proceedings of Interspeech 2017, pp. 831-i

[2] Android.com, “Download Android Studio and SDK Tools | Android Developers”, 2018. [Online]. Available: https://developer.android.com/studio/ . [Accessed: 26-May-2018]

[3] S. Wei, Y. Chen, and T. McGraw, “Computer-vision-aided lip movement correction to improve English pronunciation,” Proceedings of the 122nd ASEE Annual Conference & Exposition, Paper ID #13112, 2015.

[4] D. Deterding, “The North Wind versus a Wolf: short texts for the description and measurement of English pronunciation,” JIPA, 36: 187-196.

[5] A. Kaehler and G. Bradski, “Learning OpenCV 3”. O’Reilly: Sebastopol, CA, 2017.

[6] Eslan, “Minimal pairs for vowels,” 2018. [Online] Available: http://englishspeaklikenative.com/minimal-pairs/minimal-pairs-for-vowels/ [Accessed: 1-Jul-2018]

[7] H. Bear, and R. Harvey, “Phoneme-to-viseme mappings: the good, the bad, and the ugly”, Speech Communication, vol. 95, pp. 40–67, 2017.