The Use of 2D Real-Time Visual Feedback Software in Foreign Vowel Sound Production Training

Nobuhisa Hiraishi

Graduate School, Nagoya University, Japan
hiraishi.nobuhisa@j.mbox.nagoya-u.ac.jp

Abstract

SonicPrint™, a free 2D real-time visual feedback software for windows was adjusted and installed in a CALL room at a university in Japan and applied for English vowel pronunciation training with a total of 97 students for four times. The vowel sounds produced by the students were recorded on three occasions, pre-training, post-training, and one month after the post-training. The formant frequency measurements of eight randomly-selected students were analysed and mapped on F1-F2 vowel charts. Although a powerful L1 vowel influence was observed, the vowel training software has shown promising results.

Keywords: SonicPrint™, Real-time visual feedback, L2 vowel production, Vowel constellations, F1-F2.

1. Introduction

The English language has different vowel constellations than Japanese. It is said that there are at least ten distinctive monophthongal vowel sounds (e.g., /iː, ɪ, ɛ, æ, ɑː, ʌ, ɔː, uː, ʊ, ɚ/) in English, while Japanese has merely five (e.g., /A, I, U, E, O/). It is well known that English vowels are difficult for Japanese learners to produce distinctively. To help improve L2 vowel pronunciation training methods, the usefulness of some visual feedback applications were studied at the end of last century [1, 2, 3]. As the technology has improved dramatically in this century, visual feedback applications with an F1-F2 inversed vowel chart have recently been presented and studied [4, 5, 6]. It is usually rather expensive to implement those applications. In this study, the author introduces a free software that provides real-time 2D visual feedback for vowel production training.

2. Training software

SonicPrint™ [7], provided by Arcadia Inc., is a free software that plots the vowel formant frequencies you produced on a 2D vowel chart in real time. On the chart, F2 is shown on the x-axis and F1 on the y-axis inversely. That means lower F1 values would be shown at a higher position, and lower F2 formant data would be shown to the right. Although the F1 and F2 do not actually represent the highest point of the tongue, they still give intuitive information for learners to imagine their tongue position in the articulator. The free "French β-version" for windows is the one with which you can set your target vowel inventories by adjusting or adding vowel formant frequency information on the data files in the "regions" folder. The target vowels appear in the shape of circles or ellipses on the screen. As you start recording your utterances, your vowel sounds are plotted on the screen as red dots in real time (see Figure 1). When you replay the recorded vowel sounds, you can confirm the vowel sounds both visually (with red dots) and auditorily. As is shown in Figure 1, ten English vowels (/iː, ɪ, ɛ, æ, ɑː, ʌ, ɔː, uː, ʊ, ɚ/) have been displayed on the screen.

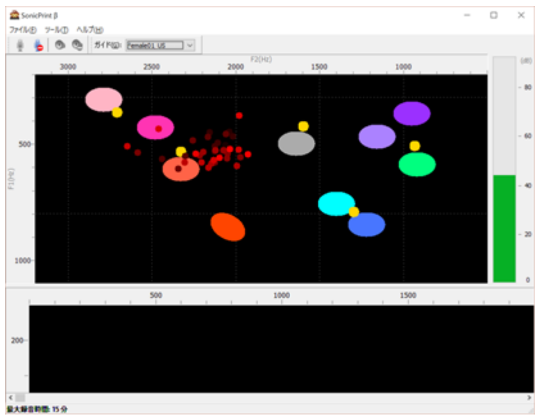

Figure 1: A screen shot of SonicPrint™. The 10 ellipses represent the target (English) vowels /iː, ɪ, ɛ, æ, ɑː, ʌ, ɔː, uː, ʊ, ɚ/, while the 5 small circles represent Japanese vowels /A, I, U, E, O/. Your uttered vowel sounds are plotted as small red dots in real time (F2 on y-axis and F1 on x-axis in inverse position).

Unfortunately, the display language is only Japanese, so those who do not understand Japanese alphabets *(Hiragana, Katakana, and Kanji)* could have some difficulties in operating the software.

10.21437/ISAPh.2016-22

3. Experiment

3.1. Training

SonicPrint™ was installed in each computer in a CALL room by the IT staff at a university in Japan during the summer holiday in 2015. Four different vowel inventories (US, LA, NY, UK) for female and male were prepared as the target vowels. The vowel data for US is from [8], LA and NY from [9], and UK from [10]. Students were allowed to try any of them in the training. A total of 97 first year students in three classes (one pre-intermediate and the other two intermediate levels with the faculty's regulation) received the training for four weeks (once a week) at the beginning of the autumn semester in September 2015. Each training session lasted for about 15 minutes. The author produced English vowel sounds as audio stimuli for students only a few times throughout the sessions. Otherwise students did not get any audio stimuli during the training. They observed plotted dots while producing the vowel sounds, and tried to get them closer to the target ellipses. They were asked to record their English vowel production with the school-adjusted version of Audacity® at the end of each training session without utilizing SonicPrint™.

3.2. Materials and recordings

A total of 11 monophthongal English vowels /iː, ɪ, ɛ, æ, ɑː(ɒ), ʌ, ɔː, uː, ʊ, ɚ(ɜː), ə/ were selected, and seven words with one or two syllables were selected for each vowel as the stimuli (e.g., 'meet', 'beat', 'deep', 'keep', 'speak' need', 'heed' for /iː/). The seventh word of each word-set has a word form of hVd except for /ʌ/ whose word-set contains 'hub' and 'hug' instead, as 'hud' does not exist in English. (For the analysis of /ʌ/, the mean value of 'hub' and 'hug' have been applied.) The vowel productions were recorded on three occasions, pre-training (In April 2015), post-training (at the end of October 2015), and four weeks after the post training session (at the end of November 2015). The five month period between the pre- and post-training was due to the CALL room regulation as the installation of any new software is available during long holidays only. At the first recording, they were asked to produce each Japanese vowel seven times in addition to the English vowels.

3.3. Subjects

The data sets of eight students (four females, four males) were selected randomly for the analysis. None of them had been abroad at the time of the experiment.

3.4. Data analysis

The formant measurements of the vowels in hVd form ('heed', 'hid', 'head', 'had', 'hawed', the mean of 'hub' and 'hug', 'hod', 'who'd', and 'hood') were taken on Praat v.6.0.14 [11] and listed on excel sheets manually. The original measurements in Hertz were normalized with the online vowel normalization tool NORM [12]. The *Lobanov* method was selected for the normalization since it has been found to be one of the most reliable methods for vowel normalization [13, 14, 15]. It applies the z-score formula for normalization.

$$F_i^N = \frac{F_i - \mu_i}{\sigma_i} \tag{1}$$

4. Results and discussions

Z-cores of the mean formant frequencies and standard deviations for the formant are shown in Table 1. As the vowel chart presents the formant data inversely, smaller negative F1 values indicate higher tongue position, while larger positive F1 values suggest lowering the tongue and jaw. Smaller negative F2 values indicate backing of the tongue while higher positive F2 values suggest tongue fronting.

Table 1: Normalized mean formant measurements (z-scores) and standard deviations of the target English vowels produced by eight JEFL learners at pre-, post-training, and one month after the post-training session (delayed).

Pre-training		iː	ɪ	ɛ	æ	ɑː	ʌ	ɔː	uː	ʊ
F1	*M*	-0.99	-0.95	-0.17	1.28	0.12	1.19	0.20	-0.74	-0.71
	SD	0.24	0.28	0.40	0.39	0.36	0.50	0.65	0.25	0.19
F2	*M*	1.58	1.48	0.84	-0.40	-1.00	-0.58	-0.97	-0.42	-0.31
	SD	0.14	0.15	0.15	0.24	0.15	0.21	0.18	0.43	0.47

Post-training		iː	ɪ	ɛ	æ	ɑː	ʌ	ɔː	uː	ʊ
F1	*M*	-1.05	-0.69	0.42	1.60	0.96	1.15	0.40	-0.93	-0.61
	SD	0.20	0.35	0.31	0.32	0.51	0.68	0.58	0.25	0.22
F2	*M*	1.64	1.41	0.77	-0.13	-0.81	-0.71	-1.10	-0.78	-0.27
	SD	0.13	0.39	0.23	0.42	0.25	0.14	0.09	0.45	0.31

Delayed		iː	ɪ	ɛ	æ	ɑː	ʌ	ɔː	uː	ʊ
F1	*M*	-1.22	-1.05	0.36	1.71	0.66	1.57	0.64	-1.04	-0.64
	SD	0.21	0.34	0.41	0.63	0.56	0.61	0.52	0.19	0.24
F2	*M*	1.68	1.57	0.94	-0.14	-1.00	-0.51	-1.03	-0.59	-0.09
	SD	0.15	0.13	0.19	0.60	0.07	0.29	0.15	0.31	0.19

Table 2: Euclidean distances between pre-training (1st), post-training (2nd), and delayed (3rd).

	iː	ɪ	ɛ	æ	ɑː	ʌ	ɔː	uː	ʊ
2nd - 1st	0.09	0.27	0.59	0.42	0.86	0.14	0.24	0.40	0.10
3rd - 2nd	0.18	0.39	0.19	0.11	0.35	0.47	0.25	0.21	0.19
3rd - 1st	0.26	0.13	0.53	0.50	0.54	0.39	0.44	0.34	0.23

Table 2 shows Euclidean distances of formant data between pre-training, post-training, and one month after the post-training recording. The largest

(0.89) is the low back vowel /ɑː/ between pre- and post-training.

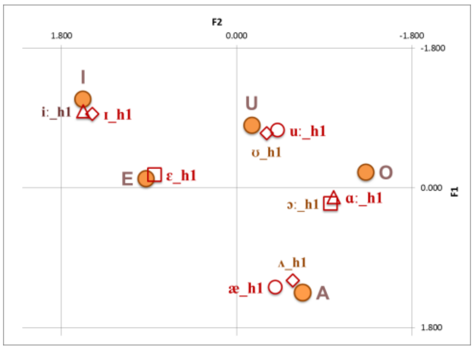

Figure 2: English and Japanese vowel constellations produced by 8 JEFL learners at pre-training session.

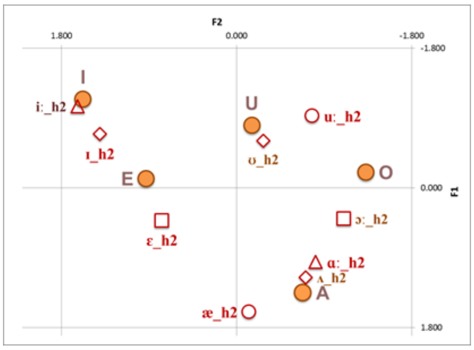

Figure 3: English and Japanese vowel constellations produced by 8 JEFL learners at post-training session.

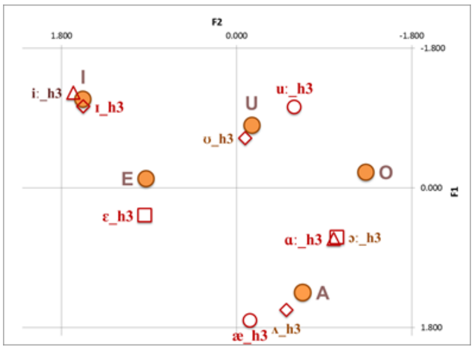

Figure 4: English and Japanese vowel constellations produced by 8 JEFL learners at the session one month after the post-training.

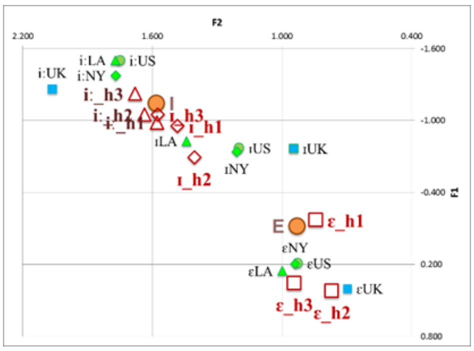

Figure 5: /iː, ɪ, ɛ/ constellations from three recording sessions (‘h1’ from pre-training, ‘h2’ post-training, and ‘h3’ delayed).

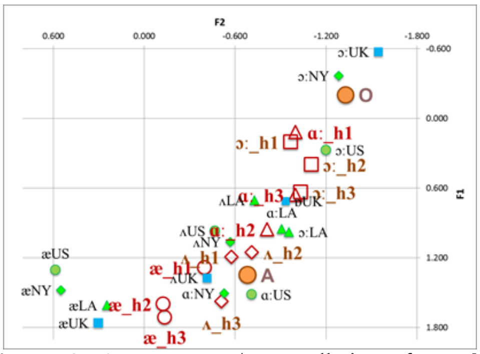

Figure 6: /æ, ɑː, ʌ, ɔː/ constellations from three recording sessions (‘h1’ from pre-training, ‘h2’ post-training, and ‘h3’ delayed).

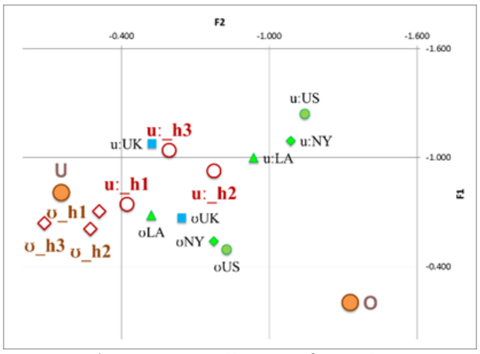

Figure 7: /uː, ʊ/ constellations from three recording sessions (‘h1’ from pre-training, ‘h2’ post-training, and ‘h3’ delayed).

The data in Table 2 were scatter-plotted with excel graphic, separated per recording session, and shown in Figure 2, 3, and 4. Figure 3, 4, and 5 present vowel constellations of /iː, ɪ, ɛ/, /æ, ɑː, ʌ, ɔː/, and /uː, ʊ/ respectively. The labels on the maps ‘h1, h2, h3’ indicate the mean values of vowel sounds in the words ‘hVd’ recorded at different sessions, i.e., ‘h1’ recorded at pre-training, ‘h2’ post-training, and ‘h3’ delayed.

As mentioned previously, the largest Euclidean distance is 0.89 with /ɑː/ between pre- and post-training. /ɑː/ has 0.54 (third largest) between pre-training and delayed. Figure 4 shows that /ɑː/ at pre-training is closer to Japanese /O/, but has shifted toward Japanese /A/ around which the target vowels /ɑː/ are scattered. This suggests that they were more or less successful in producing /ɑː/ sounds just after the visual feedback training. However, as ‘h3’ shifted backwards towards Japanese /O/, the target vowel sounds did not settle in some of the students, but the Japanese tense vowels /A/ and /O/ might have shown powerful magnet effects.

The second largest is 0.59 with /ɛ/ between pre- and post-training. /ɛ/ has the fourth largest 0.53 between pre- and delayed. Figure 3 shows that the English /ɛ/ and Japanese /E/ is almost overlapping, which indicates that students merely produced Japanese /E/ sounds when producing /ɛ/ with the word ‘head’ when they knew nothing about English vowel constellations. However, their /ɛ/ sound with 'head' is very close to the target vowels at post-training session (h2). Not only h2, but h3 is also closer to the target vowel and keeping a certain distance from h1, which indicates that they were not only successful in producing /ɛ/ with lower tongue position but also successful in preserving the sound.

Table 2 and Figure 2, 3, 4, 5 show that /ɪ/ of h1 is very close to Japanese /I/ but that gets away from /I/ towards target group directions successfully with h2. However, h3 goes back to /I/. This implies three possibilities: The first is that English /ɪ/ is difficult for JEFL learners to maintain. The second is that Japanese tense vowel /I/ has considerably powerful effects on surrounding vowel sounds like a magnet. The third is that the L2 vowel training utilizing 2D real-time visual feedback helps produce a new vowel sound that is not very far away from an L1 category. In other words, SonicPrint™ is sufficient and useful for L2 vowel training.

The traces of /æ/ and /uː/ indicate students' effort in producing English vowel sounds that are far from their L1 categories. Although both of them are halfway between Japanese vowel sounds and the target vowel sounds, h3 productions 'had' and 'who'd' do not seem to be getting back towards Japanese vowels. However, it may be too early to surmise they have created a new category on their vowel inventories. They might have had a stopover and might shift elsewhere.

/iː, ʌ, ʊ/ seem to have settled around Japanese vowels /I, A, U/ respectively. /ɔː/ sounds are between Japanese /O/ and /A/ at any time, but shifting slightly towards /A/, which is a distinctive feature of American LA (or GA) vowel sounds of /ɔː/.

The *SD* of /iː/ and /ɔː/ in Table 1 are relatively small, indicating those sounds produced by eight speakers are relatively similar in quality. The *SD* of /æ/ looks the largest, implying that the speakers are not sure of the differences between [æ, a, ɐ, a, ɑ]. Bearing that in mind, the h2 and h3 records of /æ/ suggest a positive effect of 2D real-time visual feedback as they keep some distance from /A/. Also the *SD* of F1 formants /ɑː, ʌ, ɔː/ are rather large, suggesting that English vowel sounds around Japanese /A, O/ could be problematic for Japanese learners of English.

5. Conclusions

Although some L1 tense vowels show powerful influence on L2 vowel production, the use of 2D real-time visual feedback software, SonicPrint™, has shown relatively positive results in this study. As implementing this kind of software is not usually a simple and easy process, SonicPrint™ appears to be sufficient for L2 vowel production training. The author would like to suggest that this kind of application should be utilized by those Japanese classroom teachers who are lacking confidence in improving their L2 pronunciation. Before applying to students, it would be beneficial for the teachers to play with this kind of application as it gives them an opportunity to get to know their L1 and L2 vowel constellations.

6. References

[1] Flege, J. E. 1988. Using visual information to train foreign language vowel production. *Language Learning*, 38(3), 365-407.

[2] Akahane, R., Yamada, R., Adachi, T., Kawahara, H. 1997. Second language production training using spectrographic representations as feedback. J. Acous. Soc. Japan (E), 18(6), 341-343.

[3] Dowd, A., Smith, J., Wolfe, J. 1998. Learning to pronounce vowel sounds in a foreign language using acoustic measurements of the vocal tract as feedback in real time. *Language and Speech*, 41(1), 1-20.

[4] Carey, M. 2004. CALL visual feedback for pronunciation of vowels: Kay Sona-Match. *CALICO Journal*. 21, 1-32.

[5] Kartushina, N., Hervais-Adelman, A., Frauenfelder, U. H., Golestani, N. 2015. The effect of phonetic production training with visual feedback on the perception and production of foreign speech sounds. *J. Acoust. Soc. Am.* 138 (2), 817-832.

[6] Suemitsu, A., Dang, J., Ito, T., Tiede, M. 2015. A real-time articulatory visual feedback approach with target presentation for second language pronunciation learning. *J. Acoust. Soc. Am.* 138 (4), EL382-387.

[7] SonicPrint™ ver. 1.3.1 (beta-version for French) [Computer software], Arcadia, Inc., Retrieved from http://www/arcadia.co.jp/products/sonicprint

[8] Peterson, G. E., Barney, H. L. (1952). Control methods used in a study of the vowels. *J. Acoust. Soc. Am.* 24 (2), 175-184.

[9] Misono, K. 2007. *Vowel space in English: Regional variants and their perception by Japanese learners of English.* Tokyo: Hokuseidoh Production.

[10] Deterding, D. 1997. The formants of monophthong vowels in standard southern British English pronunciation. *J. IPA.* 27(2), 47-55. DOI: 10.1017/S0025100300005417

[11] Boersma, P. 2001. Praat, a system for doing phonetics by computer. *Glot International* 5:9/10, 341-345.

[12] Thomas, E.R., Kendall, T. 2007. *NORM: The vowel normalization and plotting suite.* Online Resource: http://ncslaap.lib.ncsu.edu/tools/norm/

[13] Lobanov, B. M. 1971. Classification of Russian vowels spoken by different speakers. *J. Acoust. Soc. Am.* 49(2), 606-608.

[14] Adank, P., Smits, R., van Hout, R. 2004. A comparison of vowel normalization methods for language variation research. *J. Acoust. Soc. Am.* 116(5), 3099-3107.

[15] Flynn, N., Foulkes, P. 2011. Comparing vowel formant normalization methods. *Proceedings of ICPhS XVII*. 683-686.