Ruellot, V. (2011). Computer-assisted pronunciation learning of French /u/ and /y/ at the intermediate level. In. J. Levis & K. LeVelle (Eds.). *Proceedings of the 2nd Pronunciation in Second Language Learning and Teaching Conference,* Sept. 2010. (pp. 199-213), Ames, IA: Iowa State University.

COMPUTER-ASSISTED PRONUNCIATION LEARNING OF FRENCH /u/ AND /y/ AT THE INTERMEDIATE LEVEL

Viviane Ruellot, Western Michigan University

This study explores the impact of visual feedback derived from a speech analysis program designed to improve the pronunciation of French /u/ (as in *tout* [tu], "all") and /y/ (as in *tu* [ty], "you") in the speech of intermediate-level adult learners of French with L1 English. As /y/ is absent from the English phoneme repertoire and represents a new phone for these learners, successful distinction between French /u/ and /y/ is contingent upon experience with the language (Flege, 1987). Visual representation of articulators, with a focus on tongue position (back for /u/ and front for /y/), may help learners create distinct phonemic categories for these sounds. Students (*n*=14) in a third-year French phonetics course recorded their pronunciation of French monosyllabic words featuring /u/ and /y/. Participants in the audio-visual condition (*n*=7) received visual information about the formant trajectories of /u/ and /y/ and instruction as to their correlation with degree of mouth aperture (F1), tongue backing (F2), and lip rounding (F3). The recordings were assessed by native French speakers. Results indicate that the presence of visual feedback did not significantly improve pronunciation. The relation between the efficacy of visual pronunciation feedback, time on task, and perception skill development are discussed.

INTRODUCTION

Alternately considered marginal and as important as grammar and vocabulary, the status of pronunciation in the history of the foreign language curriculum has widely fluctuated. In the early 21st century, this skill has regained importance in instruction (Jenkins, 2004). While the efficacy of explicit pronunciation instruction has been tested (with positive results, e.g., Couper, 2003; Derwing, Munro, & Wiebe, 1998; Derwing & Rossiter, 2003), research on pedagogical tools, and more particularly instructional technology, has yielded mixed results. After a review of the relevant research in this area, this paper reports on two experiments testing the efficacy of the visual feedback generated by the speech analysis program WaveSurfer (Sjölander & Beskow, 2000-2005).

BACKGROUND

The current presentation and practice of pronunciation in textbooks and workbooks remains underdeveloped for at least two reasons. First, pronunciation presentation is commonly limited to verbal indications, unaccompanied by the support of visual elements (e.g., representations of articulators) which may help learners put into practice the information given and allow them to encode and retrieve the information through more than one cognitive channel (i.e., aural/oral and visual) (Paivio, 1971; 1991). Second, the opportunities for feedback on the learner's production are generally limited in both paper and electronic versions of instructional materials, and they rarely meet the criteria for effective feedback, which is “comprehensible, [does] not rely solely on the learner's own perception, [allows] verification of response correctness, [pinpoints] specific errors and possibly [suggests] a remedy” (Neri, Cucchiarini, & Strik, 2002, 1210).

With the development of automatic speech recognition technology, commercial pronunciation software programs have begun to address the above-mentioned issues. However, improvements are still needed, particularly regarding the adequacy of the linguistic and pedagogical content of the programs (Derwing & Munro, 2005; Levis, 2007). Still, one feature, the visual representation of speech, which is available in these programs but also in software designed for phonetics and phonology research, has caught the eye of both pedagogues and researchers who have studied the extent of the contribution of spectrograms, waveforms and pitch trackers to the development of pronunciation skills (Levis, 2007; O’Brien, 2006). A pioneer in the field, de Bot (1983) demonstrated the contribution of visual feedback to the improvement of intonation for Dutch learners of English. Hardison (2004) investigated the effectiveness of visual feedback on the learning of French prosody and segments and found significant improvement at both levels, as well as heightened awareness of prosodic and segmental features, improved listening comprehension skills, and increased overall confidence about pronunciation.

Experiments focusing on pitch contours and intonational patterns have more often yielded success and significant pronunciation improvement than have studies concentrating on segments alone (Levis, 2007). Among the oft-cited limitations of the spectrogram as a source of visual feedback on segments is the assumption that the acoustic representation of the learner’s utterance should closely match that of the target model (Neri et al., 2002). This assumption does not seem realistic due to the considerable anatomical variations that exist among speakers and which largely affect the acoustic representation of their utterances. Another limitation of spectrographic displays is related to the complex technical information they provide (Neri et al., 2002). Learners are rarely familiar with such displays and their associated technical terms (e.g., “formants,” “frequency,” “hertz,” etc.), and generally need the help of a trained informant to interpret them. Finally, one of the reasons behind the greater success of experiments focusing on pitch and intonation may be related to the iconic nature of the visual representation in which the “rising, falling, and level lines (…) usually [correspond] to rises, levels, and falls in a speaker’s voice pitch” (Levis, 2007), making the visual feedback a more direct and readily interpretable one.

These issues were addressed in an experiment (Ruellot, 2007) in which beginner American students of French practiced their pronunciation of words featuring the vowels /a/, /i/, /u/ (as in *tout* ‘all’), and /y/ (as in *tu* ‘you’). The subjects in the experimental group studied spectrograms whose degree of iconicity was enhanced through color-coding of formant trajectories correlating with articulators and which excluded technical terms (Guilloteau, 1998). The subjects in the experimental group significantly improved their pronunciation of words featuring /u/ presented in isolation and in sentences. However, none of the subjects significantly improved their pronunciation of items in /y/. Although the presence of visual feedback was met with general enthusiasm, a majority of subjects expressed the need for both additional time and interactive help from an instructor to interpret the visual displays.

The present study addressed these issues and the experimental design in the original study (i.e., Ruellot, 2007) was modified to provide subjects with additional practice time and repeated interaction with an instructor for help on display interpretation, as well as tips on how to improve pronunciation (Engwall, Wik, Beskow, & Granström, 2004). In addition, the sounds at study were limited to French /u/ and /y/, which are typically challenging for English native speakers, as /y/ does not exist in their native repertoire. Principles of Dispersion Theory (Liljencrants & Lindblom, 1972; Lindblom, 1986a; 1986b; 1990; Lindblom & Maddieson, 1988), maximal vowel dispersion (Maddieson, 1984), and predictions from the Speech Learning Model (Flege, 1987) point to the assimilation of /y/ in /u/ during the first stages of acquisition, and the development of /y/ into a distinct phonemic category as experience with the language increases. In order to favor such development, the experimental design in the present study also included a condition in which visual feedback of /u/ and /y/ was presented side-by-side, allowing subjects to concentrate on tongue position (i.e, the main contrast between these two sounds). It was assumed that such focus on a specific articulator would allow subjects to identify their mistakes (Engwall et al., 2004; Pennington & Ellis, 2000) and favor its correction.

The following two hypotheses were tested in two separate experiments in the present study:

Additional practice time and interactive explanations and feedback from an instructor will significantly help subjects receiving visual feedback improve their pronunciation of items featuring /u/ and /y/ (Experiment 1).

Visual feedback of /u/ and /y/ presented side-by-side and studied together will help develop awareness of tongue position, which will result in significantly higher scores than when feedback for /u/ and /y/ is presented and studied separately (Experiment 2).

EXPERIMENT 1: STUDY OF VOWELS IN ISOLATION

Experimental design

Participants

The students, enrolled in a third-year French phonetics course at an American university, were randomly assigned to one of two groups: the Audio feedback condition ($n = 7$), and the Audio-visual feedback condition ($n = 7$).

Instruments

Training for all subjects included a reminder of the articulatory contrasts between French /u/ and English /u/, and between French /u/ and /y/. While the subjects in the Audio condition used basic audio recording software[ix], the subjects in the Audio-visual condition used WaveSurfer (Sjölander, & Beskow, 2000-2005), the same freely available spectrographic analysis program used in the original study. The subjects in the Audio condition were asked to step outside of the classroom for 6 minutes, during which the subjects in the Audio-visual condition were instructed about the spectrographic displays they would study.

These displays were simplified in the same manner as in the original study: technical terms were not addressed and, in order to increase the iconic aspect of the displays, the first three formants of each vowel were color-coded and labeled according to their corresponding articulators.[x] The instructions for the interpretation of the visual displays were also replicated from the original study: subjects were informed that the three lines move vertically according to 1) how open the mouth is, 2) the position (front or back) of the tongue, and 3) the degree to which lips are rounded (CALLIOPE, 1989). The subjects' attention was particularly drawn to the relative distances between the native speaker's lines,[xi] and the subjects' task was to approximate these distances, not the actual 'heights' (i.e., hertz values) of the lines, following recommendations as to how to modify the position of articulators also included in the instructions. In the original study, these instructions and recommendations were provided exclusively in writing. In the present study, they were additionally presented by the instructor orally, and revisited during individual subject-instructor interactions throughout the experiment.

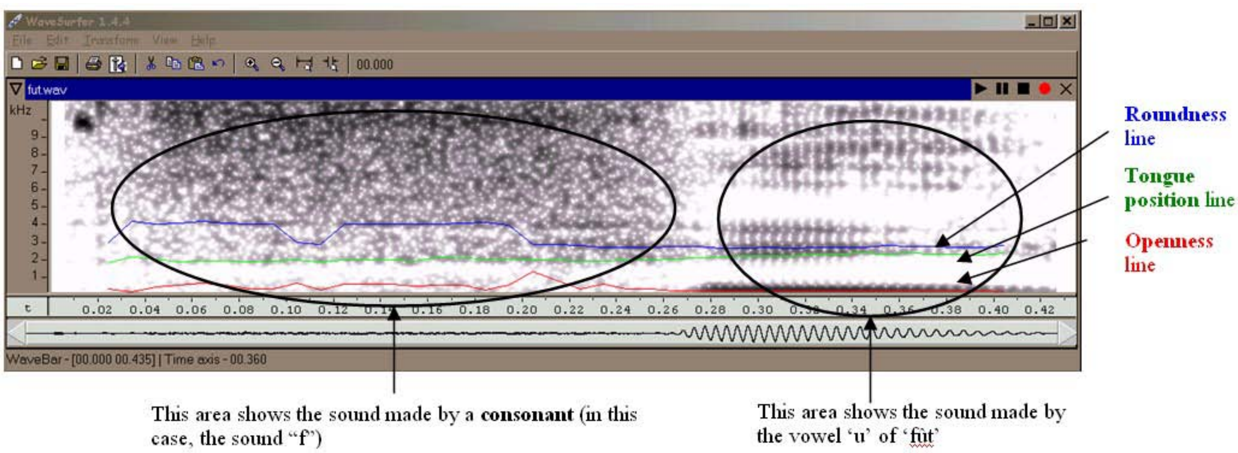

Figure 1. Model Spectrogram of fût 'barrel' uttered by a Female Native Speaker Model (Excerpt from the Instructions provided to the Subjects in the Audio-visual Condition)

Stimuli

All the subjects studied /u/ and /y/ through 10 monosyllabic words, which they recorded, along with sentences featuring them (see Table 1) in order to verify transfer of pronunciation improvement to a contextualized environment.

Table 1. *Stimuli*

Words (rehearsed)[xii]	/u/	*fou* 'crazy', *loup* 'wolf', *mou* 'soft', *nous* 'we', *sous* 'under'
	/y/	*fût* 'barrel', *lu* 'read' (past participle), *mue* 'break' (related to voice), *nus* 'naked', *sue* 'sweat'
Sentences (not rehearsed)	/u/	*Eric est fou de son chien Léon.* 'Eric is crazy about his dog Leon.'
		Le loup est un animal sauvage. 'The wolf is a wild animal.'
		Le canapé d'Anne est mou mais confortable. 'Anne's sofa is soft but comfortable.'
		Ce soir, nous dansons le rock. 'We're dancing to rock dances music.'
		Les chaussures vertes sont sous la table. 'The green shoes are under the table.'
	/y/	*Le vin reste dans un fût pendant cinq ans.* 'Wine is kept in a barrel for five years.'
		Jacques a lu un très bon livre sur les Incas. 'Jacques read a very good book about the Incas.'
		La voix des garçons mue durant l'adolescence. 'Boys' voice breaks during adolescence.'
		Boticelli aime peindre des nus. 'Boticelli likes to paint nudes.'
		Il a chaud et il sue beaucoup. 'He's hot and he sweats a lot.'

Treatment and Tests

Subjects participated in two 50 minute-long pronunciation sessions devoted to the study of /u/ (session 1) and /y/ (session 2). They listened to native speaker model pronunciations of each target word, practiced and listened to their own pronunciation as often as they wished for three minutes, after which they recorded their final version of the target word. Each session ended with the recording of unrehearsed sentences featuring the practiced words.

In the original study, time allocated to the Audio-visual condition for the study and practice of the first word in each target vowel series amounted to 12 minutes. During this time, subjects familiarized themselves with the visual displays and their accompanying written explanations. In the present study, this time was extended to 18 minutes (in addition to the 6 minutes of initial introduction to spectrographic displays mentioned above), during which a trained instructor interacted with subjects, guiding their interpretation of the visual displays and their reflection as to how to modify their articulators to approximate target pronunciations. The instructor continued to provide individual help during the rest of the sessions. The subjects in the Audio condition were also assigned a trained instructor who guided them, throughout the experiment, in their reflection on how to modify their articulators for successful pronunciation. All subjects recorded their pronunciation of the isolated words and sentences three times: before, immediately after and, in order to test for long-term improvement, seven days after treatment.

Rating Procedure

The 840 recordings obtained were submitted for assessment to three French native speakers who rated two types of recordings: 1) words pronounced in isolation and 2) the same words embedded in sentences. The latter were extracted from their sentence context so as to minimize the impact that the non-target words and sounds may have had on the ratings. The raters used a 5-point Likert-type scale (Bongaerts, Planken, & Schils, 1995) ranging from 1 – "very strong foreign accent; the comprehension of the word is very difficult; this speaker is undeniably not a native speaker of French", to 5 – "no trace of a foreign accent; the comprehension of the word is very easy; this speaker is undeniably a native speaker of French."

RESULTS

The first hypothesis posits that more time on task and interactive explanations with – and feedback from – the instructor will help learners who receive visual feedback improve their pronunciation of /u/ and /y/.

- Short-term improvement

Although all of the subjects in the experiment improved their pronunciation of words in /u/ and /y/, as well as target words in /y/ appearing in sentences (i.e., extracted words) immediately after treatment (Table 2), improvement in the Audio-visual condition was not significantly different than in the Audio condition. This can be concluded from the results of unpaired t-tests run on mean ratings obtained immediately after treatment for words and extracted words in /u/ and /y/ with condition as a between-subjects factor displayed in Table 3.

Table 2. *Mean Ratings and Mean Difference in Mean Ratings from Pre-test to Post-test for Isolated Words and Extracted Words (both Conditions combined)*

Vowel	Recording	Post-test mean	Pre-test mean	Mean difference	*t*	*df*	*p*
u	Words	3.524	2.941	0.583	3.477	13	0.004**
y	Words	3.581	3.357	0.224	1.849	13	0.001**
y	X-words	3.783	3.010	0.773	4.181	13	0.019**
u	X-words	3.636	3.092	0.543	2.672	13	0.087

**p <.01; 'X-words' refers to extracted words

Table 3. *Mean Ratings and Mean Difference in Mean Ratings at the Post-test level for Isolated Words and Extracted Words in the Audio and the Audio-visual Conditions*

Vowel	Recording	Condition	Mean	Mean difference	*t*	*df*	*p*
/u/	Words	Audio-visual	3.572	0.096	0.307	12	0.764
		Audio	3.476				
/y/	Words	Audio-visual	3.661	-0.244	-0.489	12	0.634
		Audio	3.905				
/u/	X-words	Audio-visual	3.629	0.095	0.301	12	0.769
		Audio	3.533				
/y/	X-words	Audio-visual	3.700	0.130	0.225	12	0.826
		Audio	3.571				

'X-words' refers to extracted words

- Long-term pronunciation improvement

As can be gathered from Table 4, long-term improvement for all of the subjects in the experiment was restricted to words in /y/. That is, the pronunciation improvement for words in /u/ and extracted words in /y/ obtained immediately after treatment was not maintained one week later.

Table 4. *Mean Ratings and Mean Difference in Mean Ratings from Pre-test to Delayed Post-test for Isolated Words and Extracted Words*

Vowel	Recording	Delayed Post-test mean	Pre-test mean	Mean difference	*t*	*df*	*p*
u	Words	3.009	2.941	0.068	0.519	13	0.613
y	Words	3.685	3.010	0.675	3.756	13	0.002**
y	X-words	3.367	3.092	0.275	1.505	13	0.156
u	X-words	3.228	3.357	-0.128	-0.860	13	0.405

**p <.01; 'X-words' refers to extracted words

However, the presence of visual feedback did not significantly contribute to long-term improvement of words in /y/, as indicated by the absence in Table 5 of a significant difference between conditions.

Table 5. *Mean Ratings and Mean Difference in Mean Ratings at the Delayed Post-test level for Isolated Words and Extracted Words in the Audio and the Audio-visual Conditions*

Vowel	Recording	Condition	Mean	Mean difference	*t*	*df*	*p*
/u/	Words	Audio-visual	2.960143	-0.097	-0.220	12	0.830
		Audio	3.057143				
/y/	Words	Audio-visual	3.645500	-0.080	-0.149	12	0.884
		Audio	3.725429				
/u/	X-words	Audio-visual	3.133143	-0.191	-0.653	12	0.526
		Audio	3.323714				
/y/	X-words	Audio-visual	3.390571	0.047	0.092	12	0.928
		Audio	3.343143				

'X-words' refers to extracted words

EXPERIMENT 2: STUDY OF VOWEL PRESENTED IN PAIRS

Experimental design

Experiment 2 started one week after Experiment 1 and lasted for one session. The participants were the same students enrolled in the French phonetics course who had served as subjects in Experiment 1. For the second experiment, they switched groups, and the subjects from the Audio-visual condition in Experiment 1 were now assigned to the Audio condition and exclusively worked with audio feedback, while the subjects from the Audio condition in Experiment 1 now made up the Audio-visual condition and additionally studied visual displays of /u/ and /y/ presented side-by-side. Initial training, instruments, stimuli, treatments, tests, and rating procedure were almost the same as in Experiment 1. The only difference is that Experiment 2 took place over only one session, during which /u/ and /y/ were not studied separately but simultaneously.

RESULTS

In order to test hypothesis 2 and compare the efficacy of the visual feedback on pronunciation of specific sounds when they are presented, studied and practiced separately vs. side-by-side, two series of unpaired t-tests were run. Each contrasted mean ratings from the Audio-visual condition in Experiment 1, i.e., when/u/ and /y/ were studied separately, with mean ratings by the subjects in the Audio-visual condition in Experiment 2, when /u/ and /y/ were presented side-by-side. The first series of unpaired t-tests contrasts performance in the short-term, i.e. immediately after treatment, while the second one compares long-term performance, i.e., one week after treatment.

Although the subjects who studied the visual display of /u/ and /y/ presented side-by-side significantly improved their pronunciation of words in /u/ immediately after treatment (paired $t(6) = 7.316$, $p = 0.0003$), their scores were not significantly higher than those of the subjects who studied the vowels separately, be it immediately after treatment (see Table 6) or one week later (see Table 7). In other words, studying the articulatory contrasts between /u/ and /y/ through visual displays presented side-by-side did not constitute an advantage over studying them separately.

Table 6. *Mean Ratings and Mean Difference in Mean Ratings at the Post-test Level for Isolated Words and Extracted Words in the Audio-visual Conditions in Experiment 1 (Audio-visual 1) and Experiment 2 (Audio-visual 2)*

Vowel	Recording	Condition	Mean	Mean difference	*t*	*df*	*p*
/u/	Words	Audio-visual 1	3.572	0.295	1.231	12	0.242
		Audio-visual 2	3.277				
/y/	Words	Audio-visual 1	3.661	-0.116	-0.287	12	0.779
		Audio-visual 2	3.777				
/u/	X-words	Audio-visual 1	3.629	-0.001	-0.002	12	0.998
		Audio-visual 2	3.629				
/y/	X-words	Audio-visual 1	3.700	0.472	0.877	12	0.397
		Audio-visual 2	3.229				

‘X-words’ refers to extracted words

Table 7. *Mean Ratings and Mean Difference in Mean Ratings at the Delayed Post-test Level for Isolated Words and Extracted Words in the Audio-visual Conditions in Experiment 1 (Audio-visual 1) and Experiment 2 (Audio-visual 2)*

Vowel	Recording	Condition	Mean	Mean difference	*t*	*df*	*p*
/u/	Words	Audio-visual 1	2.960	-0.602	-1.589	12	.138
		Audio-visual 2	3.562				
/y/	Words	Audio-visual 1	3.646	0.007	.012	12	.991
		Audio-visual 2	3.639				
/u/	X-words	Audio-visual 1	3.133	-0.419	-1.472	12	.167
		Audio-visual 2	3.552				
/y/	X-words	Audio-visual 1	3.391	-0.286	-.569	12	.580
		Audio-visual 2	3.677				

'X-words' refers to extracted words

DISCUSSION AND CONCLUSION

The results of the present study suggest that simplified visual feedback generated from spectrograms does not significantly contribute to pronunciation improvement of /u/ and /y/, regardless of whether they are presented and studied individually or side-by-side. The comments from the feedback questionnaire that subjects completed at the end of the experiment may shed some light on these results.

A large majority of subjects welcomed the use of computer-generated visual feedback while studying pronunciation (the sessions were rated 5.86 on a scale from 1 to 7). They also highly appreciated the oral component absent in the original study: oral presentations of the sounds' articulatory characteristics, oral explanations for the visual displays, and subject-instructor repeated interactions. Indeed, they found it considerably more helpful than the accompanying written presentations and explanations, which they reported using only as reference. Subjects also highlighted the benefits of the visual displays which provided immediate feedback on their production[xiii] and heightened their awareness of articulator position,[xiv] allowing them to compensate for perception difficulties.[xv] The subjects in the second experiment also found it very helpful to study the visual representation of /u/ and /y/ side-by-side as it enabled them to pinpoint the exact problem on which to focus.[xvi]

While all the subjects receiving visual feedback soon found it comprehensible, a minority of them expressed difficulty understanding and interpreting the visual displays when these were first introduced. Although the time allocated for training on the visual displays was double that in the original study,[xvii] it appears that this type of visual feedback, even when simplified, requires much longer initial presentation for subjects to feel comfortably familiar (O'Brien, 2006). It also seems that three minutes of practice on each target word may not be sufficient for subjects to be able to fully benefit from both the information provided by the visual displays and the interactive feedback with the instructor. Indeed, 9 of the 14 subjects who received visual feedback would have liked to have had more time to practice their pronunciation of each target word after interpretation of the visual display.[xviii] The importance of practice time has been stressed, with the recommendation to allow students to complete computer-based pronunciation practice beyond the classroom (Wang & Munro, 2004).

The benefits reported by the subjects seem to indicate that the experimental treatment in the present study partially meets the requirements of effective feedback: it does "not rely solely on the learner's own perception," it allows "verification of response correctness," and it pinpoints "specific errors" to the learner. The user's interpretation of the display, as well as interactive feedback from the instructor, provide opportunities to "suggest a remedy" (Neri et al., 2002). Additional time for familiarity with the visual displays, as well as for practice of sounds, may enhance the benefits of spectrographic visual feedback, which has often been considered hardly effective for pronunciation improvement at the segmental level (Neri et al., 2002). Moreover, several subjects pointed to the contribution of the visual displays to the development of perception skills. Considering that training on sound discrimination alone has had a positive impact on production (Bradlow, Pisoni, Akahane-Yamada, & Tohkura, 1997; Coniam, 2002; Levis, 2007), and that phonemic distinction between /u/ and /y/, both at the perception and the production level, is contingent on time and experience (Flege, 1987), the effect of these visual displays at the perception level in the long-term would also merit close examination before this type of feedback is definitively considered ineffective. Future research comparing the effect of visual feedback on the pronunciation of learners of different age categories, L2 proficiencies and/or learning styles is also warranted.

ABOUT THE AUTHOR

Viviane Ruellot began teaching French and applied linguistics at Western Michigan University in 2005. She serves as coordinator of beginning French courses and as liaison between WMU and Kalamazoo high schools. Her research focuses on the acquisition of pronunciation, more particularly that of French as a foreign/second language (L2). She looks for ways in which visual feedback may help learners bridge the gap between perception and production and improve their pronunciation. One of her goals is to find ways to simplify that information to make it available to learners not trained in speech analysis. She is also interested in the pedagogy of pronunciation teaching, as well as the stages of development in the acquisition of L2 French pronunciation.

[ix] Microsoft ® Sound Recorder, version 5.1.
[x] I.e., F1 (red): “the openness line”, F2 (green): “the tongue position line”, and F3 (blue): “the lip roundness line”.
[xi] A short distance between the red (F1) and the green (F2) lines signifies that the tongue is pushed towards the back of the mouth, which is characteristic of /u/. Contrastingly, a wide distance between the same lines indicates the tongue is massed in the front of the mouth, which is necessary for the pronunciation of /y/.
[xii] Subjects were instructed not to pronounce the final consonant letter in *loup*, *nous*, *sous*, *fût*, and *nus*.
[xiii] “I feel that due to the immediate feedback received by the programs and being able to see what my pronunciation was, I was able to correct the sounds easily.”
[xiv] “I was able to see [the] placement [of my articulators] and compare it to my own. I was able to see my mistakes and hone in on them.” The visual displays presented side-by-side contributed to raising awareness of articulators and their position for a majority of subjects: “Before this class, I had never noticed a difference between those (…) sounds, and listening to and hearing oral explanations and examples helped, but WaveSurfer helped me to put my articulators where they needed to be.”
[xv] “I have a hard time hearing the difference in my pronunciations so I used the 3-line model to see if I was doing it correctly;” “Especially with the /u-y/ practice I felt like I could really hear and feel the difference,” and “I can now distinctly tell the difference between the sounds. Before I couldn't very well.”
[xvi] “I was able to tell what exactly I needed to fix, like having my tongue closer to my teeth.”
[xvii] In the original study, training on the visual displays amounted to 12 minutes. In the present study, it lasted for 24 minutes (i.e., 6 minutes of initial introduction in addition to the 18 minutes devoted to the study of the first target word).
[xviii] “These sessions were very helpful, but overall the class felt rushed because we crammed in the worksheets so that we could get enough pronunciation time.”

REFERENCES

Bongaerts, T., Planken, B., & Schils, E. (1995). Can late learners attain a native accent in a foreign language? A test of the Critical Period Hypothesis. In D. Singleton & Z. Lengyel (Eds.), *The age factor in second language acquisition: a critical look at the Critical Period Hypothesis* (pp. 30-50). Clevedon: Multilingual Matters.

Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., & Tohkura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *Journal of the Acoustical Society of America, 101*, 2299–2310.

CALLIOPE (1989). *La parole et son traitement automatique*. J.P. Tubach, (ed.). Paris: Masson.

Coniam, D. (2002). Technology as an awareness raising tool for sensitising teachers to features of stress and rhythm in English. *Language Awareness*, *11*(1), 30–42.

Couper, G. (2003). The value of an explicit pronunciation syllabus in ESOL teaching. *Prospect*, *18*(3), 53–70.

de Bot, K. (1983). Visual feedback of intonation I: Effectiveness and induced practice behavior. *Language and Speech*, *26*(4), 331-350.

Derwing, T. M. & Munro, M. J. (2005). Second Language Accent and Pronunciation Teaching: A Research-Based Approach. *TESOL Quarterly 29* (3), 379-397.

Derwing, T. M., Munro, M. J., & Wiebe, G. (1998). Evidence in favor of a broad framework for pronunciation instruction. *Language Learning*, *48*, 393-410.

Derwing, T. M. & Rossiter, M. J. (2003). The effects of pronunciation instruction on the accuracy, fluency, and complexity of L2 accented speech. *Applied Language Learning, 13* (1), 1-17.

Engwall, O., Wik P., Beskow, J., & Granström, B. (2004). Design strategies for a virtual language tutor. *ICSLP 2004* (3), 1693-1696. Retrieved August 24, 2010, from http://www.speech.kth.se/ville/publications/icslp2004_tutor.pdf

Flege, J. E. (1987). The production of "new" and "similar" phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics*, *15*, 47-65.

Guilloteau, N. C. (1998). Modification of phonetic categories in French as a second language: Experimental studies with conventional and computer-based intervention methods. *Dissertation Abstracts International, A: The Humanities and Social Sciences*, *58* (7, Jan, 2621-A).

Hardison, D. (2004). Generalization of computer-assisted prosody training: Quantitative and qualitative findings. *Language Learning & Technology, 8*(1), 34-52.

Jenkins, J. (2004). Research in teaching pronunciation and intonation. *Annual Review of Applied Linguistics*, *24*, 109-125.

Levis, J. (2007). Computer technology in teaching and researching pronunciation. *Annual Review of Applied Linguistics, 27*, 184–202.

Liljencrants, J., & Lindblom, B. (1972). Numerical simulations of vowel quality systems: the role of perceptual contrast. *Language, 48*, 839-862.

Lindblom, B. (1986a). Phonetic universals in vowel systems. In J. J. Ohala & J. J. Jaeger (Eds.), *Experimental Phonology* (pp. 13-44). New York: Academic Press.

Lindblom, B. (1986b). On the origin and purpose of discreteness and invariance in sound patterns. In J. Perkell & D. Klatt (Eds.), *Invariance and variability of speech processes* (pp. 493-510). New Jersey: Lawrence Erlbaum.

Lindblom, B. (1990). On the notion of possible speech sound. *Journal of Phonetics, 18,* 135-152.

Lindblom, B., & Maddieson, I. (1988). Phonetic universals in consonant systems. In L. M. Hyman & C. N. Li (Eds.), *Language, speech and mind* (pp. 62-78). London: Routledge.

Maddieson, I. (1984). *Patterns of sounds*. Cambridge: Cambridge University Press.

Neri, A., Cucchiarini, C., & Strik, H. (2002). Feedback in computer assisted pronunciation training: Technology push or demand pull? *ICSLP-2002*, Denver, USA, pp. 1209-1212.

O'Brien, M. G. (2006). Teaching pronunciation and intonation with computer technology. In L. Ducate & N. Arnold (Eds.), *Calling on CALL: From theory and research to new directions in foreign language teaching* (CALICO Monograph Series, Vol. 5, pp. 127–148). San Marcos, TX: CALICO.

Paivio, A. (1971). *Imagery and verbal processes*. New York: Holt, Rinehart and Winston, Inc.

Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology, 45*, 255-287.

Pennington, M. C., & Ellis, N. C. (2000). Cantonese speakers' memory for English sentences with prosodic clues. *The Modern Language Journal*, *84* (3), 372–389.

Ruellot, V. (2007). *French pronunciation learning and computer-mediated visual feedback*. Unpublished PhD dissertation, University of Illiois at Urbana-Champaign. DAI-A 67/07, p. 2556, Jan 2007, 3223704.

Sjölander, K., & Beskow, J. (2000-2005). WaveSurfer (Version 1.8.5) [Computer software]. KTH, Stockholm. Available: http://www.speech.kth.se/wavesurfer/download.html

Wang, X. & Munro, M. J. (2004). Computer-based training for learning English vowel contrasts. *System, 32*, 539–552.