IMPLICATIONS OF ULTRASOUND TECHNOLOGY IN THE L2 CLASSROOM

Bryan Meadows
University of Arizona

The purpose of this article is to report ongoing exploratory research into the potential implications of ultrasound technology to L2 pronunciation instruction. An initial exploratory study, conducted in the spring of 2006, constructed a JFL (Japanese as a Foreign Language) classroom which featured both ultrasound-based and conventional instruction. The target articulations of this classroom were (1) long vowels, (2) nasals, and (3) flaps. The study yielded both qualitative and quantitative results. Qualitative results indicated that students responded extremely well to ultrasound-based instruction (teacher modeling and student interaction with the ultrasound) while they did not respond well to practice sentences and ultrasound model videos. Quantitative results, although not significant, did document general improvement in L2 pronunciation for all subjects over the course of the study. Results for individual target articulations were mixed. A follow-up study will adopt an experimental approach in order to begin determining if in fact a causal relationship exists between ultrasound technology and improved L2 pronunciation instruction. In addition, the follow-up study will address the shortcomings of the initial case study.

INTRODUCTION

Since 2004, the APIL (Arizona Phonological Imaging Lab)[1] located at the University of Arizona has been exploring the possibility of applying ultrasound technology to L2 instruction. This article is only a snapshot of what research has been done (a classroom case study) and what is being planned for in the immediate future (a classroom comparison experiment).This article will first discuss some of the background to this line of research along with some relevant studies appearing in the literature. Next, it will report on a case study conducted during the spring semester of 2006. To close, plans for a follow-up study will be introduced. This follow-up study will extend the basic findings of this initial study and address its shortcomings.

LITERATURE REVIEW

Recent Studies

Pronunciation training faces many difficulties in the L2 classroom. For one, it is difficult for L2 learners to detect phonemic distinctions that do not regularly occur in their native language. This phenomenon, regularly referred to as 'phonological filtering' (Trubetzkoy, 1969 in Lambacher, 1999), means that even the best-designed L2 pronunciation lesson plan will fail if it relies on audio input alone, as the L2 learner cannot perceive certain L2 differences. For this reason, classroom researchers experiment with

instructional technologies that translate auditory phonemic differences into visual ones so as to circumvent the effects of phonological filtering. This sort of technologically-guided exploration is popularly known as Computer-Aided Language Learning (CALL). One central hypothesis driving this area of research is the Noticing Hypothesis (Schmidt, 1990 in Neri et al., 2002, p.447), the prediction that "awareness of an L2 feature is necessary to acquire that feature." Thus, making explicit that which language learners often overlook (or fail to perceive) is a common component of CALL-informed L2 phonetic instruction. The current study situates itself in the strain of CALL research which is primarily concerned with classroom instruction of L2 pronunciation and will draw on the recent literature.

One recent series of experiments that tested the contribution of visualized speech to L2 speech perceptional development is chronicled in Hazan, Sennema, Iba, & Faulkner (2005). The experimenters provided training to Japanese L1 speakers who were studying English. Subjects were divided into three training conditions: auditory only 'A', audiovisual 'AV', and visual only 'V'. Performance on the post-treatment perception tasks showed that subjects receiving AV training outperformed the other two conditions when asked to distinguish between English /b/ and /v/. On the other hand, when presented with the infamous /l/ and /r/ phonemic distinction, subjects receiving A training (auditory only) showed the highest rate of accuracy. The authors speculate that AV training does not benefit the /l/ and /r/ distinction because the tongue articulations that create the phonemic distinction are not visible from the outside of the mouth.

Hirata (2004) expanded the role of visuals in her study, and tested a computer program which provided visual representations of L2 learner production in addition to visual representations of target L2 pronunciation. Hirata's subjects, eight L1 English speakers currently studying Japanese, were divided into a treatment and control group. The treatment group received training which displayed pitch contours of target Japanese words in visual format. Subjects then attempted to reproduce the target pitch pattern, and their production was interpreted by the computer program and displayed in visual format on a split-screen. This allowed students to compare their production with the target model. Side-by-side comparison of the two visuals was the only form of feedback made available to the students. Post-test results show that audiovisual training did significantly improve L2 production over the control condition. Furthermore, Hirata reports signs that improvement in L2 production was then transferred to improvements in L2 perception, but the small subject population prevents any generalizable claims. Hirata (2004) suggests that not only L2 input, but L2 feedback in visual form contributes to improved L2 speech performance.

A third report, Lambacher (1999), outlines an instructional design similar to Hirata (2004), except subjects in the experimental condition are presented with input and feedback in electronic-visual format. This system, known as *Electronic Visual Feedback* (EVF), projects on the computer screen spectogram representations that indicate sound duration and frequency range.

Like Hirata (2004), students try to mimic the target model pronunciation and then compare their performance against a visual of the original target model. The drawback to such graphs is that they are nearly indecipherable without some previous training in phonetics. To overcome this challenge, Lambacher included instructors in the lesson plan to help interpret and evaluate the spectograms for the L2 students. Building on what is offered in Hirata (2004), students in this arrangement receive feedback from three sources: the spectogram report, their initial interpretation of that report, and their teacher's informed analysis. Different from the previous two, Lambacher (1999) is not a report of an experimental study, but rather a blueprint of how EVF technology could be incorporated into an L2 classroom setting.

Advantages of Ultrasound Technology

An L2 classroom equipped with ultrasound technology can address the individual shortcomings of the studies presented above. For example, Hazan et al. (2005) report that visual input seemed to help in circumstances where phonemic difference was instigated at an external point on the mouth (e.g. the lips), but not in cases where phonemic differences were made inside of the mouth. This is an obvious limitation of visuals that originate with video cameras since the cameras sit outside of the mouth. In contrast, ultrasound technology projects soundwaves through the skin and into the mouth, making the internal areas (i.e. tongue muscle and segments of the palate) visible on a monitor screen. This type of technology bypasses a difficulty common to all L2 classrooms and experienced by all L2 instructors—how to show the students what is happening inside the mouth while the tongue is in motion.

Hirata's (2004) subjects were limited in that they did not receive any expert feedback on their production. They could only visually compare the relative similarities of their production with the target model. This can be a major concern for classroom instructors because students who do not receive informed feedback from an external source can easily begin reinforcing incorrect L2 features resulting in the dreaded 'fossilization' (Shehadeh, 2003, p.167). In fact, this was observed in one of the beginning L2 ultrasound experiments at APIL. In the pilot study, one L1 Korean speaker was successfully trained to make a phonemic distinction in English (*beet* vs. *bet*) which does not exist in his L1. Unfortunately, however, the subject reversed the distinction in his head, confusing written versions of 'beet' for 'bet' and vice versa. The most reasonable conclusion was that the subject's confusion must have been a result of no teacher feedback during training.

Lambacher's (1999) proposal does address the issue of feedback and makes provisions for teacher-student interaction during audio-visual training. However, Lambacher's (1999) article is only a suggestion of how an EVF classroom could look and does not describe the actualization of such a classroom design. The current study utilizing ultrasound technology fills that gap. Feedback is an integral instructional component of the current case study.

INITIAL CASE STUDY

Study Context and Purpose

The next section will document a first attempt at creating an ultrasound L2 classroom. This initial attempt was conducted as a case study during the spring semester of 2006 at the University of Arizona. From the beginning, the study was designed with two goals in mind. First, this attempt would serve as a pilot for subsequent studies. Based on the results, the investigative team could modify later studies in response to any problems in regards to protocol, procedure, and data collection encountered during this initial study. Second, this first study was intended to establish some baseline data for continued research. Specifically, the investigative team wanted to know how much of an effect was measurable after providing students with both conventional and ultrasound-based instruction. Once it was determined what was possible when students were provided with all tools at their disposal, subsequent studies could then begin removing them one by one in order to measure their respective effects. Although this first study did not provide significant results as were hoped for, the results were extremely valuable in setting the course for continued research.

This initial study addressed several key questions. The first question was whether or not ultrasound technology can contribute to development of L2 target articulations. Second, this study addressed the issue of what kind of instructional protocol seems to be most effective when incorporating ultrasound into lesson delivery. Third, the study considered student reactions to the ultrasound technology in the classroom. In other words, to which aspects of the ultrasound technology do they respond favorably?

The study is best characterized as a mixed-design due to its inclusion of both qualitative and quantitative data. For qualitative data collection, the Principal Investigator (PI) observed the subjects during the lessons both as an uninvolved observer and later as a co-instructor. For quantitative data collection, subjects' L2 production was recorded both prior to (pre-condition) and immediately following the instructional sessions (post-condition). These recordings were judged by native Japanese speakers for acceptability, and their judgments were quantified for comparison between the two recording times (pre and post conditions).

Key Questions of Initial Case Study:

1) What contributions can ultrasound technology make to L2 pronunciation development?

2) What kind of protocol seems to work best when using ultrasound technology as part of lesson delivery?

3) What kind of reactions do students give to ultrasound-based teaching?

Participants

Three university students who are currently studying Japanese as a foreign language participated in the study. The sample size was extremely

small, but perhaps reasonable considering the pilot-study nature of the project. Nevertheless--to be clear--the results of this study cannot be generalized to other classroom situations due to the small sample size and preliminary nature of the study. Their regular Japanese class (JPN 101) was their first formal introduction to the study of Japanese. Although new to Japanese, all students reported some familiarity with foreign language study, having attended Spanish classes in high school. The participating Japanese instructor is a native Japanese speaker. She has a Masters degree in TESOL (Teaching English as a Second or Other Language) and has been teaching Japanese at the university level for 3+ years. The PI is a graduate student at the University of Arizona and holds a Masters degree in second language education. Prior to this study, he had spent one year working as a research assistant at APIL, so he brought to the study a strong familiarity with the ultrasound technology and its imaging capabilities. During ultrasound lessons, the PI contributed to instruction by helping interpret the ultrasound images for the students.

Instructional Session Procedure

Subjects met with the instructor and PI for one-hour instructional sessions on three occasions outside of the students' regular Japanese class time. Originally, eight weekly instructional sessions were planned for, but due to scheduling conflicts the class was only able to meet a total of four times. Each instructional session consisted of one or two lessons that focused on a particular articulation that can be challenging for L1 English speakers learning Japanese (see *Table 1*).

The target articulations featured in the individual lessons were the following: (1) long vowels [o:] and [e:] because American English speakers will tend to diphthongize these vowels when held for a long duration, (2) nasals, in particular the uvular nasal [ɴ] as it does not normally occur at the end of words in English as it does in Japanese, (3) flaps [ɺ], as the word-initial and palatalized versions [ɺ ʲ] can be very challenging for American English speakers. Equipment used during the lessons included the following: whiteboard, instructional handouts, ultrasound machine, computer laptop (with video of target model articulations), and laptop projector (for enlarging the ultrasound and model video images).

Table 1. Target Articulations of each Instructional Session

Session One	Session Two	Session Three	Session Four
-- Introduction to Ultrasound and Phonetics -- Long vowel [e:]	-- Long vowel [o:] -- Nasals	-- Nasals (continued) -- Flaps / Palatalized Flaps *Note: Only Subject 2 in attendance.*	-- Nasals (continued) -- Flaps / Palatalized Flaps *Note: Only Subject 1 and Subject 3 in attendance.*

Lesson Half-One: Conventional Instructional Approaches

Each individual lesson was divided into two parts. The first part (PI-constructed handouts and teacher call-response techniques) emulated conventional approaches to L2 phonetics training, while the second half utilized ultrasound technology. To begin the conventional half of each lesson, the subjects were presented instructional handouts which highlighted each target articulation (see Appendix A). During the first two lessons, subjects were simply presented with the handouts and given time to read them over. However, over time the PI rapidly reduced the role of the handouts as the students commented in their surveys that the handouts were too 'dense' and not helpful. By the final instructional session, the PI was highlighting specific areas of interest from the handout and accenting the information with informal visuals on the classroom whiteboard.

On the other hand, L2 input from the Japanese teacher remained a constant feature of all lessons. This initial input from the teacher followed conventional teaching methods. The teacher read through the target words from the whiteboard, pronouncing each one in a clear and audible voice while pointing to the words on the whiteboard. Immediately following the L2 input, students were prompted to provide some output. Again, this followed conventional techniques. The teacher began by having the students read in unison the target words two-to-three times. Following this, each student was prompted to read each target word outloud individually (i.e. round-robin style), and the teacher provided immediate feedback to the students following each production.

One troubling area of this half of the lesson design concerned target sentences. Originally, the subjects were to practice reading sentences which featured the target articulations in them. However, with the first lesson, it became very apparent that the sentences were hindering progress rather than contributing to it. By the second lesson, we had abandoned the practice sentences all together.

Lesson Half-Two: Ultrasound Approaches

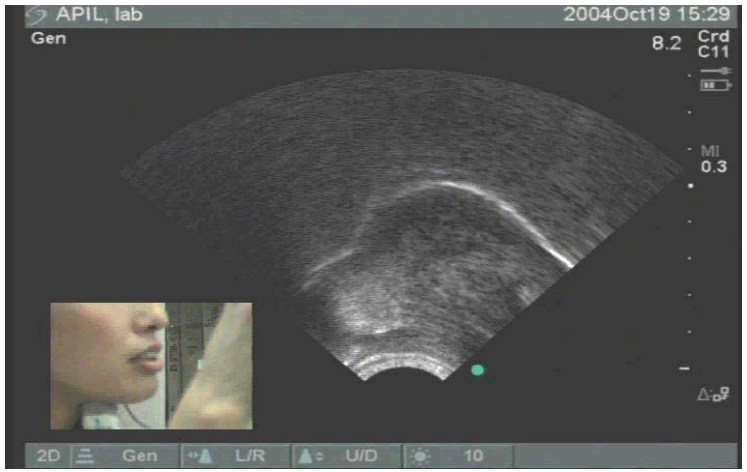
Following this traditional-style exchange of student output and teacher feedback, the lesson transitioned to ultrasound-based instruction. The first element of ultrasound-based instruction was the teacher modeling of target tongue articulations via the ultrasound. To do this, the teacher pressed the ultrasound transducer to the underpart of her chin which then projected a side-view image of the internals of her mouth (and was in turn projected to the classroom wall by way of a video projector). Most prominent in the ultrasound image is the upper portion of the tongue muscle, but segments of the palate are also visible (See *Figure 1*).

Over time, the PI began inserting commentary to the teacher's real-time ultrasound model projections by standing next to the projection screen and pointing out for students particular points of interest in the ultrasound image. At times, the ultrasound moves fast and it is sometimes difficult to know exactly what to watch for on the screen. Because the PI has over a year's

worth of experience viewing ultrasound imagery, it was theorized that his input could help direct subjects to the most salient points of the ultrasound image.

A further source of ultrasound L2 input came in the form of model videos prerecorded by the instructor and made available to the students during each lesson. Originally, the videos were loaded onto a laptop computer and set to the side, available to the students to interact with while waiting their turn with the teacher and ultrasound. During the final lesson, the PI brought the model videos to the students' direct attention by incorporating the model videos into the lesson as L2 input.

Figure 1. Screen Capture of Ultrasound Imaging of Japanese L1 Speaker



The final element of each lesson was a chance for students to sit down with the instructor and the ultrasound to receive individualized feedback on their L2 production. This triadic interaction is the integral part of the envisioned ultrasound classroom. Just as the instructor had done immediately prior, each subject placed the ultrasound transducer to the underpart of their chin, thus projecting an internal image of their mouth onto the projector screen. Subjects then practiced the target articulations with the instructor's guidance while paying close attention to their own articulations made visible on the screen. This arrangement provided subjects with two sources of feedback: one from the ultrasound as they were now able to see exactly what their tongue was doing, and two, from the instructor who was able to provide very specific feedback as she too could see what the subject was doing with their tongue. With time, the PI began providing a third source of feedback to subjects. Standing next to the projector screen, the PI was able to provide specific instructions to subjects in order to help them better mimic the target articulations for each lesson. In this manner, the PI and instructor played complementary roles. The instructor provided specific feedback on the

acoustic quality of a subject's L2 production, and the PI provided feedback on the tongue articulations of the subjects.

QUALITATIVE RESULTS

Qualitative data were derived from two sources. The PI observed and kept careful field notes of each ultrasound instructional session. Additionally, subjects filled out opinion surveys at the close of each session. The surveys asked open-ended questions such as, "What went really well for you today?" and "What part of today's instruction was not very helpful? Why?" Both observations and student surveys gave us a clear idea of what was effective teaching protocol and what was not.

Sentences Featuring Target Articulations

PI observation and student surveys quickly verified that the sentences prepared for the students were not working. In one of his survey responses, Subject One (S1) wrote, "Reading through the sentences was the least helpful today," and Subject Three (S3) wrote, "Reviewing how to make sounds that were not focused on today (sentences) was the least helpful." One most likely cause for their dissatisfaction was the fluency issue. Since these were beginning students of Japanese, reading long sentences—even if they feature the target articulations—can actually be counter-productive as the students soon get lost in the cumbersome task of decoding. The PI observed the teacher compensating by skillfully dividing the sentences into smaller chunks and then gradually increasing the length of each target sentence, but it was to no avail. The subjects did not respond well to the sentence-reading exercise and, sensing complete shut-down, the teacher suggested the lessons skip ahead. We ceased using the sentences altogether by the second instructional session.

Instructor Ultrasound Modeling

While the sentences were less than successful, PI observations and student surveys suggested that teacher modeling of target articulations on the ultrasound machine was well-received by the students. This approach seemed especially effective in demonstrating to the students stark contrasts between target articulations. Featured in the student surveys are three success stories. For example, following the lesson on Japanese long vowels, Subject One (S1) and Subject Two (S2) both remarked on their survey that the ultrasound helped to clearly see the difference between a vowel that is simply held for an extended duration and one that is allowed to turn into a diphthong. Additionally, following the lesson on uvular nasals [ɴ], S1 and S2 again remarked that seeing the teacher's tongue articulations via the ultrasound helped to see the difference between an alveolar nasal [n] (what they are already used to) and the uvular nasal [ɴ] (for them, a novel articulation).

That lesson, which also addressed the articulatory difference between the velar nasal [ŋ] and [ɴ] uvular nasal, elicited further interesting responses from S2, what could be called a second success story. In his post-lesson survey

S2 wrote, "The most helpful thing today was seeing the location difference between the [ŋ] and the [ɴ] because I can't hear the difference between the two." Such a statement speaks directly to the reasoning behind incorporating ultrasound visuals into an L2 lesson. It is precisely because of phonological filtering that S2 is having a difficult time detecting the difference in sound quality between the velar and uvular nasal. If the L2 input that he was receiving was limited to auditory information, then he may have left the classroom frustrated.

A final success story that resulted from the ultrasound modeling came during the final lesson on Japanese flaps. During the modeling stage of the lesson, the teacher demonstrated the articulatory differences between the regular flap [ɺ] and the palatalized version [ɺʲ]. The regular flap only involves a small portion of the tongue (tongue tip), while the palatalized flap involves a broader tongue surface area. Normally, this would be nearly impossible for a student to see since such a small difference is not detectable if one is looking into someone's mouth from the outside. However, this small detail is in fact magnified when subjected to the ultrasound, making the distinction visually clear. The explanatory effect of the ultrasound modeling did not go unmentioned in the PI's field notes, nor did students fail to mention the dramatic effect in their response surveys. S1 wrote, "Watching the teacher pronounce the words in person and on-screen really went well today."

Ultrasound Model Videos

While the real-time imagery of the teacher's articulations was well-received, the model videos were not so popular. During the first two instructional sessions, the PI left the model videos to the side on a laptop computer but available to students to interact with while they waited their turn with the teacher and ultrasound. However, from the first lesson, subjects showed little interest in the model videos. Instead, while waiting for their turn with the ultrasound and teacher, the subjects sat transfixed on the student-teacher interaction that was occurring in front of them. Even though they were not directly participating, students remained engaged in the activity, watching the ultrasound image projected on the large screen and listening to the teacher's feedback.

To measure subject reaction, the PI chose to include the ultrasound model video in the ultrasound half of the final lesson on flaps. He projected the ultrasound video on the large screen and played it two times through so students could hear and see the target words without interruption. Following this, the PI rewound the video to replay specific points in the video that best displayed the salient elements of the target articulations. Since *QuickTime* allows one to advance through the video frame-by-frame, the PI was able to freeze the tongue articulations at the most relevant points and discuss them with the subjects. Judging from the subject survey responses, however, this approach did not seem very well-received. S3 explicitly criticized the model video writing that it was completely unhelpful.[2] There could be some practical reasons for S3's critical response. The *QuickTime* video can advance frame-

by-frame in either direction, forwards or backwards. When one is manipulating the video at a fast-rate and rapidly reversing direction, it can become confusing to the viewer what exactly they are seeing. Perhaps if subjects themselves had control over the video, they may be able to avoid this type of confusion.

Student Interaction with Ultrasound and Instructor

Student response was overwhelmingly positive regarding the direct interaction between student and teacher/ultrasound. All subjects remarked both formally on the surveys and in passing conversation that this was the best part of each lesson. There is little reason why it would not be. The students have the visual representation of their L2 articulations available to them, and they have a teacher beside them providing immediate feedback on their L2 pronunciation. In addition, the students received commentary on their articulatory production from the PI who was helping to interpret the ultrasound imagery. For computer-aided pronunciation training, providing immediate (and accurate) evaluation of students' L2 production is a major difficulty. Neri, Cucchiarini, Strik, & Boves (2002, p.455) point out that visual readouts can be impossible for students to interpret on their own, and that teacher interaction is normally needed. Some commercially-available programs require the user to submit their recordings to an external location and wait for evaluative feedback (Pennington, 1999, p.431). This ultrasound study answers that call by pairing subjects with an instructional team for immediate and accurate feedback.

A further complication mentioned by Neri et al. (2002, p.453) is the acceptable range of articulatory variation that exists in any group of native speakers. Thus, matching one spectrogram reading (L1 model) to another (student's L2 production) may not always be necessary. Any linguistic community allows for a particular range of variation for any phoneme. This initial study answers this call too. Since the teacher is evaluating the students' pronunciation in terms of intelligibility to a native speaker, there is no pressure for the students to perfectly mimic the teacher's model articulation that is featured on the ultrasound. Although this did present some later inconsistencies during the evaluation stages (to be discussed later), the benefit of this arrangement is that students were receiving evaluative feedback that closely resembles a real-life L2 encounter. They were judged by a native-speaker who can grant them some leeway, and not by a rigid computer program that can only compare points on a grid.

PI's Side Commentary

A final element of each lesson which warrants qualitative analysis is that of the PI's side commentary both during the teacher's ultrasound modeling and during the students' practice production. The PI's commentary, drawn on his phonetics training and familiarity with ultrasound imagery, provided student subjects with corrective feedback on their L2 pronunciation specifically regarding manner of articulation. Unfortunately, the students

rarely commented directly on the PI's role as adjunct contributor. However, it is believed that the PI's contribution to interpreting the ultrasound imagery was helpful because the PI's observations were intended to help in diagnosing pronunciation problems. This was the case for S1 during the final lesson on flaps. In reference to a quick lesson instigated by the PI and in response to the subject's struggle with the flap sound, she wrote, "The most helpful today was going from the English to Japanese words to get the sounds." Furthermore, it is also believed that the PI's presence up in front of the screen helped to focus the subjects' attention on the ultrasound image in order to support their noticing of salient articulatory features of the L2, as is promoted by the Noticing Hypothesis. This strategy seemed to work for most of the students, as the PI observed S2 and S3 focusing primarily on the ultrasound screen during their interaction with the teacher. However, S1 remained transfixed on the teacher's mouth movements through to the end of the lessons.[3]

QUANTITATIVE RESULTS

Data Collection

To augment the qualitative data, a quantitative analysis was conducted to compare subjects' pre- and post-treatment L2 production. All subjects participated in a pre-treatment recording of their production of the targeted L2 articulations (long vowels [e:] and [o:], nasals, and flaps). During the recording session, subjects recorded a total of 32 word tokens that were presented to them in isolated format. Although each individual token was only read once, each target articulation featured two tokens which allowed averaging the two together for a composite score for each category. Following the three instructional sessions, all subjects recorded the same tokens again, creating the post-treatment data. A complete list of the tokens used in the recordings may be found in Appendix C.

Evaluation of Data

The subjects' recorded L2 production was presented to three native Japanese speakers for evaluation. In total, each evaluator listened to and rated 192 tokens. All three evaluators are native speakers of the Tokyo dialect of Japanese. Two of the evaluators are currently graduate students and both have at least one year of formal training in phonetics. Both have been living in the United States for 5+ years. The third evaluator has lived in the States for 3 years, but does not have any formalized phonetics training.

The evaluators rated each L2 token on a scale from 1-4 (see *Table 2*). The closer the score was to one, the closer the subject had approximated native-like L2 pronunciation. On the other hand, a rating of four meant that the token was unintelligible, or that it varied too far from the target word to be considered Japanese. When listening to the subject recordings, evaluators saw the word on the score sheet so they were aware beforehand of what the subject was attempting to say in each instance. This format was chosen because the goal was not so much intelligibility (whether or not Japanese speakers could

decipher the intended word), but rather the acceptability of the production (how close the word comes to native-like pronunciation). Of the two, acceptability is the more ambitious goal, but it was assumed that ultrasound input would help the subjects reach that goal in the short amount of time that was available. Later, it was realized that there was a discrepancy between the stated goals in the ultrasound classroom (intelligibility) and the criteria espoused during evaluation of subject L2 production (native-like acceptability). The result was an unfortunate conflict between instruction and assessment goals. Follow-up studies will address this discrepancy.

Table 2. Token Evaluator Scale

1 Native-speaker (sounds just like me!)	**2** Native-like but something a little off	**3** Clear foreigner accent, but intelligible	**4** Not Japanese (unintelligible)

Following this, ratings for each token were averaged across the three evaluators to arrive at a composite score. To facilitate easy interpretation of the data, the composite ratings were then translated into percentage numbers toward Native-Like Attainment (NLA). To begin with, the rating of *one* was assigned 100% NLA and the rating of *four* was assigned a 0% NLA. In between those two polarities, the rating of *two* was given a score of 66.66% NLA and *three* was translated to 33.33% NLA. Using those figures as benchmarks, the composite evaluation scores were changed to percentage scores. For example, a composite score of 2.67 became 44.30% NLA on the new scale. A rating of 1.33 became 89% NLA.

Inter-rater reliability was below expectations. As the reader will see in *Figure 5* below, results for S1 show a range of 4.67 points on their Pre-Test (28.00 – 23.33) and a range of 4.33 points on their Post-Test (33.33 – 29.00). On the other hand, S3 shows alarming ranges across scores: Pre-Test = 16.67 point range (52.00 – 35.33) and Post-Test = 8.33 point range (57.33 – 45.66). Nevertheless, it must be underscored that, while the actual scores did vary, the ranking of the subjects across raters did not. All evaluators placed S2 at the top, S3 in the middle, and S1 at the bottom in terms of native-like L2 pronunciation. Thus, while specific NLA scores varied in terms of inter-rater reliability, the hierarchy perceived by all three evaluators remained consistent.

Table 3. Evaluator Results Comparison Table

PRE TEST				
	Evaluator 1 (% NLA)	Evaluator 2 (% NLA)	Evaluator 3 (% NLA)	Composite Score
Subject 1 (S1)	28.00	23.33	25.00	**25.44**
Subject 2 (S2)	61.33	53.66	53.00	**56.00**
Subject 3 (S3)	52.00	41.66	35.33	**43.00**

POST TEST				
	Evaluator 1 (% NLA)	Evaluator 2 (% NLA)	Evaluator 3 (% NLA)	Composite Score
Subject 1 (S1)	33.33	29.00	29.00	**30.44**
Subject 2 (S2)	60.00	60.00	64.66	**61.55**
Subject 3 (S3)	57.33	45.66	49.00	**50.66**

General Subject Improvement

Over the course of the case study, all three subjects earned higher ratings of L2 native-like attainment according to our three Japanese evaluators (See composite scores above in *Figure 5*). Graphic representation of overall improvement for each subject may be found in Appendix D. All maintained their respective positions in the hierarchy, but S3 displayed the greatest amount of advancement in terms of percentages towards native-like attainment. S3 showed 7.6% improvement towards native-like attainment while S2 and S1 showed 6% and 5% increases respectively.

Specific Results

Moving from general results, let us turn to subject performance in individual target areas. The long vowel targets of [e:] and [o:] present somewhat conflicting results (see Appendix E, *Figure E.1*). S1 showed improvement with the long [o:], but she digressed with the long [e:]. S2 experienced pronounced improvement. Long [e:] improved 21 percentage points, and long [o:] improved by 66.51 percentage points. Despite S2's impressive results, S3 showed a slight reversal in native-like attainment. For S2, the long [o:] and long [e:] fell behind 2 percentage points and 0.67 percentage points respectively.

The results from nasals produced nearly mirror reflections of the *Long Vowel* results. Where S3 found little improvement before, S3 gives an outstanding performance (33.37 percentage-point improvement) with nasals. In contrast, S1 found little to no change while S2 actually backtracked, scoring a loss of 44.5% points in [ŋ] nasals (velar) and 22.6% points in the alveolar nasal [n]. However, curiously S2 improved with the uvular nasal [ɴ] by 22.4% native-like attainment (see Appendix E, *Figure E.2*). The improvement in the uvular nasal is not surprising as the PI observed the teacher expressing great satisfaction with his production during class. Additionally, the PI could see via the ultrasound that S2 was touching the back part of the tongue to the uvular area of the mouth. However, it is unfortunate that such control did not transpose to the alveolar and velar nasal articulations. One can only speculate that S2 gave added attention to the uvular because it was a completely novel articulation, but this attention was at the expense of other nasals that appeared in the study.

The final category of study was that of flaps, both their common (word-initial and word-medial) and their palatalized versions. The flap is notoriously difficult for American English speakers learning Japanese. To begin with, the flap is misleadingly transcribed with the alphabet letter 'r' in romanizations of Japanese script which suggests to L2 learners the bunched/retroflex American English /r/ [ɻ]. The difficulty is increased by the fact that in Japanese, the flap will range in actual articulation. Sometimes it will sound like a light /d/ to an English speaker; other times it will resemble an /l/. The reason for this variation is that there is no phonemic distinction that divides up the flap articulation into distinctive units. In some cases, the word environment will force a flap to take on some qualities of a lateral and thus sound like an [l] (i.e. *shinrigaku 'psychology'*). Other word environments dictate a [d]-like sound (i.e. *tonari 'next to'*). Regardless, to a Japanese speaker, both sounds will be perceived as the phoneme that contains all manifestations of the flap.

The majority of the data for flaps are positive. S1 recorded 22% points improvement in native-like attainment in all targeted flap environments (word-initial, word-medial, palatalized, and lateral). S2 returned slightly more robust results, recording 33% point improvement in the word-initial and palatalized environs. The word-mid was the most productive for S2 who returned 66.7% point movement toward native-like attainment. As was the case with the previous two categories, the results are not unequivocal. S3 improved with word-initial flaps (+33.27%), but digressed when placed in the word-medial position (-44.7%). Clearly, S1 and S2 benefited the most from the instructional sessions, while S3 benefited only partially (see Appendix E, *Figure E.3*). However, a deeper investigation of the data uncovers a troublespot in the flap word-medial data. One particular token /*warui*/ 'bad', experienced a 55.7% drop in native-like rating. During the pre-test, S3 scored a 100% native-like attainment, but then something happened during the post-test that resulted in a low rating of only 44.30% NLA. When placed together, this calculates to a major drop in performance. However, the other three tokens making up the word-medial section actually remain stable. Considering the fact that S3 scored so well on the pre-test and that the other word-medial tokens remained stable, it is safe to discount to anomaly status the post-test version of /*warui*/.

Statistical Analysis

Statistical analysis of the quantitative data yielded insignificant results (at the p=0.01 and p=0.005 levels). This result can be attributed to two factors: (1) the limited total amount of time each subject spent in the exploratory classroom and (2) the limited number of recorded tokens and the small number of subjects which both contributed to low power in statistical calculations. Future studies will attempt to address these issues by increasing the amount of time spent in the exploratory classroom along with an increased number of tokens recorded from each subject.

GENERAL DISCUSSION

A general discussion of this case study requires consideration of both qualitative and quantitative data that resulted from the instructional sessions. A helpful point of departure is to revisit to the original questions driving the study: 1) What contributions can ultrasound technology make to L2 pronunciation development? 2) What kind of protocol seems to work best when using ultrasound technology as part of lesson delivery? 3) What kind of reactions do students give to ultrasound-based teaching? We will address each question in turn.

Key Question One: Ultrasound Contributions to L2 Pronunciation Instruction

First, judging from observations and student survey responses, it seems that ultrasound technology can contribute valuable information to students in the form of L2 input and L2 output. During the case study, the ultrasound was able to make visible for students particular distinctions in tongue articulations that they would not be able to see otherwise. Likewise, students and instructors can see exactly what a student is doing articulatory-wise and may give feedback that is specific and well-informed. Student surveys support the warm reaction to the ultrasound. Students mentioned on every occasion that seeing the teacher model on the ultrasound was one of the best parts of each lesson. Equally well-received was the triadic interaction students shared with the teacher and the ultrasound.

At this time, the numeric data can only suggest that instructional sessions which feature both conventional and ultrasound-based instruction can contribute to L2 pronunciation development. However, due to the nature of the case study, we cannot determine what element of the instructional sessions contributed to their L2 development. Furthermore, the quantitative data is not unequivocal. While some students exhibited improvement in some areas, some students actually digressed in others. At this point such ambiguous results are acceptable as the purpose behind this case study was not to determine causality, but to set the stage for further research.

Key Question Two: Ultrasound Instructional Protocol

The second key question addressed the most effective sequence of lesson delivery. Judging from PI observations and student responses, some responsible conclusions can be reached. First, the long sentences did not work at this point in the students' L2 development. Providing opportunities to use the target words in context is a common feature of any effective pronunciation L2 curriculum (see Pennington, 1999; Neri et al. 2002), however, this case study is serves as a reminder that the sentences must be constructed at a level that is appropriate for the students. When sentences are too advanced or demands for fluency are unreasonable, the students become distracted from the real task at hand—L2 phonetic development. A second observation concerns the use of model video as L2 input for the students. In principle, this is a good

idea because it affords the student additional input. However, the student must be given complete control over the video. They must be allowed to manipulate, rewind, and replay the video at will.

Heading in the opposite direction, there were some activities that clearly worked for the students. The subjects appreciated the teacher modeling the target articulation via the ultrasound and projecting this image on the large screen. This activity was especially remarked upon when the teacher highlighted specific articulatory contrasts. One important consideration here is the fact that the teacher participating in this case study had some previous training in Japanese articulatory phonetics as well as familiarity with the ultrasound before carrying out the instructional sessions. What this means is that ultrasound-based instruction, as it stands at the moment, requires a period of teacher training prior to implementation in a classroom. For the current study, students were provided with feedback from two sources, the instructor and the PI. Ideally, however, if ultrasound-based instruction were brought to a classroom, the teacher would receive training in both articulatory phonetics and interpreting ultrasound imagery, thus incorporating both roles played by two individuals in the current study. A fully-trained teacher could perform both roles simultaneously. A second pedagogical item that worked well was the triadic interaction between the teacher/PI, student, and ultrasound. Students mentioned at the end of every lesson how helpful it was to sit with the teacher, project their tongue articulations via the ultrasound, and receive immediate feedback. On the other hand, the additional feedback component provided by the PI may or may not have been effective. Teacher intuition would say that the PI-directed feedback was effective, but lack of student response leaves one unsure.

In addition to the ultrasound component, one further strength to this study is the fact that subjects received immediate feedback on their L2 production. This study answers Neri et al.'s (2002, p.459) call for pronunciation instruction that provides immediate feedback. Not only was the feedback immediate, it also derived from three sources—the student, the instructor, and the PI. Moreover, since there was a native speaker instructor providing feedback who is a trained in L2 pedagogical techniques, this case study was able to avoid the common pitfall of computer-aided pronunciation programs that spit out erroneous or inaccurate feedback (see Neri et al., 2002, p.458). In sum, the ideal ultrasound-based instructional protocol will feature the following: (1) Teacher modeling on ultrasound, (2) Student interaction with teacher and ultrasound, (3) Immediate feedback to students regarding their L2 production.

Key Question Three: Student Response to the Ultrasound

In response to the third key question, it is reasonable to claim that the overall student response to the ultrasound was positive. Students remarked on the beneficial qualities of the ultrasound in their student surveys. The PI observed students who were fully engaged during the ultrasound lessons, remaining focused on the ultrasound images even when they were waiting

their turn. At no point during the case study did students express negative attitudes toward the ultrasound technology or to its value to the L2 classroom.

FUTURE RESEARCH

The spring 2007 study will build on the findings from the initial case study covered in detail above. First, a valid experimental environment with a control group (conventional instruction) and treatment group (ultrasound-based instruction) will be created. The control group will resemble the first half of the case study lessons (traditional instruction), and the treatment group will resemble the second half of the case study lessons (ultrasound-based instruction). The planned experimental environment will allow causal connections to be determined between the particular form of instruction and student post-experiment results.

The upcoming study will also address some of the shortcomings identified in the initial case study. This planned study will switch the subject population from beginning L2 speakers of Japanese to intermediate L2 speakers of Japanese (at least one year of formal study). This is to address two issues. First, Pennington (1999) and Neri et al. (2002) both assert that CALL pronunciation instruction should explicitly aim for one of two goals, either *intelligibility* (i.e. L1 speakers detect an accent, but communication is not impeded), or *acceptability* (i.e. L2 pronunciation closely mimics L1 speakers). As mentioned above, the current study presented students with conflicting goals. The classroom aimed for intelligibility, but the subsequent evaluative procedure demanded acceptability. To reconcile this, acceptability will be the explicit goal of the next study, a goal believed to be within reach of intermediate L2 students within the limited time of an experiment. Second, in concordance with Pennington (1999, p.436) and Neri et al. (2002, p.460), this and future studies are built on the assertion that effective L2 pronunciation instruction links communication with L2 phonetic training in the form of classroom simulations of real-life encounters. Based on the students' struggle with the L2 practice sentences featured in this initial study, it is clear that such an activity may prove to be too taxing for beginning L2 students. Instead, it is theorized that such simulations where communication goals are interlaced with phonetic ones are a clear possibility with intermediate students. Therefore, such communicative activities will be incorporated into future studies.

To increase the power of statistical calculations, the scope of the experiment protocol will be broadened in the following ways: (1) increase the number of participating subjects, (2) increase the amount of instructional time while reducing the number of target articulations, (3) increase the number of tokens recorded both prior to and following the experimental classrooms.

The follow-up study will continue with what went well in the current study. Students in both conditions (control and treatment) will receive copious amounts of L2 input, personal interaction with the instructor, and individualized feedback from a native speaker. Where subjects in the control condition will receive traditional forms of L2 input, the subjects in the

ultrasound condition will receive visual ultrasound input. Personal interaction with the instructor and subsequent feedback for the treatment group will feature the ultrasound as a mediating device, allowing the instructor to give the best-informed feedback possible to the student.

CONCLUSION

Like any other instructional tool, the ultrasound is best seen as one piece of a comprehensive L2 pronunciation instructional program. Continued study is necessary to first document the potential contributions of the ultrasound compared to traditional methods. Following that, the current research agenda will begin teasing out the individual factors of the ultrasound which contribute the most to L2 pronunciation training. Despite the large amount of research to be done, teacher intuition suggests that ultrasound technology presents undeniable benefits to L2 instruction.

REFERENCES

Hazan, V., Sennema, A., Iba, M., & Faulkner, A. (2005). Effect of audiovisual perceptual training on the perception and production of consonants by Japanese learners of English. *Speech Communication*, 47, 360-378.

Hirata, Y. (2004). Computer assisted pronunciation training for native English speakers learning Japanese pitch and durational contrasts. *Computer Assisted Language Learning*, 17, 3-4, pp. 357-376.

Lambacher, S. (1999). A CALL tool for improving second language acquisition of English consonants by Japanese learners. *Computer Assisted Language Learning*, 12 (2), 137-156.

Neri, A., Cucchiarini, C., Strik, H., & Boves, L. (2002). The pedagogy-technology interface in computer assisted pronunciation training. *Computer Assisted Language Learning*, 15 (5), 441-467.

Pennington, M. (1999). Computer-aided pronunciation pedagogy: Promise, limitations, directions. *Computer Assisted Language Learning*, 12 (5), 427-440.

Shehadeh, A. (2003). Learner output, hypothesis testing, and internalizing linguistic knowledge. *System*, 31, 155-171.

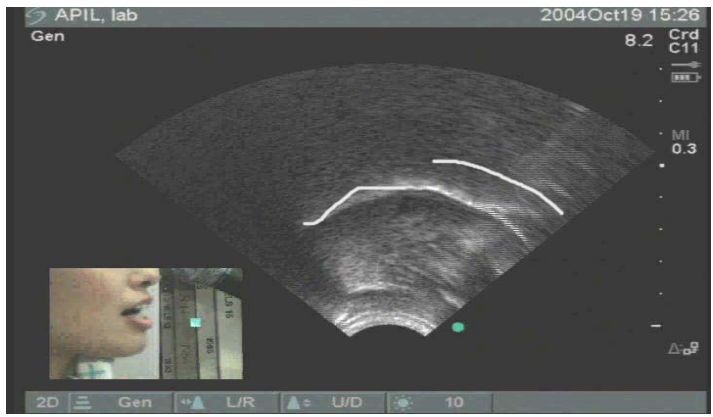
Appendix A
Lesson Handout Sample

Type One [ɺ] (flap)

Graphic:

Figure 40
Articulatory gesture for the Japanese [ɾ]

Ultrasound Image:



Recommended Steps or Description:

- Raise tip of tongue quickly—just behind the teeth ridge (the hard part of the mouth just behind the teeth).
- Shoot tongue tip forward and downwards to hit once on the teeth ridge.
- Contact is momentary and light. Contact must be loose, otherwise you get a [d]
- Body of tongue does not move
- When occurring at the front of a word (i.e., *raishuu*), it's suggested to place tongue in position for flap (lightly touching the teeth ridge this time). As you begin pronouncing the word (*raishuu*), let the outgoing air blow your tongue tip down from its position on the teeth ridge.

Connection to English: Similar to the articulation in American English of *little*, *Betty*, and *middle*.

Sample Word Lists

Word-Initial (flap occurs at the beginning of the word)		
raishuu	らいしゅう	next week
rooka	ろうか	hallway
rekishi	れきし	history
rippa	りっぱ	wonderful

Word-Medial (flap occurs in the middle of the word)		
warui	わるい	bad
kore	これ	this
yoroshiku	よろしく	'please'
dare	だれ	who
iro	いろ	color
kara	から	from
onari	となり	next to, neighbor
tori	とり	bird

APPENDIX B

Lesson Protocol Chart

Lesson Protocol Chart / USJFL Project

	Session One phonetics intro / long 'e'	Session Two long 'o' / nasals	Session Three (a) flaps / palatalized flaps	Session Three (b) flaps / palatalized flaps
Lesson Intro				
Ss read handout	x	x		
PI reviews handout		x	x	x
PI diagrams on whiteboard		x	x	x
Input				
T. writes hiragana on whiteboard	x	x	x	x
T. says target words (1 time each)	x	x	x	x
Oral Output / Feedback				
Ss say target words in unison	x	x	x	x
Ss say target words round robin-style	x	x	x	x
T. gives immediate feedback (correction=repetition)	x	x	x	x
Ss say target sentences (chunks to whole)	x			
US Input				
US model video with stop-frame function			x	x
T. models US in real-time	x	x	x	x
T models US + PI interprets on screen		x	x	x
Interaction w/US + Feedback				
Ss interact with US and teacher	x	x	x	x
Ss interact with US + dual feedback (T & PI)		x	x	x

APPENDIX C
Word Tokens
(for recording)

Section One: Long Vowels [e:] and [o:]

	[e:]	[o:]
Word-Initial	eego, eega	ookii, oosutoraria
Word-Mid	oneesan, keezai	doozo, sayoonara
Word-Final	sensee, kiree	arigatoo, benkyoo
12 words		

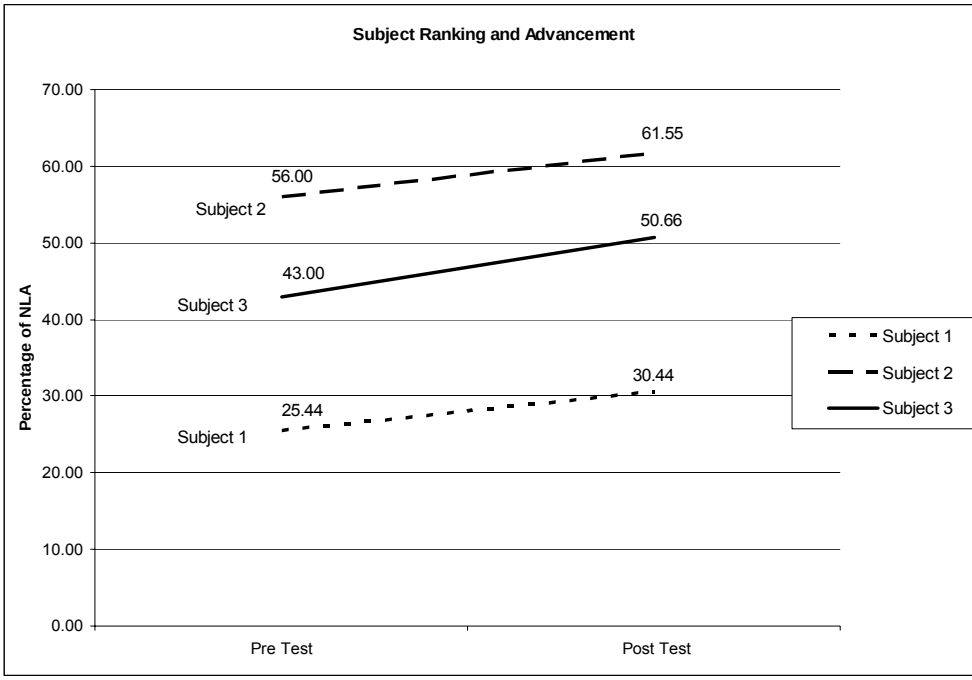
Section Two: Nasals

bilabial	alveolar	palatal-alveolar	velar	uvular (word-final)
tempura komban wa	konnichi wa	shinshitsu kenchiku	genki nihongo	kaban zabuton
9 words				

Section Three: Flaps	
Word-Initial	rippa, rooka
Word-Mid	warui, kore
Word-Final	tonari, dare
6 words	

Section Four: Palatalized Flaps	
Word-Initial	ryokoo, ryuugaku
Word-Mid	goryoushin, goryokoo
Lateral	shinrigaku
5 words	

APPENDIX D
Graphic Summary of Subject Overall L2 Advancement in the Target Articulations



APPENDIX E
Evaluative Results Graphs

Figure E.1 Change in Production of Long Vowels

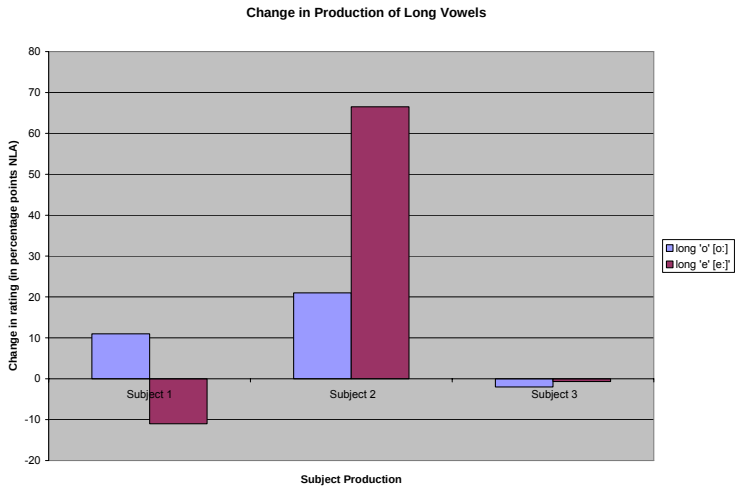
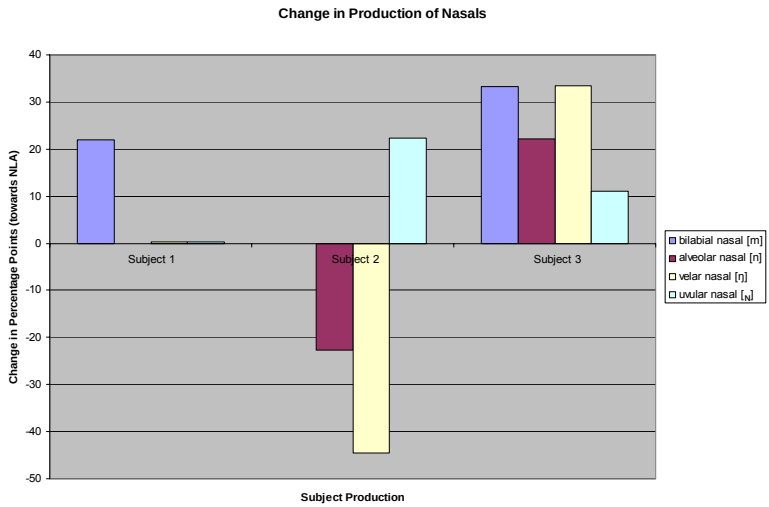
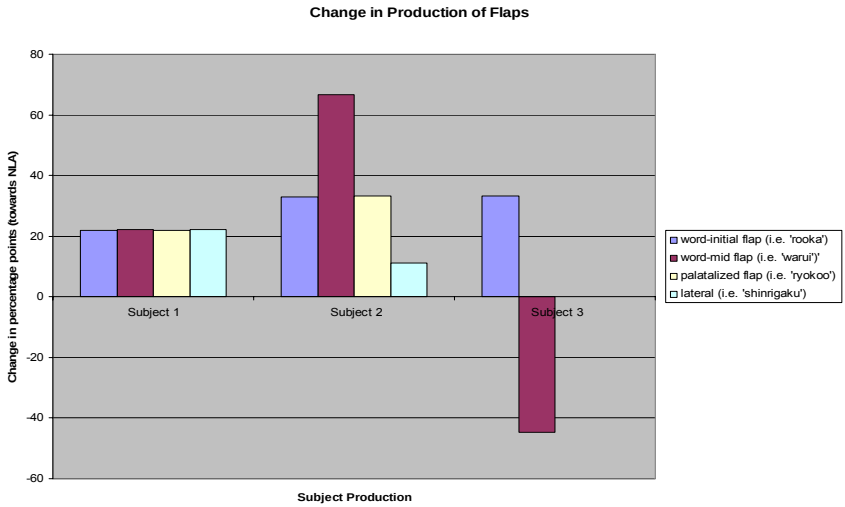


Figure E.2 Change in Production of Nasals



http://w3.coh.arizona.edu/awp/

Figure E.3 Change in Production of Flaps



[1] Many warm and sincere thanks to Dr. Diana Archangeli, Dr. Jeff Mielke, Gwanhi Yun, Peter Richtsmeier, and Adam Baker for their continued advice, suggestions, and encouraging support. They are pleasant reminders that quality academic work is never a solitary accomplishment, but a combined effort. This study was made possible by a James S. McDonnell Foundation grant.

[2] The other two subjects did not make explicit comments about the model video so it is unclear how they reacted to it.

[3] This may have been a good thing for S1. Considering the difficulty S1 was having with L2 pronunciations, the ultrasound may have been irrelevant to her at that point in time. Most valuable to her was close and direct interaction with the teacher.