COMPUTER-MEDIATED INPUT AND THE ACQUISITION OF L2 VOWELS

Michelle Minnick Fox

Department of Linguistics, University of Pennsylvania

ABSTRACT

Programs for testing and training of difficult vowel distinctions in American English were created for subjects to access via the Internet using a web browser. The testing and training data include many likely vowel confusions for speakers of different L1s. The training program focuses on one distinction at a time, and adjusts to concentrate on particular contexts or exemplars that are difficult for the individual subject. In the current study, 52 subjects participated in testing and 2 subjects participated in training. In the testing portion, results indicate that the L1 and the fluency level in English, as well as individual variability, have an effect on perceptual ability. In the training portion, subjects showed significant improvement on the contrasts on which they trained. Because these programs make extensive data collection over large populations and large distances easy, this method of research will facilitate further investigation of questions regarding second language acquisition.

1. INTRODUCTION

Several studies have shown that non-native perception of difficult phonemic contrasts can be improved through identification training that uses multiple natural tokens of the two phonetic categories and provides the learner with immediate feedback. (See Jamieson [3] for review.) In general, these studies have concentrated on homogeneous groups of subjects with a common L1, and have typically trained the subjects on a single contrast in the L2 that speakers of that L1 find difficult. These studies have had limited practical value due to this narrow scope of investigation, and also due to a lack of feasibility of implementation on a large scale for training a large number of L2 learners.

In this study, computer programs were developed which incorporate successful techniques from previous studies, with the goal of making them accessible to be used by a large number of L2 learners of English in addition to being a means to conduct research on second language acquisition. In order to accomplish this, the programs developed for this study differ from the methods used in previous studies in two important ways:

- The programs are accessible through the Internet and are controlled by the subject, with automatic data storage, so non-native speakers can use them to participate in studies from anywhere in the world. Use of the programs can also be incorporated effortlessly into the foreign language curriculum or used for individual language learning.
- The programs test and train subjects on a number of different contrasts. These contrasts were chosen to include many likely vowel confusions for speakers of different L1s: /i/-/I/, /I/-/ɛ/, /ɛ/-/ej/, /ɛ/-/æ/, /u/-/ʊ/, /ʊ/-/ʌ/, /a/-/o/, and /a/-/ʌ/. Within a given vowel contrast, the training programs automatically adjust to focus on what the individual finds problematic. Because training is customized to the individual, it can be useful for learners with different needs.

Because these programs are designed to be used by subjects with various L1s, the question of determining which contrasts a particular subject should concentrate on becomes important. Testing and training on each contrast requires a great time investment on the part of the language learner (a week or more of daily participation appears to be needed for measurable progress), and so it is important to concentrate only on those contrasts which would potentially be problematic for the individual. Best [1] and Flege [2] have suggested that the ability to perceive non-native contrasts is at least partially determined by the way that non-native phones are perceptually assimilated to native phonetic categories. If perceptual assimilation to L1 categories is a primary factor in determining L2 perceptual difficulties, then it might be expected that subjects would benefit most from training on just those contrasts that members of their L1 have difficulty with. Kingston et al. [4], on the other hand, argue that difficulty in perceiving contrasts is not primarily due to how L2 phones are perceptually assimilated to L1 categories, but instead that difficulties are based on the perceptual distance between tokens in two phonetic categories. If this is the case, then subjects should have the most difficulty with

those contrasts that are less perceptually distinct regardless of their L1, and training should concentrate on these contrasts.

For the testing portion of this study, subjects across as many L1s as possible were recruited in order to test the predictions made by the perceptual assimilation hypotheses and the perceptual distinctiveness hypothesis, as well as to evaluate other factors that may determine difficulty of L2 contrasts. For the training portion of the study, the main goal was to measure the effectiveness of the training subjects across multiple contrasts in one study.

2. METHOD

The programs for testing and training were written so that they can be accessed through the Internet from anywhere in the world. Links to samples of these programs, as well as other information regarding this study, can be found at http://www.ling.upenn.edu/~minnick/. The main interface that the subject sees is a Java applet imbedded in an HTML web page. Subjects can run the programs using an Internet browser such as Netscape or Internet Explorer, on any computer that is connected to the Internet, is Java-enabled, and has good quality sound (preferably with headphones). In a subject's first session (the pre-test session), he or she fills out a demographic form and is assigned a login id. In subsequent sessions, the subject uses the same login so that the testing and training sessions can be customized for that subject, and also so that the subject's data can be appropriately stored.

2.1 Testing

Test Format. An ABX test format was chosen so that the test would be difficult enough to avoid ceiling effects. In each question, listeners hear two words that contain different vowels, and then a test word that contains the same vowel as either the first or second words. In the first half of the test, all three words are spoken in isolation. In the second half, the first two words are spoken in isolation, and the test word is embedded in the carrier phrase "Now say X again". The inter-stimulus interval, measured from the beginning of one word to the beginning of the next within a question, is two seconds. After the beginning of the test word (test phrase), the subject is given 5 seconds (6 seconds) to click on the button corresponding to the first word or the second word.

Stimuli. Five native speakers of American English, three females and two males, produced the tokens used in the test. Each speaker produced the tokens for 8 questions on each of 8 vowel contrasts. The test words for these 8 questions consisted of two minimal pairs, each recorded both in isolation and in the carrier phrase. In addition to the test words, 4 sample words, used as the first two words in each question, were recorded for each vowel for each contrast and randomly assigned to appear with a test word. All three words for a given ABX triad were always produced by the same speaker, but the three words did not necessarily have similar phonetic contexts. The stimuli were recorded in a sound-attenuated room at 16kHz, and then converted to 8kHz, 8-bit mu-law encoding, to be compatible with the Java audio format.

Procedure. Before taking the test, subjects were given a sample test with 12 questions to familiarize them with the test format. The sample test used the contrasts /I/-/æ/ and /I/-/a/ because these contrasts are relatively easy to distinguish. During the sample test, the subjects were given exposure to both questions with the test word in isolation and with the test word in the carrier phrase. Subjects were instructed to repeat the sample test until they understood the task and felt comfortable with it.

The test itself consisted of 318 questions (8 questions per speaker per contrast × 8 contrasts × 5 speakers, with one minimal pair removed due to ambiguity). All subjects received the same questions, but the order of questions within each half of the test was randomized, thereby mixing together questions by different speakers and on different vowels. Questions were separated by a one-second pause, with a longer pause of about 5 seconds every 10 questions and a 90-second pause between the first and second halves of the test. The entire testing session lasted about 90-100 minutes.

2.2 Training

Training Format. The training stimuli were presented in a two-alternative forced-choice task. In each trial, subjects were able to play the test token as well as a sample of a word from each of the two vowel categories as many times as needed before choosing the correct vowel. The same two words (ex. *hit* and *heat*) were used as samples and as labels for the response buttons for all trials of a given contrast to reduce confusions with English orthography as much as possible and to promote the formation of robust categories across phonetic contexts. Feedback was provided after each response.

Stimuli. Words in isolation and in the carrier phrase were recorded by a total of 14 speakers for the training data. All of the speakers were different from those in the testing. To the greatest extent possible, words used in the training were different from those used in testing; this was not possible for some contrasts because they have very few minimal pairs. In order to compensate for the lack of minimal pairs in some contrasts, a few non-words were used in addition to real words. Approximately 150-200 tokens were recorded for each contrast.

Procedure. The program concentrated on one contrast at a time, and automatically began on a new contrast when the subject met criterion as measured by a progress test. The tokens for a given contrast were grouped into smaller sets of tokens, usually with the words in a given set having the same or similar phonetic contexts. Each set was concentrated on one at a time, with periodic reviews of the tokens from all previous sets. The program automatically adjusted to the subject's performance, concentrating on particular tokens that the subject found difficult. Subjects used the training program on their own for a total of about 15 sessions of 400 questions each over the course of 3 weeks.

3. RESULTS

3.1 Testing

A total of 52 subjects were tested in three different testing locations: 35 subjects were participated at the University of Pennsylvania, 13 subjects participated at the University of Maryland, and 4 subjects participated from their homes in Japan. Table 1 shows the number of subjects in each L1 and their performance on each contrast. Overall, native speakers of English answered 92% of the questions correctly, while non-native speakers only answered 69% of the questions correctly. Unanswered questions were labeled as incorrect, but these accounted for only about 5% of the questions. English speakers performed above 88% on all of the contrasts with the exception of /a/-/ʌ/. The poor performance on this contrast may be due to confusion between /a/ and /ɔ/ in the test stimuli.

A multivariate analysis was run on the overall scores of the non-native speakers using three dependent variables: years of English taken in school and college; self-reported amount of training specific to vowel contrasts; and length of time spent in the United States. The years of English and amount of training had small effects on performance, although these were not significant. The length of time spent in the United States did not have an effect. However, self-reported fluency (Beginner/ Intermediate/Advanced/Near-Fluent) was a better predictor of overall performance (Figure 1). When the fluency level was taken into consideration, the native language did not have a significant effect on the overall performance.

Figure 2 is a graphical representation of the data in Table 1. While the performance on some contrasts appears to be similar across languages, performance on other contrasts varies greatly. For the

L1	N	/i/-/ɪ/	/ɪ/-/ɛ/	/eɪ/-/ɛ/	/ɛ/-/æ/	/u/-/ʊ/	/ʊ/-/ʌ/	/o/-/a/	/a/-/ʌ/	Ave.
English	7	98	94	97	93	88	89	93	80	92
All non-native	45	68	72	69	59	56	71	73	53	65
Chinese	11	67	80	74	59	57	85	89	52	70
Japanese	9	73	68	73	67	65	74	71	54	68
Korean	7	65	77	70	51	54	80	76	53	66
Jpn. (in Japan)	4	68	53	63	51	53	59	60	56	58
Spanish	4	64	59	72	63	50	59	62	44	59
Portuguese	3	65	89	68	55	60	65	81	61	68
Other	7	72	77	66	66	55	74	72	55	67

Table 1. Number of subjects with each native language (N) and the percent correct on each of the contrasts. The category "Other" includes 2 Turkish, 1 French, 1 Hindi, 1 Italian, 1 Thai, and 1 Ukrainian. The Japanese subjects taking the test in Japan are listed separately.

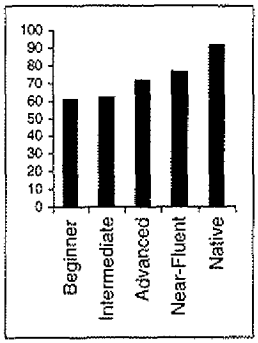


Figure 1. Performance by Self-reported fluency.

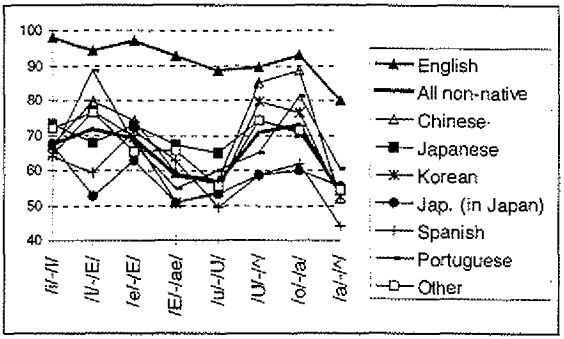


Figure 2. Average performance by native language for each contrast.

subjects tested, /ɛ/-/æ/ and /u/-/ʊ/ are generally more difficult for all languages than /i/-/I/ and /ej/-/ɛ/. This suggests that some contrasts may be intrinsically more difficult because the categories are less perceptually distinct, and in these cases perceptual assimilation may play a secondary role. Other contrasts, such as /I/-/ɛ/, /ʊ/-/ʌ/, and /o/-/a/, show great variability across languages. For example, the Portuguese L1 subjects performed nearly as well as native English speakers on /i/-/ɛ/, while the Spanish L1 subjects performed poorly on this contrast. Because there are great language-specific variations in these contrasts, it seems unlikely that the intrinsic difficulty of the contrast is the determiner of subject performance. Instead, perceptual assimilation to L1 categories may be playing a larger role.

In many cases, the subjects of a given L1 had difficulty with the same contrasts, which is what we would expect if assimilation to native phonetic categories is what causes certain contrasts to be difficult. In these cases, subjects would benefit greatly by simply studying the contrasts that are difficult for subjects from that L1. However, subjects from a single L1 do not always have difficulty with the same contrasts. As one example, Figure 3 shows the individual results for each of the 4 Spanish L1 subjects. Two of the subjects performed well on the /i/-/I/ contrast, and two performed poorly. Likewise, there is great variability in the performance on /ʊ/-/ʌ/ among the Spanish speakers. In these cases, training all subjects with the same L1 on the same contrasts would be sub-optimal. Instead, subjects would benefit most from training on the contrasts that they did poorly on in the test.

Overall performance across different speakers by non-native subjects varied less than 4%, but there were great variations across speakers for different vowels. Since each of the speakers read different words, it is impossible to determine the effect of speaker and the role of phonetic context. However, the finding that performance varies with speaker and with context is consistent with the perceptual assimilation results reported in Strange et al. [6], where significant speaker and context effects were found in the assimilation of German vowels by American subjects.

Non-native subjects on average performed slightly higher (67% vs. 64%) on the questions with the test word in isolation than with it in the carrier phrase, but this varied somewhat across speakers. After taking the test, some subjects reported that the questions with the carrier phrase were harder because the important word was in fluent speech. Others reported that the questions with the carrier phrase were easier because they were able to adjust to the speaker before hearing the test word. Strange et al. [5] found that the perceptual assimilation of long and short English vowels to Japanese vowels was affected by this type of isolated word/carrier phrase variation; when the vowel was in a phrase, durational cues were used to assimilate long and short vowels to different Japanese vowel categories. The data from the Japanese L1 subjects in the current study were compared to see if performance was relatively better on words in the carrier phrase than on isolated words for the contrasts that differ in length (/i/-/I/, /ej/-/ɛ/, /u/-/ʌ/). This was not the case; Japanese subjects on average did worse on the words in carrier phrases than for isolated words for these contrasts. This does not disprove the perceptual assimilation hypothesis, however, because it is possible that by the test format where the subject heard three words in a row spoken at about the same rate reduced the difference in the use of the durational cues in the two question types.

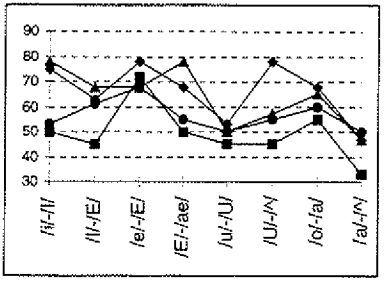


Figure 3. Performance by each of the 4 individual Spanish L1 subjects on each contrast

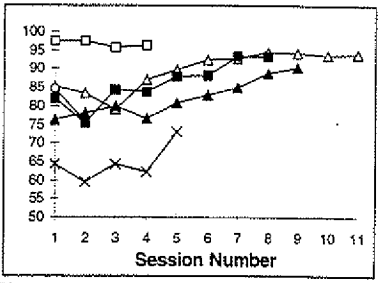


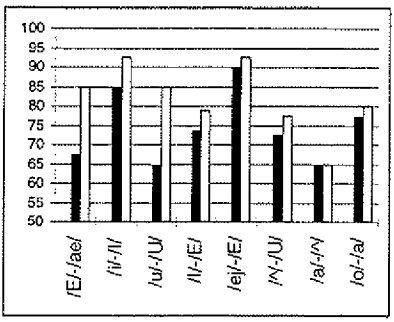
Figure 4. Performance by Session number for the contrast. Filled markers: Subject 1; open markers: Subject 2; X's: Subject 3. Triangles: /ɛ/-/æ/; squares: /i/-/I/; circles: /u/-/ʌ/.

Even in the cases where a subject performs relatively well on a contrast, performance is almost always below that of the native speakers of English. This is particularly true for subjects with lower fluency levels. This suggests that the subjects have not formed robust phonetic categories across different speakers and contrasts as the native speakers have. In this case, subjects may be able to benefit from training even on those vowel contrasts that gave the subjects less difficulty.

3.2 Training

Two subjects participated in the training program for 3 weeks. An additional subject participated in the training for one week but then could not continue due to time constraints. The results from the training sessions for each contrast trained on are shown in Figure 4. Subject 1, a 32-year old Japanese female, trained on /ɛ/-/æ/, /i/-/I/, and /u/-/ʊ/. Subject 2, a 27-year old Chinese male, trained on: /ɛ/-/æ/ and /i/-/I/. Subject 3, a 33-year old Chinese male, trained only on /ɛ/-/æ/. The initial dips after the first session are to be expected, since the tokens that the subject finds more difficult are played more frequently. Generally, improvement starts to be seen by about the fifth day of training on a particular contrast.

Figure 5 displays the results for the pre- and post-tests for Subjects 1 and 2. Subject 1 showed some improvement on all the contrasts except for /a/-/ʌ/, however there was marked improvement on the three she trained on: nearly a 20% improvement on /ɛ/-/æ/ and /ʊ/-/ʌ/, and 7% improvement on /i/-/I/. Subject 2 showed less overall improvement, and showed only a small amount of improvement on /ɛ/-/æ/, but marked improvement on /i/-/I/. Subject 2 also improved on /I/-/ɛ/, which is a contrast that he



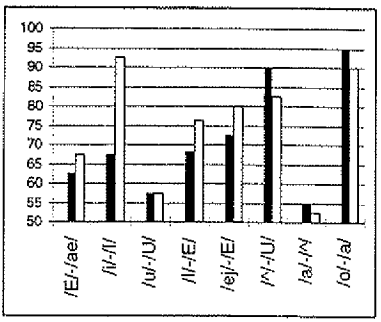


Figure 5. Performance by Subject 1(left) and Subject 2(right) on the pre-test (black columns) and the post-test (white columns).

did not train on specifically, although the contrasts that he did train on contain these two vowels. It is not clear why Subject 2 only improved a small amount on /ɛ/-/æ/, since more time was spent on that contrast than on /i/-/ɪ/, and because Subject 1 improved significantly on this contrast. Further studies with a larger number of subjects are planned to investigate the role of L1 and the length and intensity of training on improvement on different contrasts.

4. CONCLUSION

The data from the testing portion of this study is partially consistent with both the perceptual assimilation hypothesis and the perceptual distinctiveness hypothesis; some of the American English vowel contrasts studied appear to be universally more difficult than others, while other contrasts vary in their difficulty across L1s. However, there is variation in performance across subjects within an L1 group, and so the predictions of the two hypotheses are not accurate enough to reliably predict which contrasts a particular subject will find difficult. In order to make the training as effective as possible, subjects of all L1s should be tested on all of the contrasts prior to training to pin-point the contrasts that are the most difficult for the subject. Subsequent training would then concentrate first on the contrast that the subject scores the lowest on, and progress to easier contrasts after the harder ones have been learned.

Because the programs developed for this study are readily adaptable and make extensive data collection easy, this method of research will ultimately facilitate the investigation of many questions surrounding second language acquisition of phonemic contrasts. For example, the two subjects who participated in the training appear to have improved somewhat on contrasts that they did not train on. Future investigation is planned to measure the effect of training on one or more vowel contrasts on the ability to perceive other vowel contrasts. The programs can also be used to investigate questions that have not previously been easy to research because previous methods were confined to the laboratory. For example, the effectiveness of perceptual training for subjects in the United States can be compared against the effectiveness for subject in settings where English is not the dominant language. Finally, because the programs are written in Java, they are easily expandable, and can be applied to other second language acquisition questions such as the relationship between perception and production.

5. ACKNOWLEDGEMENTS

I am greatly indebted to my advisor Mark Liberman at the University of Pennsylvania and JoAnne Negrin-Cristiani for her contributions to the preliminary organization of this study and for coordination of a pilot study.

5. REFERENCES

1. Best, C. T. "A direct realist view of cross-language speech perception." In W. Strange (ed.) *Speech perception and linguistic experience: Issues in cross-language research.* York Press, Timonium, MD, 1995.

2. Flege, J. E. "Speech learning in a second language." In C. Ferguson, et al. (eds.) *Phonological development: Models, research and application.* York Press, Timonium, MD, 1992.

3. Jamieson, D. G. "Techniques for training difficult non-native speech contrasts." *Proc. ICPhS*, Vol. 4, 100-107, 1995.

4. Kingston, J., Bartels, C., Benki, J., Moore, D., Rice, J., and Thorburn, R.. "Learning Non-Native Vowel Categories." *Proc. ICSLP*, 1996.

5. Strange, W., Akahane-Yamada, R., Fitzgerald, B., and Kubo, R. "Perceptual Assimilation of American English Vowels by Japanese Listeners." *Proc. ICSLP*, 1996.

6. Strange, W., Bohn, O.-S., Trent, S., McNair, M., Bielec, K. "Context and speaker effects in the perceptual assimilation of German vowels by American listeners." *Proc. ICSLP*, 1996.