University of Pennsylvania Working Papers in Linguistics

Volume 6
Issue 1 *Proceedings of the 23rd Annual Penn Linguistics Colloquium*

Article 21

1-1-1999

Perception and Production of American English Tense and Lax Vowels by Japanese Speakers

Michelle Minnick Fox

Kazuaki Maeda

Perception and Production of American English Tense and Lax Vowels by Japanese Speakers

Perception and Production of American English Tense and Lax Vowels by Japanese Speakers

Michelle Minnick Fox and Kazuaki Maeda

1 Introduction

It has been widely recognized that non-native speakers often have difficulty perceiving and producing phonemic contrasts in a second language (L2) that do not exist in their native language (L1). Best (1995) and Flege (1992) have claimed that the ability to perceive non-native contrasts is at least partially determined by the way that non-native phones are perceptually assimilated to their native phonetic categories. In this paper, we report the results of a perceptual and production study of native speakers of Japanese (J) of the American English (A.E.) contrast between the two high front vowels /i/ and /I/. Each of these two A.E. vowels can be considered phonemically equivalent to a different vowel in J, and many J speakers of A.E. are able to satisfactorily categorize these two A.E. vowels in the most favorable circumstances. However, because the primary acoustic cues used by speakers of the two languages are different, we hypothesized that the perception and production of the A.E. vowels by J speakers would be influenced by the native cues rather than the cues used by A.E. speakers.

The perception and production experiments described in this paper were designed to focus specifically on three questions:

- To what extent do native J speakers use duration versus vowel quality to perceptually distinguish A.E. /i/ and /I/?
- What are the acoustic characteristics of A.E. /i/ and /I/ produced by J speakers?
- What is the intelligibility of J speakers' productions of these vowels as judged by native A.E. speakers?

2 Background

Both Best (1995) and Flege (1992) account for at least some of an L2 speaker's difficulty in perceiving non-native contrasts by the way that the L2 sounds are perceptually assimilated to L1 categories. For example, according to Best's Perceptual Assimilation Model (PAM), a non-native speaker's perception relies to a large extent on the native phonemic category that is

closest to the non-native category. Thus, if two different phonemes from the L2 are perceptually assimilated to the same category in the L1 (i.e. both sound like they belong to the same L1 category), and are considered relatively good fits to the category, this theory predicts that the contrast will be difficult to distinguish. A likely reason for this difficulty is that the L2 speakers continue to use the acoustic criteria which are important for discriminating phonemes in their L1.

Flege's (1995) Speech Learning Model (SLM) hypothesizes that the production of a phoneme often corresponds to the properties represented in its internal phonetic category representation. According to this theory, non-native-like production by more advanced L2 speakers is not *primarily* due to difficulty in motoric skills, although such difficulty may cause improvements in production to lag improvements in perception. Since the speaker's internal phonetic representation of a category is presumably closely related to how the speaker perceives the sound, we would expect non-native-like production to be highly correlated with non-native-like perception.

In Japanese, there are five pairs of long (two-mora) and short (one-mora) vowels. For each pair, the long and short vowel only differ in duration, while their vowel qualities are nearly identical (Shibatani, 1990). Thus, the primary acoustic cue distinguishing the two high front J vowels /i:/ and /i/ is duration. In contrast, the two high front A.E. vowels /i/ and /I/ differ in both vowel quality and in duration (Jones 1962, Klatt 1976). Many researchers consider vowel quality to be the primary cue to vowel identity in A.E., while vowel length is a phonologically redundant feature. Furthermore, the duration of an A.E. vowel is affected by various factors, including the following coda consonant, speech rate, stress, emphasis, and boundary condition. When these factors are fixed, /i/ is in general longer than /I/ (Peterson and Lehiste 1960), but when these are varied, the relative durations may vary, reducing the reliability of the duration cues.

Strange et al. (1996a, 1998) measured the perceptual assimilation of several A.E. vowels to J vowel categories by asking J subjects to indicate which J vowel was the most similar to each of the A.E. vowels used as stimuli. In addition, the subjects rated each of the vowels on a scale of 1-7 for "category goodness," with a 7 indicating the "best fit" to the Japanese category. Two types of stimuli were used in this study: a *disyllable condition*, where the vowels were presented in the context /hVba/, and a *sentence condition*, where the vowels were presented as part of the carrier phrase "I say the /hVb/ on the tape." The results that Strange et al. (1998) obtained for just A.E. /i/ and /I/ are shown in Table 1.

As the data in Table 1 indicate, in the *disyllable condition*, native speak-

	Disyllable condition			Sentence Condition		
AE vowel	Modal R	%	G	Modal R	%	G
i:	i	59	6	ii	83	6
I	i	58	3	i	77	4

Table 1. Data reported by Strange et. al (1998) on perceptual assimilation of A.E. /i/ and /I/ to Japanese categories

ers of Japanese responded most often ("Modal Response") that both the A.E. vowels /i:/ and /I/ were closest to the one-mora J vowel /i/, while in the *sentence condition*, the A.E. /i/ assimilated most often to the two-mora J vowel /ii/ and A.E. /I/ assimilated to the one-mora J vowel /i/. In both cases, the A.E. vowel /i:/ was rated higher in category goodness than /I/, indicating that the vowel quality of A.E. /i:/ is perceived as being somewhat closer to the vowel quality of the J high front vowels than A.E /I/. It is also important to note that in the absence of a "larger rhythmic context" (i.e. more than just two syllables), the subjects had difficulty using the duration information of the A.E. vowels in determining which J vowel was closest, even though the stimuli used for the disyllable condition had a similar duration contrast between the two vowels as the stimuli in the sentence condition.

If the hypotheses proposed by Flege and Best are correct, we would expect that since both A.E. /i/ and /I/ assimilate to the same vowel quality category, native J speakers will have difficulty categorizing the vowels, particularly when the vowel duration cues are removed or weakened. In the perceptual experiment, we therefore tested non-native speakers of English on their ability to categorize the two vowels both in words in isolation (weak duration cues) and in words in carrier phrases (robust duration cues). We also manipulated the duration cues in the stimuli to completely remove all duration cues and to make the duration cues contradict the vowel quality cues.

To test Flege's (1995) suggestion that experienced non-native speakers produce phonemes "correctly" according to their internal category representation, we had the same group of J subjects produce A.E. words containing the two vowels /i/ and /I/. If the subjects are both perceiving and producing the vowels according to their internal phonetic representations of these vowels, then we would expect to see a correlation between the subjects' use of acoustic cues in perception and production.

3 Perceptual Study

3.1 Stimuli

Three native speakers of A.E., two females and one male, produced the stimuli used in the perceptual portion of this study. The stimuli consisted of minimal pairs differing only in the vowel (/i/ vs. /I/), and all of the words used as stimuli were monosyllabic so that there would be no question as to which vowel was to be categorized. Since perceptual assimilation of L2 vowels to L1 categories can depend on the consonantal context as well as on the speaker (Strange et al. 1996b), the words used in the experiment were chosen to maximize the variety of consonantal contexts. The stimuli included words in isolation and words in the carrier phrase *Now say X again.*

All of the stimuli were recorded in a sound-attenuated room at 16kHz. The vowels in the stimuli were then manually labeled, and the durations of the vowels were modified in three different ways using the TD-PSOLA algorithm (Moulines and Charpentier, 1990), leaving the rest of each stimulus unmodified. This resulted in four token type variants of each stimulus:

- *natural token*: no duration modifications
- *shortened token*: length of the vowel shortened by a factor of 1/2
- *lengthened token*: length the vowel lengthened by a factor of 2
- *uniform token*: length of the vowel modified to within one pitch period of 140ms for all tokens

After the tokens were recorded and the three series of stimuli were created, a native speaker listened to all of the tokens. Due to the resynthesis involved in the duration modifications, some of the stimuli did not sound natural. If any of the stimuli of a given series was judged to sound unnatural or to not be a good exemplar of its phonetic category, all variants of that stimulus were removed from the stimulus set. The resulting set of stimulus tokens consisted of 20 minimal pairs in each of the 4 stimulus type series.

The acoustic characteristics of the *natural* tokens of the words spoken in isolation are shown in Table 2. The mean values for duration (ms), F1 (Hz), and F2 (Hz) are shown. The values in parentheses indicate one standard deviation. The acoustic characteristics of the other token types vary from these values only in their duration.

3.2 Test Format

The format of the perceptual study consisted of a two-choice forced identification task. For each question, the subject heard the stimulus, and had to

		Duration (ms)	F1 (Hz)	F2 (Hz)
Speaker 1 (F)	/i/	200 (57)	373 (25)	2743 (51)
	/ɪ/	134 (32)	532 (51)	2112 (126)
Speaker 2 (F)	/i/	156 (35)	385 (27)	2804 (94)
	/ɪ/	116 (14)	421 (52)	2503 (198)
Speaker 3 (M)	/i/	140 (44)	257 (13)	2278 (104)
	/ɪ/	129 (23)	387 (47)	1854 (75)

Table 2. Acoustic characteristics of the vowels of the words in isolation in the perception study (*natural* tokens only). The values in parentheses indicate one standard deviation.

enter on the computer whether the vowel in the word was the same as the vowel in the word *beat* or the vowel in the word *bit*. Reference to the actual words in the minimal pair was avoided to reduce confusion between the vowel sound and English orthography. The subjects were instructed to listen only for the vowel sound rather than trying to identify the word that was spoken. No feedback was provided to the subjects during any portion of the test.

3.3 Test Procedure

The perceptual test was performed on the computer in a quiet lab using headphones set to a comfortable listening level. The test was self-paced; subjects controlled the playing of the stimulus and were permitted to listen to a stimulus token more than once if needed, although they were discouraged from listening to a particular stimulus more than necessary.

The perceptual test was divided into four separate sections:

1. *Natural, lengthened,* and *shortened* stimuli; words in isolation
2. *Natural, lengthened,* and *shortened* stimuli; words in carrier phrase
3. *Uniform* stimuli only; words in isolation
4. *Natural* stimuli only; words in isolation

The questions within each section were presented to each subject in a random order, thereby mixing the stimuli produced by the different speakers (and of the stimuli of different length types in the first two sections). For sections 1, 2, and 3 of the test, the subjects were notified that the durations of the vowels

in some of the stimuli might have been modified. The entire testing session lasted approximately 30 minutes.

3.4 Results

A total of 12 native speakers of Japanese participated in the perceptual portion of the study. The speakers ranged in age from 23 to 35 (average 28.5). All were living in the United States at the time, for an average of 18.4 months.

Figure 1 shows the percent correct for each of the subjects across all sections of the test. Figure 2 shows the data for the section of natural words in isolation only (section 4) and the section of uniform words in isolation only (section 3). Several of the subjects performed at over 90% correct on the natural tokens. However, nearly all of the subjects, especially those who did well on the natural tokens, performed worse on the section of uniform tokens than they did on natural tokens. This indicates that while the non-native speakers are able to achieve high levels of performance on tokens with the duration information present, when the duration information is not available, subjects are less able to properly use spectral information to classify the tokens. Because the subjects were able to perform better on the natural tokens of words in isolation than on the uniform tokens of words in isolation, the subjects must have been able to use the duration information available, even though the cues to duration were relatively weak in the absence of the whole sentence.

Figure 3 breaks down the performance across all subjects for each of the

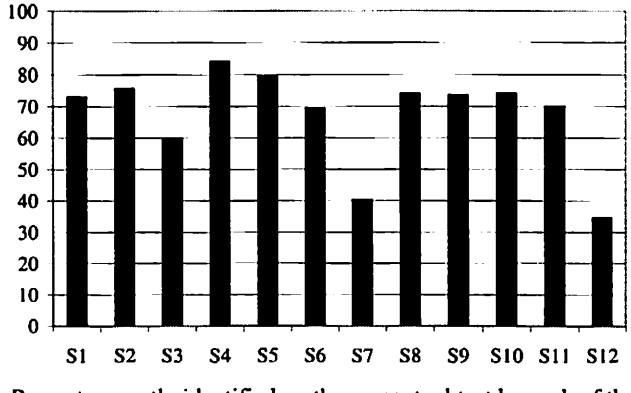

Figure 1. Percent correctly identified on the perceptual test by each of the subjects on all tokens.

token types (lengthened, natural, and shortened) for each of the vowels in section 1 (isolated words) and section 2 (in carrier phrase) of the test. Performance on these tokens is of particular interest because these tokens not only require the subjects to attend to the spectral information to correctly identify the vowel, but in the case of the shortened /i/ and lengthened /I/, in order to correctly identify the tokens, the subjects need to *disregard* the conflicting temporal information. As we would expect if the subjects rely on duration cues, the subjects performed better on the lengthened /i/ than on the natural /i/, and they performed much better on both of these than on the shortened /i/. The opposite pattern is found in the subjects' performance on

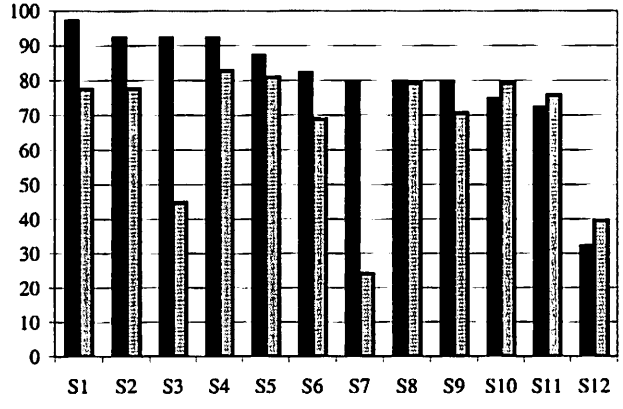

Figure 2. Percent correctly identified by each of the subjects on natural tokens (black) and on uniform length tokens (gray).

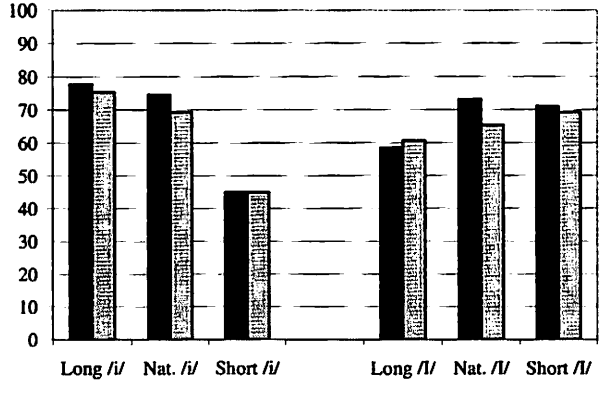

Figure 3. Percent correctly identified by token type for isolated words (black) and words in carrier phrases (gray).

/I/, with performance on lengthened /I/ worse than performance on natural /I/ and shortened /I/. Thus it appears that the subjects are attending to the duration cues. However, if the subjects were only using duration cues, we would expect a very low (close to 0%) performance on shortened /i/ and lengthened /I/. Since they perform better than this, the subjects must be able to attend to the spectral cues to a certain degree. The fact that the difference between lengthened and shortened /i/ is much greater than the difference between lengthened and shortened /I/ is consistent with the hypothesis that the subjects are able to use spectral cues to a certain extent in their categorization; the average vowel in the case of lengthened /I/ is longer than in the case of shortened /i/, so that the subjects have longer time to attend to the spectral cues in lengthened /I/.

Further evidence that the subjects rely primarily on duration cues when those cues are available, but are also able to use spectral information somewhat, is shown in Figure 4. Figure 4 shows three ratios for each of the speakers that produced the tokens used in the test:

- the ratio between the average *duration* of /i/ and /I/,
- the ratio between the average *F1* of /i/ and /I/, and
- the ratio between the average *F2* of /i/ and /I/.

Also plotted on the same graph are the overall percent correct by all subjects for each of the speakers for both the uniform tokens and the natural tokens. As the plot indicates, the performance on the natural tokens, which is in general higher than the performance on uniform tokens (with the exception of Speaker 3's tokens, where performance was nearly identical for both types of

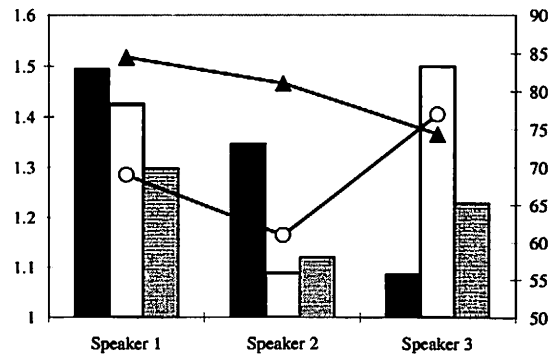

Figure 4. Columns (scale on left): Average ratios between /i/ and /I/ in duration (black), F1 (white), and F2 (gray). Lines: Performance by all subjects on each speaker's tokens; natural tokens only (triangles), uniform tokens only (circles).

tokens), is more closely related to the duration ratio; the subjects performed best on tokens produced by the speaker with the greatest average duration distinction between the two vowels. In contrast, performance on the uniform tokens, when duration information is absent, is more closely related to the F1 ratio.

Thus, in classifying natural tokens, subjects appear to rely on duration to a great degree. When the duration cues are removed, subjects use (at least to some extent) the spectral cues. Some subjects attain a high level of performance on natural tokens, but the performance on uniform tokens does not achieve native levels.

Since native speakers of A.E. perform at or near 100% for all token types[1], the results from the perception study indicate that the Japanese speakers are *not* using the same acoustic cues in perception as native speakers of A.E. However, because many of the Japanese speakers do attain a high level of correct categorization of natural tokens, it is unlikely that these speakers receive the type of feedback necessary to modify their internal phonetic category representation of the two vowels. If this is the case, we might expect this to be reflected in their production as well, causing their distinction between the two vowels to be primarily in vowel duration and only secondarily in vowel quality.

4 Production Study

4.1 Recording Procedure

Seven of the subjects from the perception experiment participated in the production portion of the study. Each of the vowels were recorded in the context /h_C/, where there were three different following coda consonants /C/: /p/ (unvoiced stop), /d/ (voiced stop) and /m/ (nasal). Each of the words was recorded three times each in each of three situations:

- word in isolation
- word in the carrier phrase *Now X is the word I say*
- word in the contrastive phrase *The word is X* (ex. *"heed"), not Y ("hid")*

Each of the words or phases was presented to the subject on a computer screen in randomized order, with an interval of five seconds between words.

[1]Two native speakers of A.E. took the perceptual test. One performed at 100% for all token types, and the other performed at 100% for all tokens types other than shortened /i/, where she performed at 98%.

The subjects were recorded in a sound-attenuated room. The recordings were digitized at 16kHz with 16-bit quantization on a Sun workstation, and the acoustic measurements were taken using Entropic Research Laboratories Waves+ and ESPS software. Vowel duration was measured by manually labeling the beginning and the end of each vowel. Formant values were calculated using the results of a 12th order LPC analysis that had been visually checked with wide-band spectrograms and corrected as necessary by the authors. The means of F1 and F2 in the central one third of each vowel were recorded for later analysis.

4.2 Intelligibility Procedure

In addition to an objective measure of the performance of the subjects according to the duration, F1, and F2 characteristics of the two vowels, an intelligibility test was also conducted of the recordings of the words in isolation produced by the J subjects. For each of the subjects, there were 2 vowels * 3 contexts * 3 repetitions = 18 tokens to be evaluated. Four native speakers of A.E. listened to each of tokens produced by each of the subjects 3 times. The tokens were presented in randomized order, to prevent the evaluators from adjusting their criteria according to the Japanese speaker as much as possible. The evaluators were instructed to respond with either /i/ or /I/ according to which vowel the production *sounded* more like, not according to what they thought the speaker intended to produce. In the rare cases that the vowel sounded more like a vowel other than /i/ or /I/, the evaluators were forced to select which one of these two vowels it sounded *more* like.

The format for the evaluation was similar to that of the perception test; the evaluators were able to listen to each token as many times as desired before selecting one of the two vowels.

4.3 Results

Table 3 shows the average durations of the vowels produced by the J subjects, and Figure 5 shows plots of the average F1 and F2 values. Because the phrase type did not have a significant effect on either the durations or the formants of the productions, data from all phrase types are included together. The following coda consonant had an effect for the duration, but not for the formants, so only the duration information is separated according to coda consonant. The effect of the coda type is consistent with the results of other studies of native A.E. speakers' productions (for example, Peterson and Lehiste, 1960); the mean duration of the vowel is approximately 30% longer

	/hip/	/hid/	/him/	/hIp/	/hId/	/hIm/	Ave. ratio
S1	146 (27.9)	212.6 (45.4)	210.9 (45.4)	116.8 (23.6)	140.3 (37.3)	132.7 (30.5)	1.46
S3	180.4 (31.1)	260.3 (25.2)	261.8 (38.7)	90.6 (14.3)	112 (13.6)	114.8 (22.2)	2.20
S5	143.6 (31.1)	192 (37.5)	174.9 (46.3)	83.2 (12.0)	122.9 (16.8)	107.8 (14.2)	1.61
S6	187.4 (32.5)	232.4 (29.6)	255.6 (47.3)	99.7 (23.0)	118.6 (17.4)	144.6 (52.3)	1.86
S8	188.4 (39.2)	267 (46.3)	212.6 (35.0)	147.3 (35.0)	176.4 (36.7)	172.4 (29.5)	1.36
S10	144.1 (18.3)	201.2 (27.2)	202.9 (22.0)	115.7 (13.9)	129.2 (12.2)	129.3 (11.3)	1.46
S12	150 (35.2)	204 (48.2)	216.9 (31.6)	104.9 (14.7)	120.8 (19.1)	113.9 (18.1)	1.69

Table 3. Duration of vowels (in ms) produced by each subject by context.

when the coda is a voiced stop than when the coda is a voiceless stop. Across all speakers, the ratio of /i/ to /I/ was 1.66 to 1. Individual speakers ranged from a smallest ratio of 1.36:1 (Subject 8) to a largest ratio of 2.20:1 (Subject 3). Thus, the J subjects consistently distinguished the two vowels in duration, and the ratio was greater than that of the vowels produced by the native A.E. speakers for the perceptual test (see Figure 4).

As the plots in Figure 5 indicate, with the exception of subject S12, all of the J subjects' average /i/ had a lower F1 and a higher F2 than the average /I/. This tendency follows the difference in F1 and F2 values as produced by native speakers of English. However, many of the J subjects whose two vowels had a *significantly different mean value* did not produce the vowels in a manner that would allow for easy categorization of the two vowels according to their vowel quality. This is shown by the fact that for F2, subjects S3, S6, and S8 do produce mean differences with the proper tendency, but there are overlaps in the distributions of the two vowels, as measured by the error bars of one standard deviation overlaid on the means.

The extent of overlap in the vowel quality of /i/ and /I/ is more clearly seen in the plots of individual tokens as a function of the first and second formants in Figure 6. The plot for some of the subjects, in particular S1, shows two distinct distributions for the two vowels, while the plot for other subject, in particular S12, shows a clear lack of separation of the two vowels. The plot of individual tokens produced by a native speaker of A.E. also

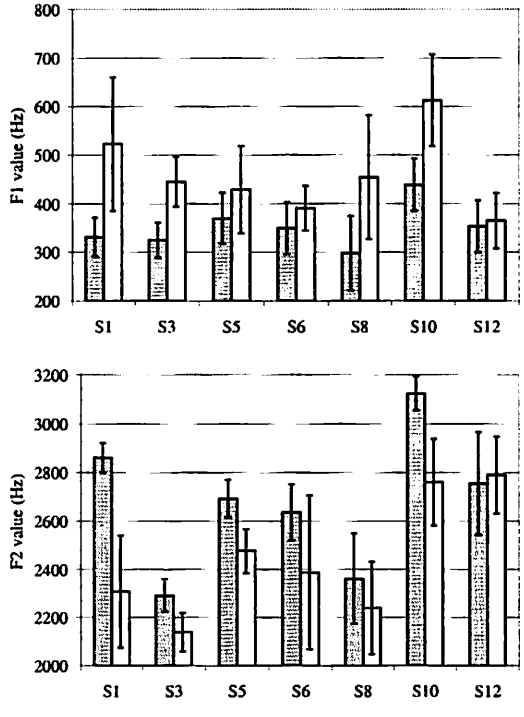

Figure 5. Average F1 (top plot) and average F2 (bottom plot) for each of the J subjects for all productions of the vowel /i/ (gray) and the vowel /I/ (white). The error bars indicate one standard deviation.

shows two distinct distributions for the two vowels, as we would expect.

Although the relative distributions of F1 and F2 for the two vowels clearly indicate that vowel quality cannot be used to distinguish /i/ and /I/ produced by some of the speakers, the F1/F2 plots cannot be used exclusively to measure whether the subjects have mastered the production of A.E. /i/ and /I/. Instead, the extent to which the productions are intelligible to native speakers of A.E. is a better gage. Figure 7 shows the average intelligibility score for each subject as well as the intelligibility of each of the two vowels. The correlation between the ratio of each of the acoustic characteristics (duration, F1 and F2) and overall intelligibility was measured:

- Duration and overall intelligibility: $r = 0.158$
- F1 and overall intelligibility: $r = 0.429$
- F2 and overall intelligibility: $r = 0.775$

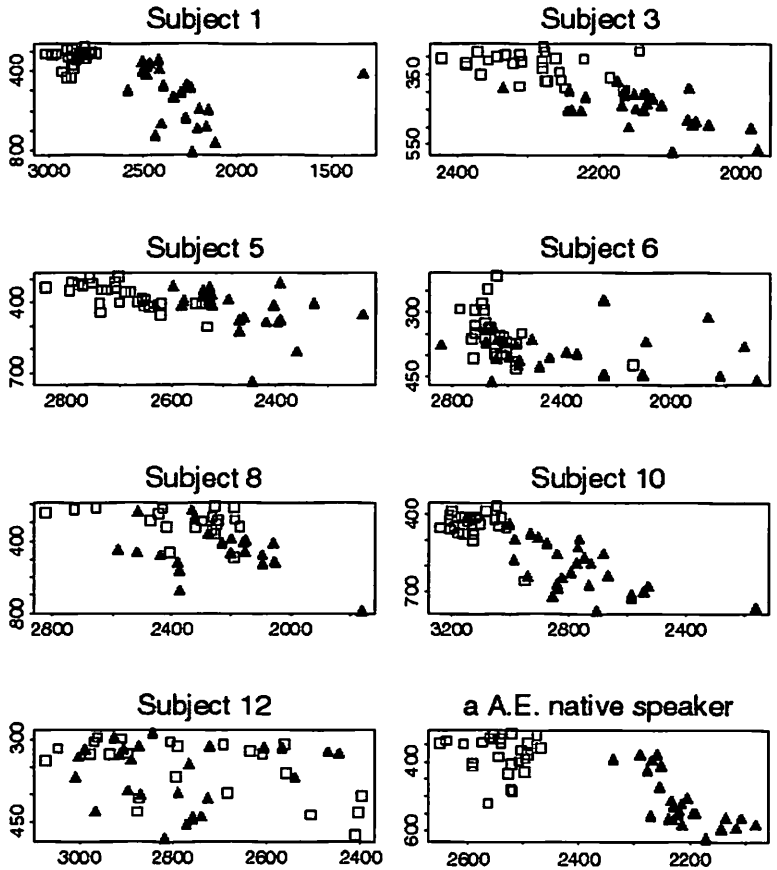

Figure 6. Vowel quality as measured by F1 (y axis) and F2 (x axis) for each of the Japanese subjects and one native speaker of English. Open squares: /i/; Solid triangles: /I/.

As expected, the ratio of average duration and overall intelligibility are not closely correlated, but the ratio of the formants (and in particular the ratio of average F2's) have much higher correlation factors. This confirms that the subjects who make the most distinction in vowel quality are in general the most intelligible.

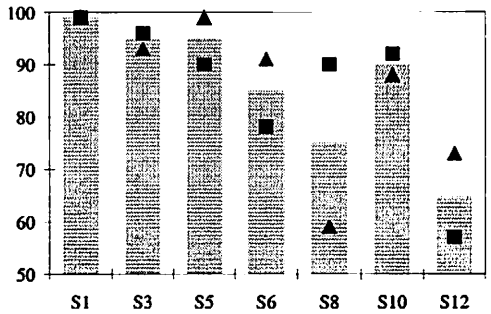

Figure 7. Overall intelligibility (column), intelligibility of /i/ (triangles), intelligibility of /I/ (squares).

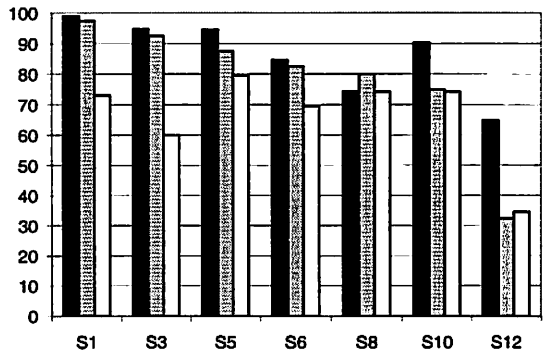

Figure 8. Average intelligibility (black), perceptual performance on natural tokens only (gray), and overall perceptual performance (white).

5 Discussion

In both perception and production, the J subjects consistently distinguish /i/ and /I/ according to duration, but the extent of vowel quality distinction varies by subject. Despite varying levels of vowel quality use, nearly all of the subjects performed above chance on the perception of uniform tokens and made a *statistically significant* difference in the mean values of the formants, although the distributions sometimes overlapped. Therefore, the subjects have begun to learn to use vowel quality to some extent, which is particularly interesting since several of the subjects reported that they were unaware of any vowel quality distinction between the two vowels.

Figure 8 shows a comparison of overall intelligibility and performance on the perceptual portion of the study for each of the subjects. In addition,

the correlation was calculated between the overall intelligibility and the perceptual performance on three different types of tokens:

- Perception of natural tokens and overall intelligibility: $r = 0.863$
- Perception of uniform tokens and overall intelligibility: $r = 0.422$
- Overall perception and overall intelligibility: $r = 0.667$

Thus, the subjects' perceptual performance on just natural tokens is highly correlated with overall intelligibility. This result is consistent with Flege's SLM which hypothesizes that improved production should follow improved perception; in general, the subjects who perceived the distinction the best were also able to produce it most intelligibly.

However, a comparison of the correlations across the three token types is somewhat surprising. Since the SLM predicts that native-like production requires native-like perception, we would expect that the subjects who are best able to use the proper acoustic cues in production must also be able to use the same cues in perception. As shown above, intelligibility is more related to vowel quality than to vowel duration distinction, so those subjects who are the most intelligible (i.e. use vowel quality effectively in production) should be able to use vowel quality the most effectively in perception. Following this line of reasoning, overall intelligibility should be highly correlated with the perception of the *uniform* tokens, since these tokens have only a vowel quality distinction. Surprisingly, of the three token types considered, the perception of uniform tokens is the weakest predictor of intelligibility.

Although our results appear to be inconsistent with the SLM's prediction that the use of vowel quality in perception should develop before the use of a vowel quality distinction in production, this is not necessarily the case. In the present study, we focused on including a great deal of variation in phonetic context and speaker to ensure accurate measurements of the subjects' overall use of cues. However, to keep the perceptual test to a reasonable length, this choice limited the number of different duration types that could be included, and our stimuli sets only contained tokens with either (1) natural duration cues, (2) no duration cues, (3) exaggerated duration cues, or (4) the "wrong" duration cues. As we saw in the production portion of the study, J subjects' productions include a complex relationship between the use of duration and the use of vowel quality. A perceptual study similar to the present one with more gradation in duration cues, particularly between the *natural* token types and the *uniform* token types, may reveal that a similarly complex relationship in perception exists. Such a study may also show that the most intelligible speakers tend to use vowel quality to a greater degree in perception than less intelligible speakers do, even though they still use duration cues as well.

References

Best, Catherine T. 1995. A direct realist view of cross-language speech perception. In *Speech Perception and Linguistic Experience: Issues in Cross-Language Research* (ed. W. Strange), 171-204, Timonium, MD: York Press.

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

Flege, James E. 1995. Second language speech learning: Theory, findings, and problems. In *Speech Perception and Linguistic Experience: Issues in Cross-Language Research* (ed. W. Strange), 233-277, Timonium, MD: York Press.

Jones, Daniel. 1962. *An outline of English phonetics*, 9th ed. Cambridge: Heffer.

Klatt, Dennis. 1976. Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America*, 59, 1208-1221.

Peterson, Gordon. E. and Lehiste, Ilse. 1960. Duration of syllable nuclei in English. *Journal of the Acoustical Society of America* 32, 693-703.

Shibatani, Masayoshi. 1990. *The Languages of Japan.* Cambridge, UK: Cambridge University Press.

Moulines, Eric and Charpentier, Francis. 1990. Pitch-synchonous waveform processing techniques for Text-to-Speech synthesis using diphones. *Speech Communication* 9, 453-467.

Strange, Winifred, Akahane-Yamada, Reiko, Fitzgerald, Brett, and Kubo, Rieko. 1996a. Perceptual assimilation of American English vowels by Japanese listeners. In *Proceedings of the Fourth International Conference on Spoken Language Processing,* 2458-2461. Philadelphia, Pennsylvania.

Strange, Winifred, Bohn, Ocke-Schwen, Trent, Sonja, McNair, Melissa, and Bielec, Katherine. 1996b. Context and speaker effects in the perceptual assimilation of German vowels by American Listeners. In *Proceedings of the Fourth International Conference on Spoken Language Processing,* 2462-2465. Philadelphia, Pennsylvania.

Strange, Winifred, Akahane-Yamada, Reiko, Kubo, Rieko, Trent, Sonja, Nishi, Kanae, and Jenkins, James. 1998. Perceptual assimilation of American English vowels by Japanese listeners. *Journal of Phonetics* 26, 311-344.

Michelle A. Minnick Fox

Department of Linguistics
619 Williams Hall
University of Pennsylvania
Philadelphia, PA 19104
minnick@unagi.cis.upenn.edu

Kazuaki Maeda
Dept. of Computer and Information Science
Phonetics Lab, Dept. of Linguistics
619 Williams Hall
University of Pennsylvania
Philadelphia, PA 19104
maeda@unagi.cis.upenn.edu