

# ACOUSTIC ANALYSIS OF JAPANESE-ACCENTED VOWELS IN ENGLISH

Kimiko Tsukada

Speech, Hearing and Language Research Centre  
Macquarie University

## ABSTRACT

In an attempt to characterize Japanese-accented English, vowels in monosyllabic English words produced by 12 Australian talkers and 24 Japanese talkers were analyzed acoustically. The results show clear temporal and spectral differences between the two groups which may be perceived as a "foreign accent" in English produced by the Japanese talkers. These differences are in agreement with a notion of intermediate nature of non-native speech production.

## INTRODUCTION

Phonetic characteristics of vowels are influenced by various factors. In English, in particular, and in many other languages, vowel duration is found to vary as a function of the voicing of the following consonant (Chen, 1970). However, in languages such as Arabic, Polish and Czech, such voicing effects are shown to be absent or negligible (Keating, 1985, Mitleb, 1984). Many studies on cross-language production have demonstrated smaller voicing effects by non-native speakers of English (Crowther and Mann, 1992, Elsendoorn, 1985, Flege, 1993, Mack, 1982, Munro, 1993). Such consistent differences between native and non-native speakers of a language could contribute to the perception of "foreign accent" by native listeners.

Although modelling "foreign accent" is problematic, it is generally accepted that non-native adult learners of a second/foreign language attain limited mastery when it comes to pronunciation and, as a result, their speech is characterized by varying degrees of "foreign accent". There is some evidence that linguistically untrained listeners are capable of detecting small distortions of the phonetic norms (Flege, 1984). The aims of this study are 1) to investigate Japanese and English vowels produced in comparable contexts by native speakers of each language and compare them to the interlanguage data, i.e., the same English vowels produced by the Japanese speakers, and 2) to search acoustic cues which native English listeners may use in their judgements on non-native speakers' accentedness. These cues, if established, have important implications from the point of view of second/foreign language acquisition.

## METHOD

### Subjects

This study includes 12 talkers (6 male, 6 female) of Australian English (AE) and 25 talkers (11 male, 14 female) of Japanese (J). The AE group only produced the English materials (henceforth the AE-corpus) and with the exception of one speaker, the J group produced both the English materials (the JE-corpus) and the Japanese materials (the J-corpus). The majority of the Japanese speakers in the present study have either completed university education and are continuing their postgraduate studies or are currently enrolled in their undergraduate degree studies at Macquarie University. Every Japanese participant except for one speaker has studied English for a minimum of 6 years during their secondary and tertiary education. Their length of stay in Australia or any other English-speaking country (i.e., New Zealand for this group of speakers) ranged from 3 weeks to 16 years.

### Materials

In this study, 2 databases were constructed. The first included English materials consisting of 8 vowel types in CVt/CVd words (1. /i/ in beat/bead, 2. /ɪ/ in bit/bid, 3. /E/ in bet/bed, 4. /A/ in bat/bad, 5. /O/ in cot/cod, 6. /U/ in put/good, 7. /ʌ/ in but/bud, 8. /u:/ in suit/sued). The second were the Japanese materials consisting of 5 vowel types in CVto/CVdo words (1. short /i/ in kito/kido, long /i/ in shiito/shido, 2. short /e/ in meto/medo, 3. short /a/ in kato/kado, long /a/ in kaato/kaado, 4. short /o/ in noto/nodo, 5. short /u/ in

suto/tsudo, long /u/ in ruuto/fuudo). As in English, these are real words commonly used in Japanese. A typical Japanese syllable is CV where C can be zero. Bisyllabic words ending with '-to/da' were used in this database, since consonants (except for two morae, i.e., /N/ and /Q/) do not appear in the final position. A phenomenon of vowel devoicing or deletion has been widely reported in the literature. High vowels are particularly affected when they are surrounded by voiceless consonants, resulting in fewer tokens for the words 'kito' and 'suto'. Japanese differs from English in that segmental duration is phonemically contrastive in the former language. Only 3 long vowels were included in this study, although each short vowel has a corresponding long vowel. The accent pattern for the test words is HL (high - low) in the standard Tokyo dialect. The majority of the speakers used this intended HL pattern.

The speech data were recorded in a sound-treated studio and digitized at 20 kHz on SUN workstations at the Speech Hearing and Language Research Centre, Macquarie University. Subjects read each word as it was presented 7 times in a randomized order on the computer screen. Japanese orthography was used to transcribe the Japanese test words. Prior to recording, the word lists were shown to the JE talkers so that they could resolve any uncertainty about pronunciation of the English test words. Mispronounced tokens were eliminated. The speech signal processing package waves+ was used for data preparation, i.e., segmentation and labelling. Speech samples were phonetically labelled by hand using a combination of waveform and spectrographic cues. Labelling criteria follow those set for the ANDOSL database (Croot et al., 1992). The mu+ speech database analysis system was then used for acoustic analyses and statistical tests. Two kinds of parameters were investigated: the acoustic duration of the vowel preceding the voiced and voiceless consonants; and the first two formant frequencies (F1, F2) averaged across the vowel per token. The focus of the study is on comparing the non-native productions (i.e., the JE-corpus) with the two native productions of English (AE-corpus) and of Japanese (J-corpus).

RESULTS

Vowel duration

1) English database

Vowel durations by AE and JE talkers are given in Figure 1. In both AE- and JE-corpora, speakers produced significantly longer vowels before /d/ than before /t/ (AE:  $t=-28.4$ ,  $df=1339$ ,  $p=0$ , JE:  $t=-17.0$ ,  $df=2461$ ,  $p=0$ ). However, as in many previous studies (Chen, 1970, Homma, 1973, 1981, Elsendoorn, 1985, Munro, 1993, Flege, 1993), the durational differences between vowels preceding voiced and voiceless consonants were larger in the AE-corpus (mean ratio: 1 to 1.59) than in the JE-corpus (1 to 1.25) except for /U/ and /V/ where not much difference was observed between the two groups.

When vowels were followed by a voiceless consonant, the mean vowel durations for JE talkers were longer than for AE talkers and the reverse was the case when vowels were followed by a voiced consonant, resulting in the small context effect for the non-native group. It was also evident from AE-, but not JE-, data that long vowels are more strongly influenced by the voicing status of the following consonant. In all vowel categories except for /i:/ in the AE-corpus, speakers were more variable before /d/, and the JE group was consistently more variable than the AE group.

The mean duration of vowels in 56 /t/-final words (8 vowels x 7 tokens) spoken by each talker was subtracted from the mean duration of vowels in 56 /d/-final words. The average vowel duration differences produced by AE talkers ranged from 55 to 104 ms. Of all JE talkers, 6 (25%) produced vowel duration

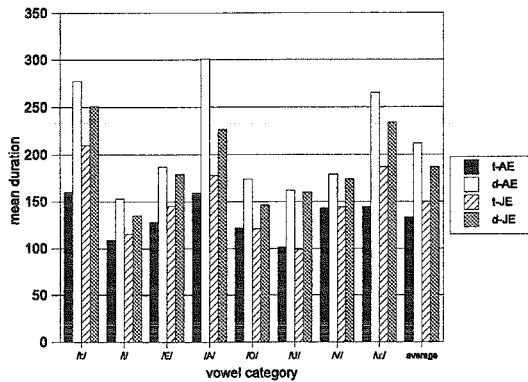


Figure 1. Mean duration (in ms) of English vowels (t-AE: vowel duration before /t/ by AE group, d-AE: vowel duration before /d/ by AE group, t-JE: vowel duration before /t/ by JE group, d-JE: vowel duration before /d/ by JE group).

differences that fell within the AE range. These 6 non-native speakers produced much smaller vowel-lengthening differences in their native Japanese (-7 to 22 ms). Therefore, as far as the voicing effects of the following consonants are concerned, JE talkers successfully suppressed their L1 durational characteristics in their productions of English vowels.

Eisendoorn (1985) found that phonemically short vowels (such as /i/ and /u/) preceding voiced consonants have similar durations to those found for long vowels preceding voiceless obstruents in the native production. This trend was also true in the present study, in which /i/ in 'bid' (153 ms) was not much different from /i:/ in 'beat' (162 ms) and /u/ in 'good' (162 ms) was longer than /u:/ in 'suit' (144 ms) when produced by AE talkers (see Figures 1 and 3). In the present study, the length contrast seems somewhat suppressed before /t/ in the AE-corpus. In the JE production, on the other hand, the length contrast is maintained regardless of the following consonant, and before a voiceless consonant it is more pronounced (difference in vowel duration between 'beat' and 'bit': 83 ms) than in the AE data (49 ms). To native listeners, 'beat' produced by JE talkers may sound too long. This possibility was suggested in Takebayashi (1970).

## 2) Japanese database

The vowels produced by J talkers in CVto/CVdo are shown in Figure 2. In the J-corpus, too, a voicing effect was observed consistently ( $t=-2.6$ ,  $df=2592$ ,  $p=0.0095$ ), but to a much less degree compared to the AE database (J ratio: 1 to 1.06 vs AE ratio: 1 to 1.59). The mean vowel-lengthening differences before voiced and voiceless consonants ranged from -13 to 27 ms in Japanese. It may be the case that some of the durational differences are too small to have perceptual relevance. In this respect, it is quite interesting that previous studies (Fujisaki et al., 1975; Klatt, 1975 reviewed in Takahashi, 1987) reported a smaller noticeable difference of duration in Japanese (10% of the original segment duration) than that found in English (20%). In other words, Japanese speakers hear smaller changes in segmental duration than English speakers.

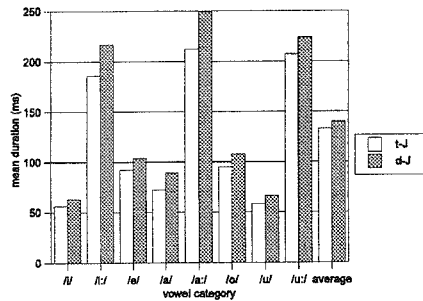


Figure 2. Mean duration (in ms) of Japanese vowels (t-J: vowels before /t/, d-J: vowels before /d/).

As mentioned above, devoicing or deletion of high vowels, i.e., /i/ and /u/, occurs when they are flanked by voiceless obstruents, and fewer tokens were available (113 and 48 for /i/ and /u/, respectively, where the maximum number of token is  $25 \times 7 = 175$ ). This makes it difficult to assess the contextual effect of the following consonant. However, a small voicing effect was present in other non-high vowels which are not subject to devoicing to the same extent as high vowels.

The order of vowel length from shortest to longest was  $i < u < a < e < o$ . Although the exact order obtained in the present study is slightly different from Han (1962) and Homma (1973), in which the order of  $u < i < o < e < a$  and  $u < i < a < e < o$  were reported, respectively, it agrees with previous work in that higher vowels were shorter than lower vowels. Furthermore, in the present study, front vowels were shorter than non-front vowels. According to Homma (1973), in Japanese the manner of articulation and voicing status of preceding consonants influence vowel duration to a larger extent than those of following consonants, and nasal consonants were found to induce lengthening of the following vowel. The slight difference in the order could be due to the fact that in the present study /e/ and /o/ were preceded by /m/ and /n/, respectively.

Length contrasts in each corpus are shown in Figure 3. On average, Japanese long vowels were slightly more than 3 times as long as their short counterparts (216 vs 69 ms). Also, we can see that values for short vowels in the JE-corpus lie between the short and long vowels in the J-corpus, indicating that the

Japanese talkers as a group successfully modified the absolute length of their Japanese vowels to approximate English durations. One explanation could be that the durational differences between the AE short vowels and the J short vowels are perceptually salient enough to inhibit the two being equated and the potential transfer of vowel duration from J to JE is blocked. Similarly, in her review of Flege's study (1979), Schmidt (1988) notes that the native Arabic speakers of English produced vowels that were midway between the vowel durations for the average Arabic long vowel (179 ms) and the average Arabic short vowel (98 ms).

### Vowel quality

#### 1) English database

The first two formant frequencies in AE- and JE-corpora are given in Table 1. These values were tracked at the midpoint of the vowel portion.

	/i:/	/I/	/E/	/A/	/O/	/U/	/V/	/u:/
F1 (AE)	326 382	331 387	446 532	653 832	586 648	352 407	696 853	311 386
F2 (AE)	1980 2535	1985 2505	1825 2316	1636 1976	1019 1114	1000 1121	1266 1490	1622 1971
F1 (JE)	286 320	342 369	547 551	720 902	611 627	396 442	734 905	326 399
F2 (JE)	2246 2807	2130 2679	1895 2356	1664 1815	1083 1166	1448 1621	1446 1620	1583 1798

Table 1. Mean F1 and F2 values (in Hz) for 8 vowel types in English (figures in the upper row are mean formant frequency values of male speakers and those in the lower row indicate the corresponding values by female speakers).

The results of the spectral analyses indicated significantly higher values for F1, F2 or both in most instances in the JE-corporus compared to the AE-corporus. If we compare the above values to those obtained in Hillenbrand et al. (1995), their American English shows higher F1 values than Australian English for vowels in question except for /A/ and /V/. /A/ in the American variety has considerably higher F2 values (male: 1952 Hz, female: 2349 Hz) than both AE and JE counterparts, indicating a more advanced tongue position. For /V/, American and Australian English have similar F1, F2 values to each other while JE has higher values on both F1 and F2. It is plausible that higher formant frequencies (F1 in particular) in the JE-corporus are an influence of American English, but this possibility requires further verification.

A merger of /A/ and /V/ is noticeable in the JE-corporus (middle panels in Figure 4). This is very much expected, since Japanese has only one vowel /a/ in this region (right panels in Figure 4). /U/ posed great difficulty for JE speakers, too, as only 4 out of 24 JE speakers managed to produce both F1 and F2 values that fall within the AE range. It appears that most JE speakers used their Japanese /u/, which is phonetically [u] as described below. Observation of each speaker in the JE-corporus showed considerable inter- and intra-speaker variability. While some JE speakers successfully produced 8 distinct vowel types, their vowel quality was not necessarily AE-like, possibly due to the varieties of English they had been exposed to. It would be interesting to explore how these stable but deviant JE speakers might be perceived by native listeners, because they may be highly intelligible but heavily accented at the same time. The

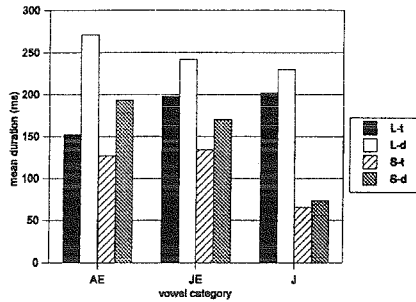


Figure 3. Mean duration (in ms) for long and short vowels in each corpus (L-t: long vowels before /t/, L-d: long vowels before /d/, S-t: short vowels before /t/, S-d: short vowels before /d/). For AE and JE-corpora, long vowels are /i:/ and /u:/ and short vowels are /I/ and /U/. For J-corpora, long vowel are /i, a, u:/ and short vowels are /i, a, u/.

study shows that the vowel space is not utilized as efficiently in the JE-corpus as in the other two corpora: there is a good deal of overlap of vowel ellipses in the formant plane and the ellipses are considerably more spread than in either of the native corpora (Figure 4).

## 2) Japanese database

The first two formant frequencies in the J-corpus are given in Table 2 for general comparison with the AE-corpus.

	long /i/	short /i/	/e/	long /a/	short /a/	/o/	long /u/	short /u/
F1	304 317	292 299	510 536	737 948	714 889	491 579	327 344	342 359
F2	2121 2696	2102 2688	1873 2372	1277 1524	1347 1652	1082 1289	1361 1655	1478 1796

Table 2. Mean F1 and F2 values (in Hz) for 5 vowel types in Japanese (figures in the upper row are mean formant frequency values of male speakers and those in the lower row indicate the corresponding values by female speakers).

Japanese has a much simpler vowel system compared to English. It consists of five vowels /i, e, a, o, u/, each contrasting in length. The phoneme /u/ which is phonetically transcribed as [ɯ] lacks lip-rounding and is characterized by high F2 values. Its tongue position is fairly advanced (aligned with /a/ in F2 values) and the two right panels in Figure 4 suggest that /o/ is the only rounded back vowel in Japanese.

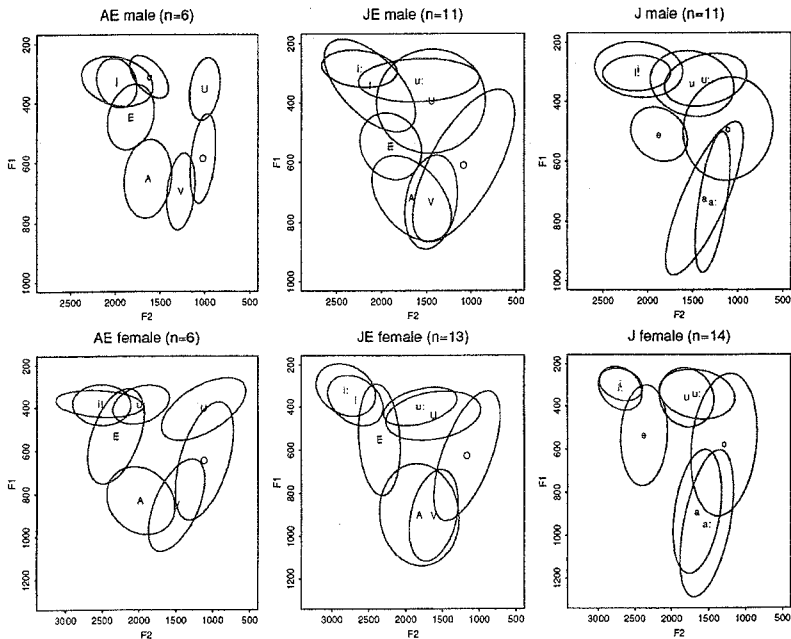


Figure 4. Mean F1 and F2 values (in Hz) for each corpus. The radius of the ellipse is 2.45 times the standard deviation of the mean, covering approximately 95% of the data points.

## CONCLUSION

The present study primarily focused on the interlanguage data collected from Japanese speakers of English with reference to both native Australian English and native Japanese vowels. In summary, as we have seen, the durational differences between vowels preceding voiced and voiceless consonants were largest in the AE-corpus (mean ratio: 1 to 1.59), least in the J-corpus (1 to 1.06) and intermediate between these two in the JE-corpus (1 to 1.25). This suggests a strong L1 influence on the production of English vowel durations by Japanese speakers. If the Japanese transfer this durational characteristics to their production of English vowels, it might be an impediment to their intelligibility (Mack, 1982). Since native English listeners require a durational change of 20% or more before they perceive any difference in length, it is highly likely that lack of sufficient vowel-lengthening differences before voiced and voiceless consonants in the JE production becomes perceptually noticeable. Their failure to execute this phonetic rule in English may be attributed to the lack of knowledge and exposure to authentic spoken English.

As for vowel quality, in addition to having generally higher F1 and F2 values compared to the AE-corpus, JE vowels were characterized by a good deal of clustering on the one hand and scatteredness on the other. This exemplifies an inadequate and inefficient use of the vowel space by the non-native speakers. In general, JE speakers are unstable and their vowels lack a clear target whereas the two native corpora show more clearly separate vowels in the vowel space.

It is necessary to examine the production of individual JE talkers in further detail. In particular, understanding 'advanced' non-native speakers' production is crucial in finding out possible sources of 'foreign accent'. Perception experiments are currently in progress to identify the significance of some of these findings for listeners' perceptions of "foreign accentedness".

## REFERENCES

- Chen, M. (1970). *Vowel length variation as a function of the voicing feature of the following consonant environment*, *Phonetica*, 22, 129-159.
- Croot, K., Fletcher, J., and Harrington, J. (1992). *Levels of segmentation and labelling in the Australian Database of Spoken Language*, Proceedings of the Fourth Australian International Conference on Speech Science and Technology, 86-91.
- Crowther, C., and Mann, V. (1992). *Native language factors affecting use of vocalic cues to final consonant voicing in English*, *J. of Acoust. Soc. Am.* 92, 711-722.
- Eisendoorn, B.A.G. (1985). *Production and perception of Dutch foreign vowel duration in English monosyllabic words*, *Language and Speech*, 28, 231-254.
- Flège, J.E. (1984). *The detection of French accent by American listeners*, *J. of Acoust. Soc. Am.* 76, 672-707.
- Flège, J.E. (1993). *Production and perception of a novel, second-language phonetic contrast*, *J. of Acoust. Soc. Am.* 94, 1589-1608.
- Han, M. (1962). *The feature of duration in Japanese*, *Study of Sounds*, 10, 65-75.
- Harrington, J., Cassidy, S., Fletcher, J., and McVeigh, A. (1993). *The mu+ system for corpus based speech research*, *Computer Speech and Language*, 7, 305-331.
- Hillenbrand, J., Getty, L., Clark, M., and Wheeler, K. (1995). *Acoustic characteristics of American English vowels*, *J. of Acoust. Soc. Am.* 97, 3099-3111.
- Homma, Y. (1973). *An acoustic study of Japanese vowels: Their quality, pitch, amplitude and duration*, *Study of Sounds*, 16, 347-368.
- Keating, P. (1985). *Universal phonetics and the organization of grammars*. In V.A. Fromkin (ed.), *Phonetic Linguistics: Essays in Honor of Peter Ladefoged*. New York: Academic Press, 115-132.
- Mack, M. (1982). *Voicing-dependent vowel duration in English and French: Monolingual and bilingual production*, *J. of Acoust. Soc. of Am.* 71, 173-178.
- Mittleb, F. (1984). *Voicing effect on vowel duration is not an absolute universal*, *Journal of Phonetics*, 12, 23-27.
- Munro, M.J. (1993). *Production of English vowels by native speakers of Arabic: Acoustic measurements and accentedness ratings*, *Language and Speech*, 36, 39-66.
- Schmidt, A.M. (1988). *The Acquisition of Some American English Duration Parameters by Non-native Speakers of English*, Ph.D. dissertation. University of Florida.
- Takahashi, Y. (1987). *Language Universals and Transfer Effects: Acquisition of English vowel durations by Japanese learners*, Ph.D. dissertation. Stanford University.
- Takebayashi, S. (1970). *The vowels of Japanese and English: A contrastive study*, *Lexicon*, 4, 49-67.