Native vs Non-native Production of English Vowels in Spontaneous Speech: An Acoustic Phonetic Study

Kimiko Tsukada

Speech, Hearing and Language Research Centre (SHLRC)
Department of Linguistics
Macquarie University
kimiko@shlrc.mq.edu.au

Abstract

This study aims to examine acoustic characteristics of English vowels produced by 1 Australian English talker and 3 Japanese learners of English in spontaneous speech. Primary stressed vowels in multi-syllabic words were extracted from five 15-minute interview sessions. While there was a considerable overlap between different vowel categories both in native and non-native vowel spaces, centroids were more clearly separated in the former than in the latter. All three Japanese learners' vowel spaces were widely spread in the F2 direction.

The Australian talker showed a moderate spectral separation in two pairs /i - ɪ/ and /a - ʌ/. Although this appears contrary to the spectral overlap commonly reported for these pairs in Australian English, it is consistent with the notion that short vowels are more susceptible to reduction than their long counterparts which are less likely to be undershot in various consonantal contexts.

1. Introduction

In speaking spontaneously, speakers, both native and non-native, are unlikely to monitor their speech as closely as when they read aloud. Second or foreign language learners who are less skilled or experienced in the target language than native talkers may not be able to maximally utilize all the linguistic and non-linguistic resources available to them. In order to deal with the immediate demands of the interaction, they may pay attention to only salient features.

It is thus intuitively plausible to assume that the vowel production of non-native learners diverge to a greater extent from that of native talkers in spontaneous speech than in citation-form or read speech. Indeed, a recent perception study [9], in which the production of English vowels by Italian talkers in conversational speech was examined, reported results that would reinforce such an assumption. Another study [10] reported a higher rate of flap substitution (i.e., using L1 flap sound) for English liquids /l, r/ in spontaneous than in reading task by Japanese learners.

As was pointed out in previous work [9], studying conversational speech poses a challenge in terms of experimental control. However, it is perhaps a valid assumption that second/foreign language learners' ultimate goal would be to attain communicative competence in the target language. In other words, we would expect them to be more concerned about how they sound when their speech is unrehearsed as in spontaneous conversation than in other situations. In this study, an attempt was made to provide quantitative data on English vowels produced by 1 Australian English (AE) talker and 3 Japanese learners of English (JE) in interview sessions.

2. Method

2.1. Talkers and speech materials

Speech data were collected from interviews recorded on several occasions in 1982 between JE talkers and native Australian English talkers. The interviews were tape-recorded in a quiet room at Curtin University of Technology (Perth, Western Australia). Data from 3 male Japanese learners (JE1, JE2, JE3) and 1 male AE talker (AE1) were analyzed in this study. JE talkers were all ELICOS (English Language Intensive Courses for Overseas Students) students. None of them had started their tertiary education yet at the time of recording. Two of them (JE1 and JE3) came to Australia upon completing their senior high school. JE1 was from Yokohama, south-west of Tokyo. JE2 was from Wakayama near Osaka and he came to Perth before finishing high school. In the Japanese education system, students gain 6 years (3 years each in junior and senior high schools) of English language learning by the time they graduate from high school. JE1's length of stay in Perth was unknown, but presumably less than 6 months. JE2 and JE3 had lived in Perth for about 5 and 8 months, respectively, when they were recorded. Interview topics were not specified, but, in general, they revolved around the interviewees' English language study at the university, their future plans, and so forth. Primary stressed vowels in multi-syllabic words such as *friendly*, *holiday* and *fishing* were examined. There were often multiple tokens of the same words in the interview.

2.2. Data processing and analysis

The speech data were digitized at 20 kHz in Speech, Hearing and Language Research Centre (SHLRC) at Macquarie University. Segmentation of the speech data were carried out using a signal processing package xwaves+ and each utterance was phonetically labelled on the basis of waveform and spectrographic information available in EMU speech database system [1]. Labelling criteria described in detail in [3] were followed. Table 1 shows the number of vowel tokens each talker produced. Statistical packages Splus and SPSS were used for analysis and graphic display. For each vowel, a series of one-way ANOVAs with Bonferroni adjustments were conducted to see if the talker identity had any effects on F1, F2 and vowel duration. The significance level of .01 was adopted. In particular, the question of interest was whether the AE talker is significantly different from all of the JE talkers on a given acoustic parameter on the

Table 1: *Number of vowel tokens per talker. No /a/ tokens were available for JE3.*

	vowel type									
talker	ɪ	ɛ	æ	ɒ	ʌ	i	a	ɔ	u	ɜ
JE1	27	53	31	14	20	24	5	18	13	9
JE2	29	35	14	15	30	9	15	13	18	12
JE3	29	27	3	16	12	15	–	9	5	2
AE1	48	56	37	27	32	34	8	7	11	7

one hand and whether JE talkers significantly differ from one another on the other hand. A decision was made to use talker identity rather than talker group as a factor, as the number of talkers was too small to be representative of either native or non-native group. In addition, Bayesian classification experiments were conducted to examine how JE vowel categories are separated in the acoustic vowel space defined by the AE talker. Methods of classification experiments were as described in previous studies [2, 6].

3. Results

3.1. Vowel quality

3.1.1. Acoustic analysis

Figure 1 shows the acoustic vowel space of each talker on the F1/F2 plane. Formant values were obtained at the vowel's acoustic midpoint. It is clear that all JE talkers' vowel spaces differ in many respects from that of the AE talker. At the same time, they differ from one another considerably, reinforcing the notion that non-native talkers show a great degree of *inter*-talker variability. In addition, there is substantial *intra*-talker variability. As for the AE talker, there is a good resemblance between his vowel space and that of other Australian talkers in comparable contexts [5]. While there was a considerable overlap between different vowel categories both in AE and JE vowel spaces, centroids were more clearly separated in AE than in JE data. All three JE talkers' vowel spaces were widely spread in the F2 direction. This observation agrees with a suggestion that "backness seemed to be the reason for the L2 accent instead of vowel height" [8].

The /æ/ vowel was as fronted as /ɛ/ in the AE vowel space whereas the same vowel in JE data was much more retracted in relation to /ɛ/ (450 Hz on average) and merging with vowel types such as /ʌ/, /ɜ/ and /a/. This merger was observed in citation-form speech [11] and reflects a strong influence from their Japanese vowel space in which only one vowel /a/ occurs in the low central region. /i/ and /ɪ/ were more fronted in JE than in AE and so was /u/ for two of the JE talkers (JE1, JE3). Previous studies showed that the fronted or centralized /u/ was a clear marker of AE, keeping it distinct from other varieties of English [6, 7]. /u/ by the two JE talkers in this study was even more fronted than that of the AE talker by approximately 250 Hz. Back vowels such as /ɒ/ and /ɔ/ showed a considerable F2 variability in JE speech. All JE talkers produced the /ɔ/ vowel which was variable primarily in F2. In contrast, AE /ɔ/ showed a large variability in both F1 and F2. AE /ɒ/ had significantly higher F1 values (573 Hz) than that of JE /ɒ/ (493 Hz), indicating that the former was produced with a lower tongue position than the latter. JE1 and JE2 produced /ɔ/ and /ɒ/ which hardly displayed any spectral differences and, as is presented in Table 3, it appears that they relied solely on vowel duration as a

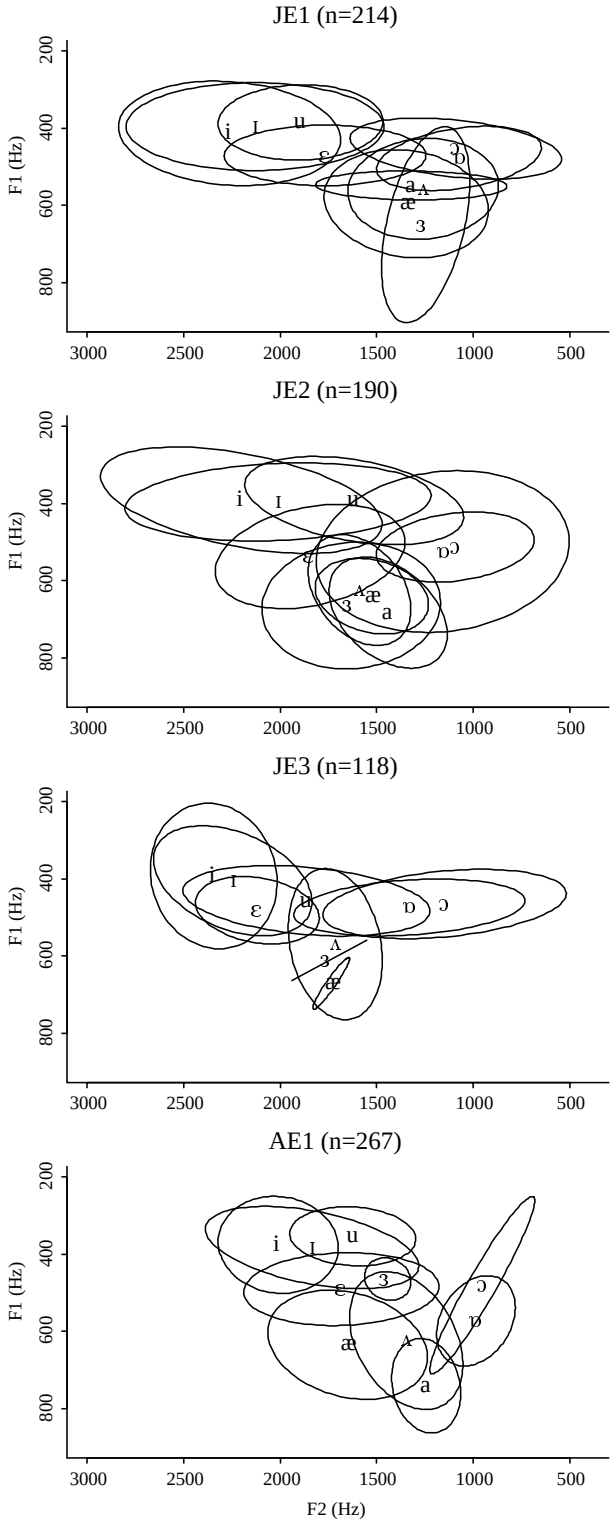

Figure 1: *F1 and F2 values (in Hz) for 10 vowel types per talker. For data having a Gaussian (normal) distribution, the radius of the ellipse is 2.45 times the standard deviation of the mean, covering approximately 95% of the data points. The centroids are the average values of those distributions.*

Table 2: *Results of ANOVA: vowel quality*

Vowel	Formant	*df*	*F*-ratio	*P*-value
ɪ	F2	3, 129	18.4	.0000
ɛ	F1	3, 167	20.2	.0000
	F2	3, 167	30.7	.0000
æ	F1	3, 81	4.4	.0067
	F2	3, 81	21.6	.0000
ɒ	F1	3, 68	15.7	.0000
	F2	3, 68	11.4	.0000
ʌ	F1	3, 90	6.4	.0005
	F2	3, 90	53.9	.0000
i	F2	3, 78	14.5	.0000
a	F1	2, 25	21.5	.0000
	F2	2, 25	6.4	.0058
u	F1	3, 43	7.4	.0004
	F2	3, 48	6.6	.0009
ɜ	F1	3, 26	11.9	.0000
	F2	3, 26	14.4	.0000

cue to distinguish these two categories. JE and AE /ɜ/ differed greatly from each other in F1 values. JE talkers produced it as a low vowel in their vowel space. Although, acoustically, /ɜ/ may be categorized as a 'new' vowel [4], the results would be more interpretable if we assumed that it is perceptually assimilated to their Japanese /a/.

The AE talker showed a moderate spectral distinction in two pairs /i - ɪ/ and /a - ʌ/. This may appear contrary to a spectral overlap commonly reported for these pairs in Australian English. However, it is consistent with the notion that short vowels are more susceptible to reduction than their long counterparts.

3.1.2. Statistical analysis

Table 2 presents the results of the ANOVAs for vowel quality. All vowel categories with the exception of /ɔ/ showed significant overall Talker effects for F2. For F1 values, the effect was observed for all vowels except for /ɪ/, /i/ and /ɔ/. In fact, /ɔ/ was the only vowel for which neither vowel quality nor duration significantly differed according to the talker.

AE1 had significantly lower F2 for /ɪ/ than all JE talkers. Among the JE group, the difference between JE2 and JE3 reached significance. The /ɛ/ vowel showed significant Talker effects for both F1 and F2. JE2 had higher F1 than the rest of the talkers who did not differ from one another. For F2, JE3 had significantly higher values than the other three talkers. The difference between JE2 and AE1 also reached significance with the former having higher F2 values. For /æ/, the only significant F1 difference was observed between AE1 and JE1 with the native talker showing higher values. For F2, JE1 was separated from the other 3 talkers for having significantly lower values. /ɒ/ showed strong Talker effects for both F1 and F2. AE1 had significantly higher F1 values than all JE talkers who did not differ from one another. Thus, the two talker groups were clearly divided. As for F2, all JE talkers had higher values than AE1, and JE2 and JE3 significantly so. The difference between JE3 (highest F2) and JE1 also reached significance. For /ʌ/, both AE1 and JE2, who did not differ from each other, had significantly higher F1 than JE1. No other contrast reached significance. JE3 had significantly higher F2 than all the other talkers. JE2 had higher F2 than JE1 and AE1 who did not differ from each other. The /i/ vowel showed significant Talker effects only for F2. The two talker groups were divided, with

Table 3: *The mean vowel duration (in ms) per talker. Standard deviations are in parentheses.*

	talker			
vowel	JE1	JE2	JE3	AE1
ɪ	64 (18)	72 (50)	53 (15)	47 (12)
ɛ	74 (20)	86 (26)	86 (35)	56 (26)
æ	88 (20)	102 (43)	109 (7)	88 (29)
ɒ	64 (22)	70 (26)	62 (17)	65 (14)
ʌ	82 (22)	78 (19)	73 (17)	59 (12)
i	102 (39)	88 (24)	95 (34)	95 (35)
a	118 (66)	113 (33)	— (—)	124 (23)
ɔ	115 (45)	112 (36)	120 (32)	108 (35)
u	76 (20)	78 (33)	71 (22)	74 (35)
ɜ	116 (35)	119 (17)	201 (100)	85 (16)

Table 4: *Results of ANOVA: vowel duration*

Vowel	*df*	*F*-ratio	*P*-value
ɪ	3, 129	5.8	.0009
ɛ	3, 167	13.7	.0000
ʌ	3, 90	9.5	.0000
ɜ	3, 26	7.4	.0009

AE1 showing significantly lower values than all three JE talkers. For /a/, JE1 had significantly lower F1 values than both JE2 and AE1 who did not differ from each other. As for F2, the only significant difference was found between JE2 and AE1, with the former showing higher values. For the /u/ vowel, JE3 had significantly higher F1 than the other 3 talkers who did not differ from one another. JE1 had significantly higher F2 values than JE2 and AE1, but not JE3. /ɜ/ showed no significant F1 difference among the JE talkers. JE1 and JE2, but not JE3, had significantly higher values than AE1. As for F2, both JE2 and JE3 had significantly higher values than JE1, but did not differ from each other.

In summary, the native vs non-native division was observed for /ɒ/ whose F1 was lower in JE and for /i/ which, with higher F2, occupied a more peripheral position in the JE vowel space.

3.2. Vowel duration

3.2.1. Acoustic analysis

Table 3 shows the mean vowel duration per talker. Overall, none of the JE vowel types differed appreciably ($<$ 30 ms) in the mean duration from AE vowels with the only exception being /ɜ/ which was clearly longer in non-native than in native speech. Some vowels (/ɪ, ɛ, ʌ, ɜ/) were shorter when produced by the AE talker, but not others. No consistent talker-to-talker variations were observed. There was no evidence that vowel duration was more variable in JE than in AE data.

3.2.2. Statistical analysis

Only 4 vowels /ɪ, ɛ, ʌ, ɜ/ showed significant Talker effects on vowel duration (Table 4). For /ɪ/, JE2 was significantly longer than AE1 and JE3, but not JE1. The /ɛ/ vowel showed a clear division between AE and JE groups with the former significantly shorter than the latter. For /ʌ/, AE1 was significantly shorter than JE1 and JE2, but not JE3. No significant differences were found in the JE group. Finally, JE3 was separated from the other talkers for having much longer /ɜ/ (by at least 80 ms).

Table 5: *Confusion matrix (in %). Figures in **boldface** indicate a match between intended and classified vowels.*

	ɪ	ɛ	æ	ɒ	ʌ	i	a	ɔ	u	ɜ
ɪ	**32**	7	0	0	0	55	0	0	5	1
ɛ	6	**60**	10	0	0	18	0	0	0	6
æ	0	6	**25**	4	54	0	8	2	0	0
ɒ	0	24	4	**27**	4	0	0	29	2	9
ʌ	2	13	52	3	**23**	0	2	5	0	2
i	40	15	0	0	0	**46**	0	0	0	0
a	0	5	35	0	35	0	**25**	0	0	0
ɔ	0	18	0	23	8	0	0	**43**	3	8
u	25	17	0	0	3	22	0	0	**31**	3
ɜ	0	9	48	4	22	0	13	4	0	**0**

Table 6: *Confusion matrix (in %). The model includes durational information. Figures in **boldface** indicate a match between intended and classified vowels.*

	ɪ	ɛ	æ	ɒ	ʌ	i	a	ɔ	u	ɜ
ɪ	**57**	11	0	0	0	28	0	0	4	1
ɛ	10	**56**	14	0	0	17	0	0	0	4
æ	0	6	**44**	13	21	0	17	0	0	0
ɒ	2	31	2	**36**	7	0	0	20	2	0
ʌ	0	18	57	8	**11**	2	2	2	0	2
i	10	8	0	0	0	**81**	0	0	0	3
a	0	5	35	10	15	5	**30**	0	0	0
ɔ	3	23	0	3	3	0	8	**58**	3	3
u	31	17	0	0	3	19	0	0	**28**	3
ɜ	0	4	48	13	0	4	26	4	0	**0**

3.3. Classification

The results of open tests (i.e., trained on the AE data and tested on the JE data) are shown in Table 5. The first column gives the vowel that was intended. Each row shows how the vowel was actually classified in the model. The table reveals a general pattern of misclassification, that is, which JE vowels were confused in the acoustic space. The model is based on the first two formant frequencies tracked at the vowels' acoustic midpoint. All three JE talkers' vowels were combined in the testing data. Overall, the JE vowels were correctly classified only 36%. The only vowel which was classified more than 50% was /ɛ/. Vowels such as /æ, ʌ, a, ɜ/ were very poorly classified, as is seen in Figure 1. When durational information was included in the model (Table 6), the overall classification score increased to 45%. The addition clearly helped to separate /i/ and /ɪ/. The scores for /ɛ/ and /ʌ/ deteriorated possibly due to the fact that these vowels were significantly shorter in AE than JE speech. It is noted that /ɜ/ was never correctly classified with or without durational information and was most frequently misclassified as /æ/.

4. Discussion and conclusion

We observed sufficient evidence that JE talkers' vowels are acoustically very different from those of the AE talker. Talker effects were significant for most vowels on one or more acoustic parameters with F2 being the most robust of the three. For /ɛ, ʌ, ɜ/, the effects were significant on all three (F1, F2, duration).

With respect to the global characteristics, we saw a reasonable spectral distinction in the AE vowel space even in spontaneous speech, but not in the JE counterpart. There was also a suggestion in AE that short members of the vowel pairs (/ɪ/, /ʌ/) may be reduced/centralized to a greater extent than the long members (/i/, /a/). Results from acoustic and statistical analyses suggest that F2 may be a particularly useful cue in studying native vs non-native differences in vowel production. Durational measurements were less clearly differentiated between the two talker groups and may be of secondary importance compared to spectral cues. Needless to say, these findings must be interpreted with caution, as the number of talkers was very small.

Although the objective of this research was to present experimental data by AE and JE talkers in spontaneous speech, it is necessary in future work to include phonetically balanced materials produced in continuous speech. From the viewpoint of interlanguage phonetics, it would be interesting to compare how various phonetic factors (i.e., voicing, place and manner of articulation of neighbouring sounds) are realized in native and non-native speech. The inclusion of citation-form speech by the same talkers is essential also for gaining insights into the issue raised in Introduction which is whether or not the difference in speech style affects the extent to which JE talkers approximate to or diverge from the phonetic norms set by AE talkers.

5. References

[1] Cassidy, S. and Harrington, J., "EMU: An enhanced hierarchical speech data management system", Proc. 6th Aus. Int. Conf. on Sp. Sci. and Tech. (SST-96): 361–366, 1996.

[2] Cassidy, S. and Watson, C., "Dynamic features in children's vowels", Proc. 5th ICSLP, 3: 959–962, 1998.

[3] Croot, K., Fletcher, J. and Harrington, J., "Levels of segmentation and labelling in the Australian Database of Spoken Language", Proc. 4th Aus. Int. Conf. on Sp. Sci. and Tech. (SST-92): 86–91, 1992.

[4] Flege, J.E., "The production of "new" and "similar" phones in a foreign language: Evidence for the effect of equivalence classification", J. of Phonetics, 15(1): 47–65, 1987.

[5] Harrington, J. and Cassidy, S., Techniques in Speech Acoustics, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1999.

[6] Harrington, J. and Cassidy, S., "Dynamic and target theories of vowel classification: Evidence from monophthongs and diphthongs in Australian English", Language and Speech, 37(4): 357–373, 1994.

[7] Ingram, J. and Pittam, J., "Auditory and acoustic correlates of perceived accent change: Vietnamese schoolchildren acquiring Australian English", J. of Phonetics, 15(2): 127–143, 1987.

[8] Iwasaki, Y.E., "Japanese speakers and American English vowels: Spectrographic analysis of learners at three different stages of acquisition", The ORTESOL J., 20: 21–50, 1999.

[9] Piske, T., Flege, J.E., MacKay, I.R.A. and Meador, D., "Non-natives' production of vowels in conversational speech", Collected Paper from the Joint Meeting "Berlin 99", 1999.

[10] Riney, T., Takada, M. and Ota, M., "Segmentals and global foreign accent: The Japanese flap in EFL", TESOL Quarterly, 34(4): 711–737, 2000.

[11] Tsukada, K., "An Acoustic Phonetic Analysis of Japanese-accented English", Unpublished doctoral dissertation, Macquarie University, 1999.