SOME ACOUSTIC CHARACTERISTICS OF AUSTRALIAN ENGLISH /aɪ/ AND JAPANESE /ai/ IN NATIVE AND NON-NATIVE SPEECH PRODUCTION

Kimiko Tsukada
Curtin University of Technology
School of Languages and Intercultural Education

tsukadak@spectrum.curtin.edu.au

ABSTRACT: An acoustic comparison was made between Australian English /aɪ/ and Japanese /ai/ produced by native and non-native talkers. Vowel duration and formant frequencies (F1, F2) at the two targets were measured with a view to characterizing cross-linguistic similarities and differences and to examining the non-native production of these sounds in comparison with that of native talkers. There was an interesting finding that, for spectral characteristics, Australian learners of Japanese approximated to the phonetic norms of Japanese to a greater extent than Japanese learners of English did to the Australian English norms.

INTRODUCTION

In studying the acquisition of speech sounds by adult learners, it is important to understand the relationship between the first language (L1) and target language (TL) sound systems. If a TL sound is equated with an L1 sound which is similar, but not identical to it via the process of 'equivalence classification' (Flege, 1987), then the lack of separate categories for L1 and TL sounds may result in foreign-accented speech. Conversely, if the learner succeeds in forming distinct categories for similar L1 and TL sounds, then the same learner may eventually produce the target sound without a detectable foreign accent.

In this study, we look at two similar sounds which are likely to be considered 'equivalent': /aɪ/ in Australian English (AE) and /ai/ in Japanese (J). While AE /aɪ/ has a phonemic status, the situation is somewhat different in Japanese. Some Japanese phoneticians regard /ai/, /oi/ and /ui/ as being phonetic diphthongs. However, Kasuya and Sato (1990) argue that what is usually termed diphthong in Japanese by analogy from English should be considered a vowel sequence. Likewise, Kubozono (1998) explains that, by introducing the durational concept of mora, both long vowels and diphthongs can be analyzed as a sequence of vowels in Japanese. There is also sufficient evidence for considerable acoustic phonetic differences between English diphthongs and their counterparts in Japanese (Hirasaka and Kamata, 1981; Roberge and Inoue, 1988; Kasuya and Sato, 1990). Nevertheless, it would not be unreasonable to hypothesize that the two sounds in question, AE /aɪ/ and J /ai/, are equated, at least in the first instance, in the learner's interlanguage system.

We aim to examine whether the two non-native groups, Australian learners of Japanese (AJ) and Japanese learners of English (JE), differ in the extent to which they approximate to or diverge from the phonetic norms of the native talkers.

METHODOLOGY

Talkers

L1 data were collected from 9 Australian English talkers (4 male and 5 female) and 9 Japanese talkers (2 male and 7 female). AE talkers' accent type can be classified as being general on the scale of broad to cultivated which is conventionally used in the description of AE accent types (Bernard, 1967; Cox, 1996; Mitchell and Delbridge, 1965). While J talkers were from different parts of Japan, they all read the given speech materials using the standard Tokyo accent pattern. A subset of these talkers produced materials in their foreign language: 6 talkers (2 male and 4 female) for Japanese English and 5 talkers (2 male and 3 female) for Australian Japanese. All AJ talkers were studying Japanese at second or third-year levels at Curtin University of Technology. One of the male AJ talkers used unnaturally fast tempo when he read English materials. His L1 productions were, therefore, excluded from the study. JE talkers were undertaking tertiary education in Australia at the time of recording. Some of the JE talkers

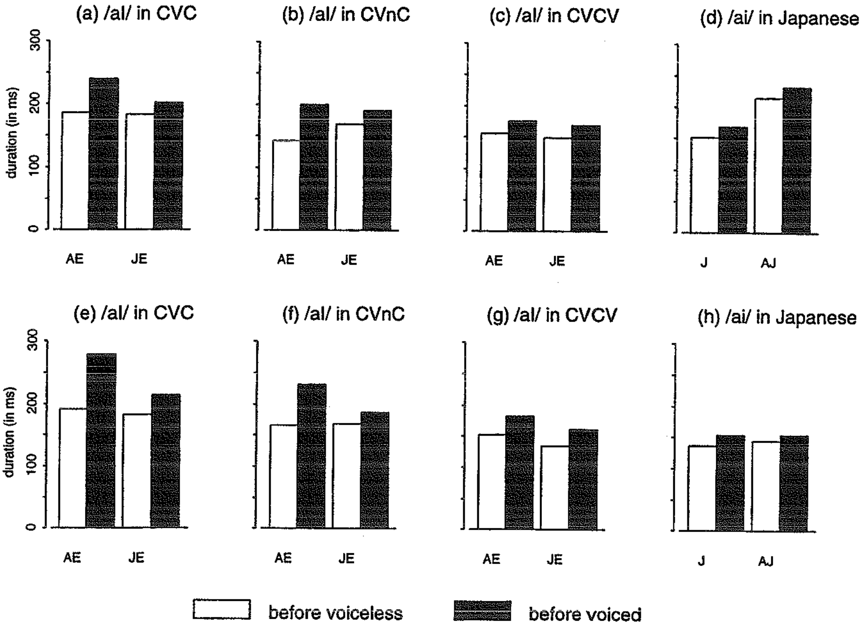

Figure 1: The mean duration (in ms) of /aɪ/ and /ai/ by native and non-native talkers. The top panels show data by female talkers and the bottom panels by male talkers.

who participated in the study at Macquarie University were enrolled in the postgraduate studies while others were qualifying year students. Those who were recruited at Curtin University of Technology were ELICOS (English Language Intensive Courses for Overseas Students) students.

Speech materials

English

The English diphthong /aɪ/ occurred in /CVC/ (n=10), /CVnC/ (n=5) and /CVCə/ (n=9) words which formed near minimal pairs with the second C contrasting in voicing. The contextual consonants included /p, b, t, d, k, g/. Each word was repeated from 2 to 7 times.

Japanese

Japanese /ai/ was placed between the first and the second consonants in 30 multi-syllabic minimal pair words with the second consonant differing in the voicing feature. For the Japanese test words, the contextual consonants were /t, d, k, g, s, z/ and each item was repeated 3 times.

Collection and processing of data

Both English and Japanese test words were embedded in short phrases and sentences and they were recorded in the sound-proofed studios in Speech, Hearing and Language Research Centre (SHLRC) at Macquarie University and Curtin University of Technology. The speech data were digitized at 20 kHz and stored on the SUN workstations in SHLRC. The signal processing package xwaves+ and the EMU speech database system (Cassidy and Harrington, 1996) were used for data processing (segmentation and phonetic labelling) and the acoustic targets (T1, T2) were marked for /aɪ/ and /ai/. Segmentation criteria set out in Croot et al. (1992) were followed using spectrographic and auditory cues.

RESULTS AND DISCUSSION

Temporal characteristics

The barplots in Figure 1 show the mean duration of English /aɪ/ and Japanese /ai/ as a function of the voicing feature of the post-vocalic or word-final consonants averaged across all speakers within each

	Japanese		English		difference (E - J)	
talker	-vd	+vd	-vd	+vd	-vd	+vd
AJ1 (f)	215	231	158	178	-57	-53
AJ2 (f)	210	221	156	163	-54	-58
AJ3 (f)	221	246	152	161	-69	-85
AJ4 (m)	157	172	167	189	10	17
JE1 (f)	172	193	178	197	6	4
JE2 (f)	121	134	139	153	18	19
JE3 (f)	156	167	144	167	-12	0
JE4 (f)	141	164	138	158	-3	-6
JE5 (m)	150	174	158	187	8	13
JE6 (m)	123	135	120	142	-3	7

Table 1: The mean durational difference (in ms) between L1 and target language production. English vowels were taken from /CVCə/ words.

talker group. Panels (a) to (c) and (e) to (g) reveal that the extent of the voicing effect in AE is largely dependent upon the syllable type, i.e. larger in monosyllabic words (/CVC/ and /CVnC/) than in disyllabic words (/CVCə/). A cross-linguistic comparison in comparable contexts (i.e., panels (c), (d), (g) and (h) in Figure 1) shows that while the effect of consonantal voicing appears marginally smaller in Japanese, the mean duration is fairly similar when the two sounds are produced by native talkers (AE and J).

AE and JE

Temporal characteristics of English /aɪ/ by AE and JE talkers have been reported elsewhere (Tsukada, 2000) and are only briefly reviewed here. The difference in the extent of the voicing effect is very clear in /CVC/ and /CVnC/ syllables (panels (a), (e) and panels (b), (f), respectively) in which AE talkers' /aɪ/ tokens are much longer before a voiced than before a voiceless consonant. In /CVCə/ contexts, the talker group difference was considerably reduced. Although this study did not focus on the production of the nasal consonant, it was interesting to observe that not only /aɪ/ but also /n/ lengthened when it was adjacent to a voiced consonant in the AE but not in the JE production.

J and AJ

Both J and AJ talkers showed a very small voicing effect. Unlike female AJ talkers who produced much longer and more variable diphthongs than did the J group, male AJ talkers' temporal patterns were quite native-like both in the mean duration and the magnitude of the voicing effect.

Dissociation from L1 by bilingual talkers

Table 1 shows the mean cross-language durational difference for each bilingual talker. All three female AJ talkers produced longer vowels in Japanese than in English both before voiced and voiceless consonants. This probably reflects their lack of experience in producing Japanese sounds spontaneously. They tended to monitor their speech closely and speak slowly. The remaining talkers' vowel duration differed very little in the two languages. As the cross-linguistic durational difference is very small when the syllable type is controlled, using L1 characteristics is unlikely to result in a negative transfer in this position for either group of non-native talkers. JE learners, however, need specific instructions so that they would not bring their L1 speech habits in producing monosyllabic English words.

Spectral characteristics

Figure 2 shows the ellipse plots for the two targets of AE /aɪ/ and J /ai/. For spectral data, the target sounds in two voicing contexts were combined. Only /CVCə/ words were included for the AE data so that phonetic environments would be comparable in English and Japanese. The leftmost and rightmost panels show clear cross-linguistic differences between AE /aɪ/ and J /ai/ at both targets in that the two targets are more clearly separated in J than in AE.

At T1, both male and female AE talkers had lower F2 values than their J counterparts which indicates a more retracted tongue position for AE. At T2, higher F1 and lower F2 values characterized the AE data. In particular, the difference in F2 values was quite substantial, reaching 616 Hz for male and 609 Hz for female talkers, respectively. These findings are consistent with the study conducted by Roberge and Inoue (1988) and we can confirm that the two targets occupy a more peripheral position in the acoustic

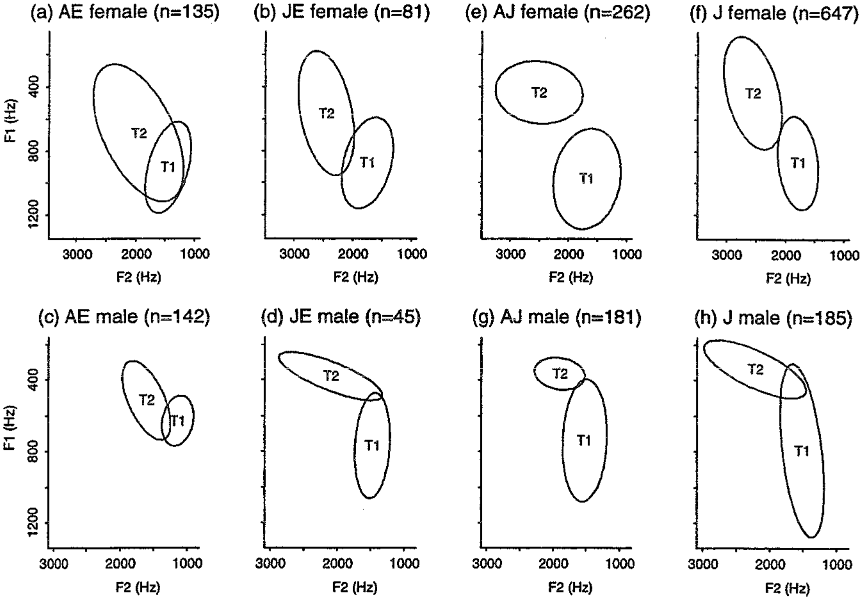

Figure 2: Ellipse plots of the two targets (T1, T2) in Australian English /aɪ/ and Japanese /ai/ by native (AE, J) and non-native (JE, AJ) talkers. For data having a Gaussian (normal) distribution, the radius of the ellipse is 2.45 times the standard deviation of the mean, covering approximately 95% of the data points. The centroids are the average values of those distributions.

vowel space for Japanese /ai/ than for Australian English /aɪ/. We now turn to comparisons between native and non-native data.

AE and JE

Panels (a) to (d) in Figure 2 show the ellipse plots in the F1/F2 plane for the two targets of English /aɪ/: panels (a) and (c) by native and panels (b) and (d) by non-native talkers, respectively. At a first glance, we note a stark contrast between the two talker groups. For both male and female AE talkers, two targets largely overlap in the F1 values with the second target showing a larger variability than the first target. For female AE talkers, T2 almost completely impinges upon T1 (panel (a) in Figure 2). A large variability for T2 is consistent with previous studies (Jha, 1985; Harrington et al., 1997; Harrington and Cassidy, 1999) that noted that the second target of diphthongs is much more variable than the first and often not fully attained. The group data need to be interpreted with caution due to the large inter-speaker variability even for L1 data. In particular, a closer inspection of individual AE talkers' data revealed that female talkers varied considerably from one talker to another while male talkers were more homogeneous.

A large overlap between the two targets we observed for AE data was absent in the JE production. In fact, there was hardly any overlap in male JE talkers' data (panel (d)). For female JE talkers (panel (b)), the overlap was limited to the F1 direction. JE talkers' T1 was not sufficiently retracted (higher F2) and their T2 occupied a more peripheral position (lower F1 and higher F2) in comparison with the AE production. In particular, at T2, JE talkers produced F2 values which are much closer to their L1 counterparts than to the target AE values. The differences between JE and J for the F2 values at T2 were 112 Hz for male and 107 Hz female talkers, respectively. On the other hand, JE differed from AE by as much as 494 Hz (male) and 502 Hz (female) on the same acoustic parameter. JE talkers appear to identify AE /aɪ/ and J /ai/ and this may originate from the way in which diphthongs are transcribed by

		Japanese			English			difference (E - J)	
talker		F1	F2		F1	F2		F1	F2
AJ1 (f)	T1	1000	1818	T1	990	1587	T1	-10	-231
	T2	401	2763	T2	714	2231	T2	313	-532
AJ2 (f)	T1	969	1605	T1	954	1540	T1	-15	-65
	T2	444	2259	T2	705	1778	T2	261	-481
AJ3 (f)	T1	947	1599	T1	941	1460	T1	-6	-139
	T2	460	2489	T2	790	1914	T2	330	-575
AJ4 (m)	T1	784	1543	T1	691	1103	T1	-93	-440
	T2	360	1983	T2	528	1730	T2	168	-253
JE1 (f)	T1	850	1699	T1	838	1632	T1	-12	-67
	T2	464	2452	T2	481	2356	T2	17	-96
JE2 (f)	T1	775	1883	T1	828	1696	T1	53	-187
	T2	543	2439	T2	683	2410	T2	140	-29
JE3 (f)	T1	990	1779	T1	981	1849	T1	-9	70
	T2	457	2667	T2	615	2500	T2	158	-167
JE4 (f)	T1	750	1872	T1	800	1768	T1	50	-104
	T2	303	2676	T2	465	2530	T2	162	-146
JE5 (m)	T1	679	1572	T1	751	1527	T1	72	-45
	T2	280	2446	T2	333	2357	T2	53	-89
JE6 (m)	T1	899	1464	T1	780	1437	T1	-119	-27
	T2	385	2002	T2	409	1927	T2	24	-75

Table 2: The mean spectral difference (in Hz) between L1 and target language production.

the phonetic symbols. Kasuya and Sato (1990) state that it is not appropriate to think of the English /aɪ/ as a combination of two vowels /a/ and /ɪ/, as /aɪ/ functions as an independent vowel just like /a/ or /ɪ/ or any other English monophthong.

J and AJ

Panels (e) to (h) in Figure 2 shows the ellipse plots for the two targets of /ai/ by J and AJ talkers. Native vs non-native acoustic differences were also observed for Japanese /ai/. Female AJ talkers had a tendency to produce T1 which was more retracted and male AJ talkers produced T2 which was not sufficiently fronted compared with J talkers. However, as we see below, the two targets of /ai/ are well differentiated in the AJ production and there was more resemblance between the AJ and J spectral patterns than between the AE and JE spectral patterns.

Dissociation from L1 by bilingual talkers

Table 2 shows the mean cross-language spectral difference for each bilingual talker. The extent of shift from one language to another is larger for AJ than for JE talkers, especially for T2. Bilingual AJ talkers clearly demonstrated their awareness for cross-linguistic phonetic differences. There was a much greater separation between the two targets in AJ than in AE data. This is because in their native English, T2 was extremely variable showing a large spread in the F1 direction whereas their T2 in Japanese was more tightly clustered. In producing the Japanese /ai/, every AJ talker shifted their second target to the more peripheral position by producing lower F1 and higher F2 values. Although, in general, JE talkers also showed a shift in the right direction towards the English spectral values, the extent of this approximation was more modest compared to AJ talkers. One JE talker (JE3) changed her production in the opposite direction by producing a slightly more retracted T1 for Japanese than English.

SUMMARY

We saw that the effect of consonantal voicing on the duration of English /aɪ/ was dependent on the type of syllable in which the target sound occurred. Native and non-native groups differed considerably when they produced monosyllabic words, but the talker group difference was negligible in their production of disyllabic words. The duration of Japanese /ai/ was not much affected by the voicing of the following consonant regardless of the talker group. With respect to the spectral characteristics, our preliminary analyses indicated that AJ talkers approximated to the J norms to a greater extent than JE talkers did to the AE norms. While Japanese talkers appeared to equate the two targets of English /aɪ/ and Japanese /ai/ in their production, a clearer separation between the two sets of targets in AJ /ai/ and AE /aɪ/ suggests that Australian talkers have distinct categories for the two targets according to the language they speak.

Why do bilingual Australian talkers dissociate /aɪ/ and /ai/ better than bilingual Japanese talkers do? Perhaps J /ai/ is perceptually more salient or 'marked' due to the extreme position of its second target in the acoustic vowel space, facilitating AJ talkers to notice cross-linguistic phonetic differences. Another possibility is the phonological distribution of AE /aɪ/ and J /ai/ relative to other categories within the respective sound systems. While /nai/ 'there is no ...' and /nae/ 'seedling' are two distinct words in Japanese, there is no such functional difference between /aɪ/ and /aɛ/ in English. In other words, there is a greater need for J /ai/ than AE /aɪ/ to reach the second target more precisely.

We have uncovered numerous acoustic phonetic features that characterize AE /aɪ/ and J /ai/ in both native and non-native talkers' speech. Findings from this study have pedagogical implications, as second/foreign language learners need to acquire the relevant knowledge of the target language norms and develop skills to effectively put it into practice in their speech production.

ACKNOWLEDGEMENT

I am indebted to the members of the Speech, Hearing and Language Research Centre (SHLRC) at Macquarie University [http://www.shlrc.mq.edu.au/] for the use of facilities and technical support. Special thanks go to Associate Professor Jonathan Harrington. Without his continuous support, this study would not have been possible.

REFERENCES

Bernard, J. R. (1967), Length and the identification of Australian English vowels, *Journal of Australasian Universities Modern Language Association (AUMLA)* **27**, 37–58.

Cassidy, S. and Harrington, J. (1996), Emu: An enhanced hierarchical speech data management, *in* P. McCormack and A. Russel (eds), *Proceedings of the Sixth Australian International Conference on Speech Science and Technology*, Australian Speech Science and Technology Association, ASSTA, Adelaide, pp. 361–366.

Cox, F. (1996), *An Acoustic Study of Vowel Variation in Australian English*, PhD thesis, Macquarie University, Sydney, Australia.

Croot, K., Fletcher, J. and Harrington, J. (1992), Levels of segmentation and labelling in the Australian Database of Spoken Language, *in* J. Pittam (ed.), *Proceedings of the Fourth Australian International Conference on Speech Science and Technology*, Australian Speech Science and Technology Association, ASSTA, Brisbane, pp. 86–91.

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

Harrington, J. M. and Cassidy, S. (1999), *Techniques in Speech Acoustics*, Kluwer Academic Publishers, Dordrecht, The Netherlands.

Harrington, J. M., Cox, F. and Evans, Z. (1997), An acoustic phonetic study of broad, general, and cultivated Australian English vowels, *Australian Journal of Linguistics* **17**(2), 155–184.

Hirasaka, F. and Kamata, S. (1981), English and Japanese diphthongs: An acoustic approach, *Sophia Linguistica* **8/9**, 183–195.

Jha, S. K. (1985), Acoustic analysis of the Maithili diphthongs, *Journal of Phonetics* **13**(1), 107–115.

Kasuya, H. and Sato, S. (1990), Long vowel, vowel sequence and diphthong, *in* M. Sugito (ed.), *Japanese and Japanese Language Education*, Vol. 3, Meiji Shoin, Tokyo, Japan, pp. 178–197.

Kubozono, H. (1998), *Seminar Series in English Linguistics*, 1, 2 edn, Kuroshio, Tokyo.

Mitchell, A. and Delbridge, A. (1965), *The Speech of Australian Adolescents*, Angus and Robertson, Sydney.

Roberge, C. and Inoue, M. (1988), A perceptive and acoustic comparison between the English diphthong [aɪ] and the Japanese vowels [ai], *Sophia Linguistica* **26**, 176–183.

Tsukada, K. (2000), Effects of consonantal voicing on English diphthongs: A comparison of L1 and L2 production, *Proceedings of the 6th International Conference on Spoken Language Processing*, Beijing, China.