A Comparison of the Extent of F0, F1 and F2 Movements in Native and Non-native English Vowels: Potential Cues to a Foreign Accent

Kimiko Tsukada

Division of Speech and Hearing Sciences, University of Alabama at Birmingham
e-mail address: tsukada@uab.edu

Abstract

The present study examined the extent of pitch and formant movement in English vowels produced by Australian and Japanese speakers. In an earlier perception study, stimuli elicited from the same speakers were presented to 19 native English listeners who were asked to identify the native speaker in a forced-choice task. The stimuli – monosyllabic English words presented in pairs – had been controlled such that the vowels by native and non-native speakers were matched for F1, F2 at the midpoint and vowel duration. The results revealed that native listeners were able to differentiate native from non-native speakers accurately even when acoustic variability was highly controlled. The aim of the present study was to search for residual acoustic cues that the listeners may have used to detect a foreign accent and determine the identity of the speakers. The Australian and Japanese groups differed in F0 and F2 ranges, suggesting a potential contribution of these cues to a foreign accent.

1 Introduction

Native speakers are known to be capable of detecting small divergences from the phonetic norms of their language [6]. Not only sophisticated listeners, but untrained listeners reliably detected a French accent in English speech materials ranging from phrases (e.g., two little dogs) to very short stimuli which did not even sound speech-like (i.e., the first 30 ms of /t/). The accuracy with which accent was detected did not differ between read and spontaneous speech. These findings were interpreted as suggesting that listeners develop very detailed phonetic category prototypes against which to evaluate speech sounds in their native language.

In another study, it was demonstrated that native English listeners were able to correctly identify the native speaker when they heard pairs of monosyllabic English words, one produced by an Australian speaker (AE) speaker and the other by a Japanese speaker (JE) almost 80% of the time [19, 20]. The stimuli were controlled such that the JE vowels were matched numerically for F1, F2 at the midpoint and duration to the AE tokens. A possible explanation for listeners' ability to distinguish between the two is that they used the acoustic cues that were left to vary (e.g., the extent of formant movements) in making perceptual judgments.

Although the difficulty in directly relating complex acoustic events to perceptual judgements has been pointed out [5, 6], an attempt was made in this study to gain an insight into what might constitute the perceptual phonetic norms by carefully assessing acoustic properties of native and non-native vowels.

This study focused on English vowels elicited from Australian English speakers and Japanese second language (L2) learners. A brief summary of some characteristics of Japanese-accented English is provided below.

1.1 Characteristics of Japanese-accented English vowels

Japanese has five short vowels (/i e a o u/) and five long vowels (/i: e: a: o: u:/) and vowel length is phonemic (e.g., *kado* "corner" vs. *kaado* "card" vs. *kadoo* "flower arrangements"). On the other hand, regardless of dialectal variation, English has more vowel phonemes than Japanese. This requires Japanese L2 learners to divide their existing vowel space more finely in order to accommodate all the English vowels.

Certain English vowels are expected to be more difficult for Japanese L2 learners to perceive and produce than other vowels. For instance, /æ, ʌ, ɜ, a/ all occur in a vowel space occupied by a single Japanese vowel /a/ (or its long counterpart /a:/) and the productions of these English vowels are poorly differentiated in the L2 vowel system [20]. An overlap between /ʊ/ and /u/ was also common among the JE speakers, possibly because they associate these English vowels with a short and long Japanese /u/ and neutralize a vowel quality contrast. In general, JE speakers' vowel space showed varying degrees of resemblance to their L1 vowel space.

ISBN 1-876346-48-5

It is well known that L2 learners from various linguistic backgrounds (e.g., Arabic, Dutch, French) do not lengthen vowels before voiced consonants as much as native English speakers do [4, 8, 15, 16]. In other words, non-native speakers show a smaller voicing effect of a following consonant on vowel duration than native English speakers do. The same has been observed for JE speakers [11, 12, 20]. Finally, a large amount of inter- and intra-speaker variability was evident in the JE productions as is usually the case with interlanguage data [20].

1.2 The present study

The present study is a follow-up of the perception study conducted earlier [19, 20], which sought to determine whether native English listeners are capable of detecting a foreign accent in isolated monosyllabic English words when speaker-related temporal and spectral acoustic variability was experimentally controlled. The earlier study found that native English listeners are capable of differentiating native from non-native English speakers even in the absence of variability relating to F1, F2 and vowel duration. Thus, an explanation needed to be sought in other acoustic cues.

In the present study, the extent of F0, F1 and F2 movements was assessed for their potential contribution to encoding a foreign accent. This decision was motivated by previous research that suggested that English vowels show more extensive formant movement than languages such as Spanish and German [7]. Another study [17] reported that in Canadian English, not only diphthongs, but monophthongs such as /ɪ/, /ɛ/ and /æ/ showed significant formant frequency changes in F1 and F2. An overlap between /ɛ/ and /æ/ in the F1/F2 plane in some dialects of American English was reduced by including the information pertaining to formant trajectories [10], suggesting a crucial role played by formant movement in the identification of these vowels.

It is not clear if Japanese vowels show more or less formant movement than English vowels, as very few studies have examined formant trajectories of Japanese vowels [13] and virtually no cross-linguistic comparison between English and Japanese vowels with respect to the extent of formant movement seems available. However, if L2 learners' vowels substantially differ from native speakers' vowels in the extent of dynamic formant movement, the difference may be audible to native English listeners. The potential contribution of formant movement to the authenticity in L2 learners' vowel production was pointed out in [1].

The question addressed in this study is whether JE vowels show a larger or smaller formant movement than AE vowels. Given the design of this study and nature of speech materials examined, whatever between-group acoustic differences are likely to be subtle, subcategorical ones. Such subcategorical deviations may not be of immediate risk to communicative efficiency. Nevertheless, the knowledge gained from understanding the minimum amount of information required to detect a foreign accent will be useful in research areas such as speech recognition, speech synthesis, forensic phonetics as well as second language acquisition.

vowel	before /t/	before /d/
/ɪ/	bit	bid
/ɛ/	bet	bed
/æ/	bat	bad
/ɒ/	cot	cod
/ʊ/	put	good
/ʌ/	but	bud
/i/	beat	bead
/a/	Bart	barred
/ɔ/	bought	board
/u/	suit	sued
/ɜ/	Burt	bird

Table 1: *Target words produced by AE and JE speakers.*

2 Method

2.1 Talkers and speech materials

Six female AE and 14 female JE speakers were originally recruited to record eleven English vowel categories /ɪ, ɛ, æ, ɒ, ʊ, ʌ, i, a, ɔ, u, ɜ/ in /CVt/ and /CVd/ words (Table 1). Of all the speech materials recorded and digitized at 20 kHz using a signal processing package ESPS/Waves (12th order LPC analysis), 66 tokens (11 vowels x 2 contexts x 3 tokens) each were selected from AE and JE productions on condition that they did not differ from each other by more than 5% on vowel duration, F1 and F2 at the vowel's acoustic midpoint. The resulting tokens were presented in pairs to 19 native English listeners for accent detection. The listeners correctly identified the native speaker about 80% of the time.

acoustic parameter	AE	JE
vowel duration (ms)	203 (79)	207 (78)
F1 at midpoint (Hz)	581 (187)	582 (186)
F2 at midpoint (Hz)	1747 (612)	1769 (599)

Table 2: *The mean acoustic characteristics of the stimuli. The values are averaged over eleven vowels, six tokens each. Standard deviations are in parentheses.*

By design, there was no significant between-group difference in the three acoustic parameters controlled in the stimuli (Table 2) and this was verified by two-way ANOVAs (2 groups x 11 vowels): $F(1, 110) = 0.01$, *ns* for F1, $F(1, 110) = 0.76$, *ns* for F2 and $F(1, 110) =$

ISBN 1-876346-48-5

0.12, *ns* for vowel duration. The main effect of Vowel was highly significant for all three acoustic parameters. However, there was no interaction between the two main factors for any of the parameters, indicating that the acoustic characteristics of vowels did not differ according to the speaker groups in this study.

2.2 Data analysis

A total of 132 tokens used in the accent detection task were analyzed using Praat [18]. Phonetic boundaries for the vowels of interest were marked on the basis of waveform and spectrographic information according to the criteria set in [3]. The minimum and maximum values between the onset and offset of the vowel portion were identified for F0, F1 and F2. The extent of pitch and formant movements was estimated by subtracting the minimum value from the maximum value.

3 Results

Acoustic measurements were submitted to two-way ANOVAs (2 groups x 11 vowels). The dependent variables were F0, F1 and F2 ranges. Table 3 shows the mean values of the stimuli.

acoustic parameter	AE	JE
F0 range (Hz)	40 (20)	63 (24)
F1 range (Hz)	182 (104)	197 (101)
F2 range (Hz)	487 (229)	421 (210)

Table 3: *The mean ranges of F0, F1 and F2. The values are averaged over eleven vowels, six tokens each. Standard deviations are in parentheses.*

3.1 F0 range

There was a significant main effect of Group [$F(1, 110) = 40.3$, $p < 0.001$]. The JE group had a larger mean F0 range than the AE group (63 Hz vs. 40 Hz). The Vowel effect was not significant [$F(10, 110) = 1.1$, *ns*], but for both groups, long vowels such as /a/ and /u/ tended to have a larger F0 range than a short vowel such as /ʌ/. There was no interaction between Group and Vowel effects [$F(10, 110) = 1.6$, *ns*].

It is interesting to note that Japanese uses pitch accent to differentiate word meaning, e.g., *kata* (high – low) "shoulder" vs. *kata* (low – high) "pattern)". Perhaps because of this, Japanese L2 learners are influenced by non-linguistic pitch movement in English vowels to a greater extent than native English speakers are. This may partially explain a slightly larger F0 range for JE than for AE speakers.

3.2 F1 range

The main effect of Group was non-significant [$F(1, 110) = 1.5$, *ns*], as the mean F1 ranges for AE and JE differed very little (182 Hz for AE and 197 Hz for JE). The Vowel effect was highly significant [$F(10, 110) = 16.1$, $p < 0.0001$]. However, it did not interact with the Group effect [$F(10, 110) = 1.8$, *ns*].

3.3 F2 range

The main effect of Group was significant [$F(1, 110) = 4.4$, $p < 0.05$]. The JE group had a smaller mean F2 range than the AE group did (487 Hz vs. 421 Hz). The main effect of Vowel was highly significant [$F(10, 110) = 6.5$, $p < 0.001$], but there was no interaction between the two factors [$F(10, 110) = 1.7$, *ns*].

Both AE and JE groups produced a wide F2 range (> 600 Hz) for high back vowels such as /ʊ/ and /ɔ/ and the narrowest F2 range for the low front vowel /æ/ (< 300 Hz). Back vowels typically have low F2 values [14]. Perhaps the observed wide F2 range for these back vowels was due to a coarticulatory effect of the final consonant which pulled F2 towards the locus for alveolar stops at the vowel offset.

The /i/ vowel by the AE speakers had the largest F2 range at 778 Hz of all vowel types examined here and this was more than 300 Hz greater than the same vowel by the JE speakers (414 Hz). An extensive F2 movement for /i/ is characteristic of Australian English, as various studies reported a long onglide starting at a lower position in the vowel space and delayed target for this vowel in Australian English [2, 9, 21].

4 Discussion

We have seen that Japanese L2 learners produced a significantly larger F0 range and smaller F2 range in English vowels than did Australian English speakers. Although the observed between-group differences were only modest and their perceptual relevance is unclear, even small acoustic differences could be of importance. This is because the AE and JE tokens were chosen to have matched vowels. The JE vowels examined here are expected to be highly intelligible, because only those tokens that fell within the AE range were included. It may be that, in the absence of obvious differences in vowel quality, small differences in the extent of pitch and formant movements may become audible and cue a foreign accent.

The speech materials examined here did not capture typical characteristics of non-native vowels, as only those tokens that satisfied stringent criteria were considered. It was not the primary purpose of the study to exhaustively analyze all the possible non-native vowel tokens. It would nonetheless be interesting to conduct a thorough acoustic analysis of many more tokens, given the complex nature of speech sounds as pointed out in [5]. The results of this study point to the potential perceptual relevance of the native vs. non-native

ISBN 1-876346-48-5

differences in F0 and F2 ranges. Although range differences reflect the extent of dynamic pitch and formant movement and may provide useful information with respect to the degree of coarticulation, they may still be a relatively coarse measure. If examined in further detail, there may be other acoustic cues that serve to differentiate the native from non-native groups more efficiently. In this study, formant frequency contours were not considered. It is possible that the AE and JE vowel tokens differed in this respect.

It may also be interesting to examine the L1 vowel production by the same JE speakers. If it is the case that the Japanese language demands that the speaker realize F0 with a wide range, the extent to which an extended F0 range carries over to an L2 may differ according to factors such as age and/or experience with L2. In other words, the adult Japanese L2 learner may be more strongly influenced by their L1 than the child Japanese L2 learner, who may exhibit distinct patterns of F0 range depending on the language he or she is using (i.e., narrow in English vs. wide in Japanese).

5 Summary

The present study provided a follow-up analysis of the stimuli that had previously been judged by native English listeners in the accent detection task. The JE group had a larger F0 range and a smaller F2 range than the AE group, suggesting the possibility that these between-group acoustic differences affected native listeners' judgements of a foreign accent.

References

[1] Bohn, O.-S. and Flege, J. E. "The production of new and similar vowels by adult German learners of English", *Studies in Second Language Acquisition,* ***14***(2), pp. 131–158, 1992.

[2] Cox, F., *An Acoustic Study of Vowel Variation in Australian English.*, Unpublished doctoral dissertation, Macquarie University, 1996.

[3] Croot, K., Fletcher, J. and Harrington, J., "Levels of segmentation and labelling in the Australian Database of Spoken Language", *Proceedings of the Fourth Australian International Conference on Speech Science and Technology*, pp. 86–91, 1992.

[4] Elsendoorn, B. A. G., "Production and perception of Dutch foreign vowel duration in English monosyllabic words", *Language and Speech,* ***28***(3), pp. 231–254, 1985.

[5] Flanagan, J., "A difference limen for vowel formant frequency", In Lehiste, I. (ed.), *Reading in Acoustic Phonetics.* Cambridge, Mass.: The M.I.T. Press, 1967, pp. 288–291.

[6] Flege, J. E., "The detection of French accent by American listeners", *J.A.S.A.,* ***76***(3), pp. 692–706, 1984.

[7] Flege, J. E., Bohn, O.-S. and Jang, S., "Effects of experience on non-native speakers' production and perception of English vowels", *J. of Phonetics,* ***25***(4), pp. 437–470, 1997.

[8] Flege, J. E. and Port, R., "Cross-language phonetic interference: Arabic to English", *Language and Speech,* ***24***(2), pp. 125–146, 1981.

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

[10] Hillenbrand, J., Getty, L. A., Clark, M. J. and Wheeler, K., "Acoustic characteristics of American English vowels", *J.A.S.A.,* ***97***(5), pp. 3099–3111, 1995.

[11] Hiramatsu, K., "Timing acquisition by Japanese speakers of English", *Reading University, Speech Research Lab. Work in Progress,* ***6***, pp. 49–76, 1990.

[12] Ichizaki, K., "Durational features of English sounds: Comparison between native speakers of English and Japanese college students", *J. of the Phonetic Society of Japan,* ***1***(1), pp. 25–37, 1997.

[13] Imaizumi, S., Kiritani, S., Hirose, H., Togami, S. and Shirai, K., "Preliminary report on the effects of speaking rate upon formant trajectories", *Ann. Bull., Research Institute of Logopedics and Phoniatrics,* ***21***, pp. 147–151, 1987.

[14] Ladefoged, P., *A Course in Phonetics*, Fort Worth, TX: Harcourt, Brace & Company, 1993.

[15] Mack, M., "Voicing-dependent vowel duration in English and French: Monolingual and bilingual production", *J.A.S.A.,* ***71***(1), pp. 173–178, 1982.

[16] Munro, M., "Productions of English vowels by native speakers of Arabic: Acoustic measurements and accentedness ratings", *Language and Speech,* ***36***(1), pp. 39–66, 1993.

[17] Nearey, T. M. and Assmann, P. F., "Modeling the role of inherent spectral change in vowel identification", *J.A.S.A.,* ***80***(5), pp. 1297–1308, 1986.

[18] http://www.praat.org

[19] Tsukada, K., "Japanese-accented English: A perception study", *Asia-Pacific J. of Speech, Language and Hearing* ***3***(1), pp. 43–65, 1998.

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

[21] Watson, C. I., Harrington, J. and Evans, Z., "An acoustic comparison between New Zealand and Australian English vowels", *Australian J. of Linguistics,* ***18***(2), pp. 185–207, 1998.

ISBN 1-876346-48-5