An acoustic comparison of vowel length contrasts in Arabic, Japanese and Thai: Durational and spectral data

Kimiko Tsukada
Department of International Studies, Macquarie University
North Ryde NSW 2109, Australia
Fax: 61-2-9850-7046
e-mail: kimiko.tsukada@mq.edu.au

Abstract

In our earlier perception study, we observed that familiarity with first language (L1) phonemic length contrasts in Japanese does not transfer optimally into an unknown language, Arabic. We hypothesized that this finding is related to cross-language differences in how vowel length contrasts are phonetically realized. The present study compares acoustic characteristics (i.e., vowel duration, first two formant frequencies (F1, F2)) of the /a/ and /a͞/ vowels that are phonemic in three typologically unrelated languages, i.e., Arabic, Japanese and Thai. We sought to understand the extent to which vowel length contrasts are similar or dissimilar in these languages. Acoustic measurements showed short and long categories were clearly differentiated in duration in all three languages. The effect of length was much more limited for F1 and F2. The finding that the short-to-long ratio did not substantially differ across languages suggests that listeners attend to more than just acoustic vowel duration in making perceptual judgments on short vs. long vowels in a discrimination task.

Keywords

vowel length, Arabic, Japanese, Thai, duration, formants

1. Introduction

1.1 Previous cross-language speech perception study

Using an AXB discrimination test frequently employed in previous speech perception research (Harnsberger 2001; Højen and Flege 2006; Wayland and Guion 2003), we have examined the perception of vowel length contrasts by listeners from diverse language backgrounds including native speakers of Arabic (NA), Japanese (NJ), non-native learners of Japanese (NNJ) and speakers of English who know neither Arabic nor Japanese.

Despite the expectation that NA and NJ listeners who are familiar with phonemic vowel length in their own language would be able to accurately perceive vowel length contrasts in each other's language, their discrimination accuracy was optimal only in their L1.

As Figure 1 shows, the NA and NJ listeners showed an opposite discrimination pattern. What was noteworthy was the finding that the NJ listeners did not show an advantage over the NNJ listeners in discriminating the Arabic vowel length contrasts. While neither the NJ nor NNJ listeners had any knowledge of Arabic, the NJ listeners were thoroughly familiar with the short vs. long vowel contrasts in their L1.

These findings are somewhat surprising in that listeners who do not use vowel duration contrastively in their L1 (e.g., Spanish) have been found to rely more on duration than spectral information when they categorized the English vowel contrast /i/-/I/ (as in 'heat' vs. 'hit') (Bohn and Flege 1990). Why is it that the NA and NJ listeners who are familiar with vowel length contrasts did not take advantage of their L1 knowledge to process vowel length contrasts in an unknown language?

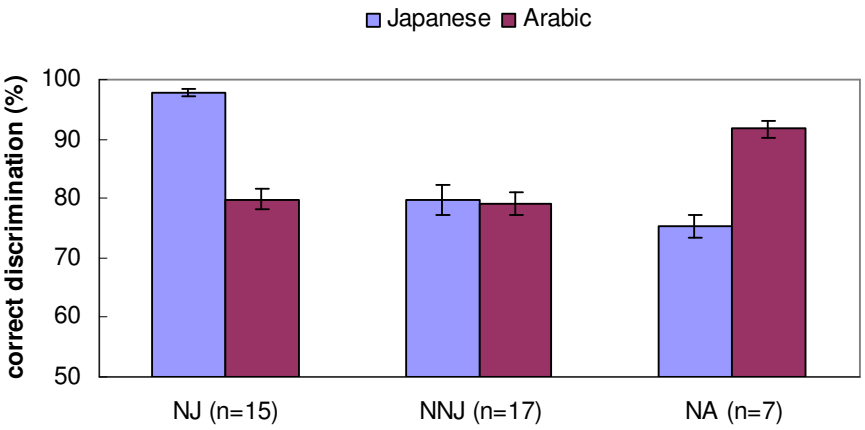

Figure 1. Mean discrimination accuracy (%) for vowel length contrasts in Arabic and Japanese by NJ, NNJ and NA listeners. The brackets enclose ± 1 standard error.

1.2 This study

The aim of the present study was to compare acoustic phonetic characteristics of vowel length contrasts in Arabic, Japanese and Thai. Mean vowel duration, F1 and F2 at the acoustic vowel midpoint are reported. Although all three languages employ vowel duration contrastively, how the short and long categories affect the vowel's duration and spectral characteristics may not be identical. Such cross-language phonetic differences, if found, may affect listeners' perception of what is short and what is long, a concept apparently familiar to them in their L1.

2. Method

2.1 Speech materials

Speakers read $C_1V_1C_2V_2$ words in Japanese and $C_1V_1C_2$ words in Arabic and Thai (where V_1 was either /a/ or /a⏌/ in each language). While this discrepancy is not desirable, the occurrence of the CVC sequence is quite limited in Japanese unlike Arabic or Thai. Furthermore, given that, in Arabic, vowel length is neutralized in open syllables including the word-final position (Thelwall and Sa'adeddin 1990), it was not an option to use CVCV words in Arabic to match the syllable structure across the languages. The syllable structure is fairly simple in Japanese (predominantly V or CV where V is a vowel and C is a consonant) (Shibatani 1990). Words produced in isolation were used for this study.

For Arabic, four minimal pairs in which C_1 was /b d s Σ/ and C_2 was /b m n/ were used. For Japanese, three minimal pairs in which C_1 was zero or /k/ and C_2 was /b t d/ were used. Japanese uses two rhythmic units, i.e., mora and syllable. While all the Japanese test words had two syllables, the test words containing the short /a/ had two morae and the test words containing the long /a⏌/ had three morae. This is because a syllable with a long vowel counts as having two morae in Japanese. For Thai, 15 minimal pairs in which C_1 was /c c^h d h s/ and C_2 was /p t k/ (phonetically unreleased [p| t| k|]) were used. Thai is a tonal language with five contrastive tones. All the Thai test words used in this study were pronounced with a low tone (indicated by /ˆ/).

Test words (all real words) were presented visually to each speaker on the computer screen one word at a time. All words were written using appropriate Arabic, Japanese or Thai script, so there was no ambiguity as to how each word should be read by mature, proficient native speakers. The speakers were asked to read the target words naturally at their normal speaking rate. Each speaker repeated the test words up to three times in randomized orders. The recorded speech materials were digitized at 44.1 kHz using CoolEdit, amplitude normalized to 50% of the peak and stored in separate files per word.

Tables 1-3 show the test words used in Arabic, Japanese and Thai, respectively.

Short: /a/	Long: /a⏌/
بن /ban/ 'coffee (beans)'	بان /ba⏌n/ 'to appear'
دم /dam/ 'blood'	دام /da⏌m/ 'to keep on'
سب /sab/ 'to curse'	ساب /sa⏌b/ 'to leave'
شب /Σab/ 'young man (colloquial)'	شاب /Σa⏌b/ 'young man (standard)'

Table 1. Arabic words used in this study.

Short: /a/	Long: /a⏌/
下部 /kabu/ ‘lower part’	カーブ /ka⏌bu/ ‘curve’
角 /kado/ ‘corner’	カード /ka⏌do/ ‘card’
後 /ato/ ‘later’	アート /a⏌to/ ‘art’

Table 2. Japanese words used in this study.

Short: /a/	Long: /a⏌/
จับ /càp/ ‘to catch’	จาบ /cà⏌p/ ‘weaver-bird’
จัด /càt/ ‘to organize’	จาด /cà⏌t/ ‘zinc oxide’
จัก /càk/ ‘to split’	จาก /cà⏌k/ ‘to leave’
ฉับ /c^{h}àp/ ‘quickly’	ฉาบ /c^{h}à⏌p/ ‘to paint’
ฉัตร /c^{h}àt/ ‘tiered umbrella’	ฉาด /c^{h}à⏌t/ ‘sound of clapping’
ฉัก /c^{h}àk/ ‘sound of a train (onomatopoeia)’	ฉาก /c^{h}à⏌k/ ‘curtain’
ดับ /dàp/ ‘to extinguish’	ดาบ /dà⏌p/ ‘sword’
ดัด /dàt/ ‘to bend’	ดาษ /dà⏌t/ ‘paper’
ดัก /dàk/ ‘to trap’	ดาก /dà⏌k/ ‘plug’
หับ /hàp/ ‘to close’	หาบ /hà⏌p/ ‘to carry on one’s shoulder’
หัด /hàt/ ‘to practice’	หาด /hà⏌t/ ‘beach’
หัก /hàk/ ‘to break’	หาก /hà⏌k/ ‘if’
สับ /sàp/ ‘to chop’	สาบ /sà⏌p/ ‘stink’
สัตว์ /sàt/ ‘animal’	สาด /sà⏌t/ ‘to throw’
สัก /sàk/ ‘to tattoo’	สาก /sà⏌k/ ‘to be coarse’

Table 3. Thai words used in this study.

2.2 Speakers

Native speakers of Arabic, Japanese and Thai participated in the recording sessions lasting between 45 and 60 minutes. All of them reported normal hearing and had no history of language problems in their L1s. They were paid for their participation.

2.2.1 Arabic

Nine (6 male, 3 female) native Arabic speakers with a mean age of 40.8 years participated. They were all born in Arabic-speaking countries and had lived in Sydney, Australia for 12

years on average at the time of recording. They were recorded in the Audio-Visual recording studio and the Centre for Language Sciences (CLaS) recording studio at Macquarie University, Sydney.

The Arabic speakers' dialectal backgrounds were Lebanese (n=5), Egyptian (n=1) and Saudi Arabic (n=3). Vowel length is phonemic in all these varieties of Arabic. While some cross-dialectal differences in vowel quality have been reported (Alghmadi 1998), phonemic short and long vowels in different dialects are clearly distinguished durationally.

2.2.2 Japanese

Eight (4 male, 4 female) native Japanese speakers with a mean age of 25.4 years participated. With the exception of one participant who had lived in the US for 11 months, none of them had lived overseas for an extended period of time. They were recorded in the sound-treated room in the Department of Linguistics, University of Kobe, Japan. These speakers primarily came from the Western part of Japan including Ehime, Fukuoka, Hyogo, Kagawa, Kanagawa and Tottori prefectures. Vowel length is phonemic to all the speakers. The intended pitch pattern for the test words was HIGH on the first syllable and LOW on the second syllable.

2.2.3 Thai

Twenty-two (5 male, 17 female) native Thai speakers with a mean age of 31.5 years participated. With the exception of two participants who were born in Sydney to native Thai-speaking parents and went back to Thailand to spend their childhood, they were all born in Thailand. They had lived in Sydney, Australia for 5.7 years on average at the time of recording. The recording sessions took place in the MARCS Auditory Laboratories recording studio at the University of Western Sydney, Australia.

2.3 Data Processing and Analysis

The first two formant frequencies (F1, F2) of the target vowels were automatically tracked at the temporal midpoint using the signal processing package ESPS/Waves (http://www.speech.kth.se/software/). The settings were 12th order linear predictive coding analysis, cosine window, 49-ms frame size, and 5-ms frame shift. The EMU speech database system (http://emu.sourceforge.net/) was used for phonetically labelling the speech segments of interest and the formant values were calculated in the R statistical environment (http://cran.r-project.org/). The beginning and end of each vowel token was identified by inspection of wide-band spectrograms and time domain waveforms following the criteria used in our previous study (Hirata and Tsukada 2009).

3. Results

3.1 Vowel duration

Mean vowel durational values were submitted to a two-way ANOVA (Analysis of Variance) in which Group (Arabic, Japanese, Thai) served as a between-subjects factor and

Length (short, long) served as a within-subjects factor. The dependent variable was the mean vowel duration (in ms) by each speaker.

Table 4 and Figure 2 show the mean duration for /a/ and /aː/ in each language. Table 4 shows the short-to-long (S/L) ratio, as well.

	Short: /a/	Long: /aː/	S/L ratio
Arabic	108 (18)	250 (54)	0.43
Japanese	82 (13)	211 (39)	0.39
Thai	147 (22)	324 (45)	0.45

Table 4. Mean vowel duration (in ms) in Arabic, Japanese and Thai. Standard deviations are in parentheses.

While there was some between-language variation in the short-to-long ratio, the long /aː/ was more than twice as long as the short /a/ in all three languages. A clear difference of this magnitude is expected to be highly salient and audible at least to proficient native speakers of each language. The mean vowel duration values obtained in this study are in good agreement with the values reported in previous studies (Abramson 1962, 1974, 2001; Alghmadi 1998; Gandour 1984; Hirata 2004a; Hirata and Tsukada 2009).

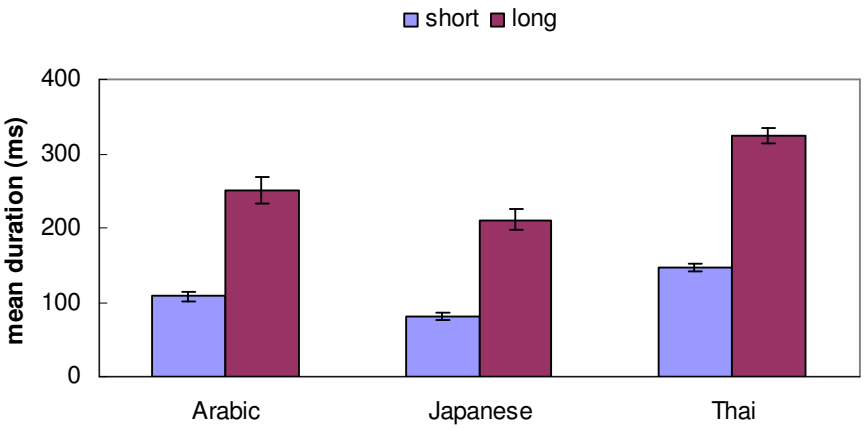

Figure 2. Mean duration (ms) for /a/ and /aː/ in Arabic, Japanese and Thai. The brackets enclose ± 1 standard error.

Group x Length ANOVA yielded significant main and interaction effects [G: $F(2, 36) = 28.7$, $p < 0.001$, L: $F(1, 36) = 587.5$, $p < 0.001$, G x L: $F(2, 36) = 7.9$, $p < 0.01$]. As can be clearly seen in Figure 2, in all three languages, the short /a/ was much shorter than the long /aː/. In each language, native speakers were fairly consistent as indicated by very small standard errors. However, both short and long vowels were relatively longer in Thai than in the other two languages, which presumably led to the significant two-way interaction.

The simple effect of Group was significant for both short and long categories [Short: $F(2, 36) = 36.8$, $p < 0.001$, Long: $F(2, 36) = 20.9$, $p < 0.001$]. For the short category, the Thai /a/

was significantly longer than the Arabic /a/, which, in turn, was significantly longer than the Japanese /a/. For the long category, the Thai /aˉ/ was significantly longer than both Arabic and Japanese /aˉ/, which did not differ significantly from each other. It is not entirely clear if these cross-language differences are inherent within each language or specific to the participants of this study as the speaking rate was not strictly controlled.

The simple effect of Length was significant for all three groups [Arabic: $F(1, 8) = 90.5$, $p < 0.001$, Japanese: $F(1, 7) = 149.4$, $p < 0.001$, Thai: $F(1, 21) = 676.7$, $p < 0.001$]. Short and long categories were clearly separated in each language.

3.2 F1

Mean F1 values were submitted to a two-way ANOVA in which Group (Arabic, Japanese, Thai) served as a between-subjects factor and Length (short, long) served as a within-subjects factor. Male and female data were analyzed separately. The dependent variable was the F1 values (in Hz) averaged across the target words by each speaker measured at the vowel's temporal midpoint.

Figure 3 shows the mean F1 for /a/ and /aˉ/ in each language. For data produced by female speakers, both main effects and two-way interaction were significant [G: $F(2, 21) = 11.7$, $p < 0.001$, L: $F(1, 21) = 6.1$, $p < 0.05$, G x L: $F(2, 21) = 6.4$, $p < 0.01$]. The simple effect of Group was significant for both short [$F(2, 14) = 11.3$, $p < 0.01$] and long [$F(2, 14) = 10.7$, $p < 0.01$] vowels. For both /a/ and /aˉ/, F1 had the highest values for Thai. The simple effect of Length was significant in Japanese [$F(1, 21) = 16.4$, $p < 0.001$] and Thai [$F(1, 21) = 6.3$, $p < 0.05$], but not in Arabic. For both the Japanese and Thai groups, the F1 value was higher for /aˉ/ than for /a/, suggesting that the vowel was produced with a lower tongue position and/or wider jaw opening (Hirata and Tsukada 2009).

For data produced by male speakers, only the main effect of Group [$F(2, 12) = 11.8$, $p < 0.01$] was significant. F1 was highest in Thai, lowest in Arabic and intermediate in Japanese. Although in all three languages, the mean F1 was higher for /aˉ/ than for /a/, the main effect of Length did not reach significance. It may be the case that the Thai speakers generally have a more careful speaking style with precise articulatory gestures than do the Arabic and Japanese speakers.

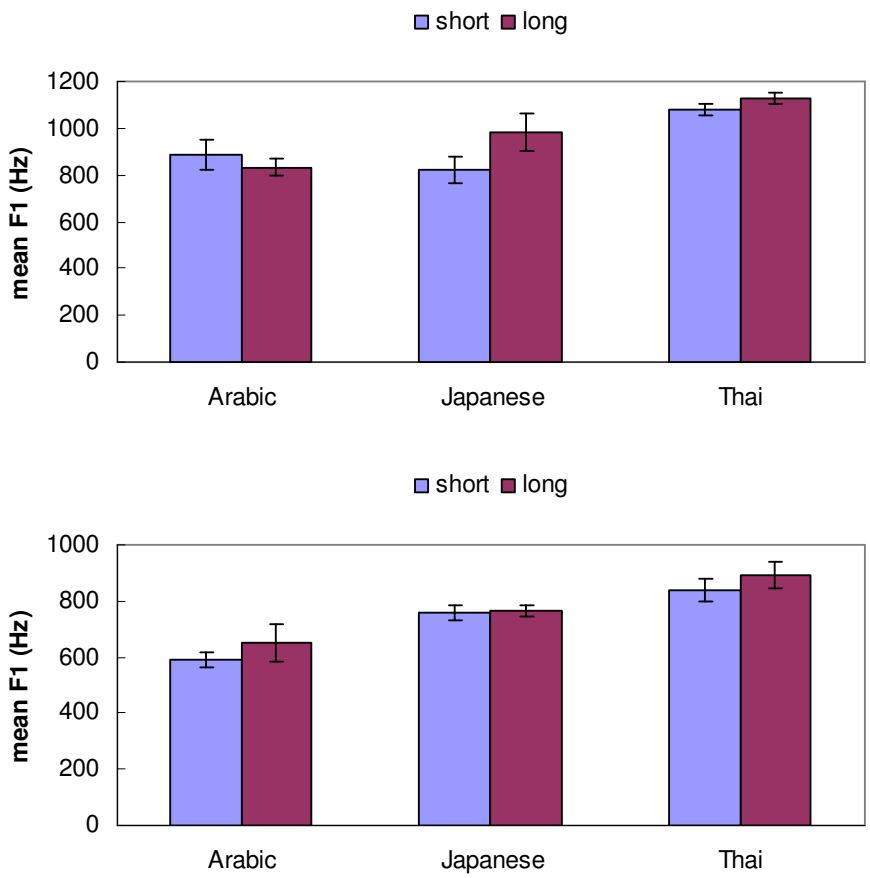

Figure 3. Mean F1 (Hz) at the temporal midpoint for /a/ and /a˞/ in Arabic, Japanese and Thai by female (top) and male (bottom) speakers. The brackets enclose ± 1 standard error.

3.3 F2

Mean F2 values were submitted to a two-way ANOVA in which Group (Arabic, Japanese, Thai) served as a between-subjects factor and Length (short, long) served as a within-subjects factor. Male and female data were analyzed separately. The dependent variable was the F2 values (in Hz) averaged across the test words by each speaker measured at the vowel's temporal midpoint.

Figure 4 shows the mean F2 for /a/ and /a˞/ in each language. The effect of Length was more limited for F2 than for F1. The pattern of results of statistical analyses was essentially the same for male and female speakers. Only the main effect of Group was significant [female: $F(2, 21) = 9.1$, $p < 0.01$, male: $F(2, 12) = 7.0$, $p < 0.01$]. F2 was highest in Arabic, lowest in Japanese and intermediate in Thai.

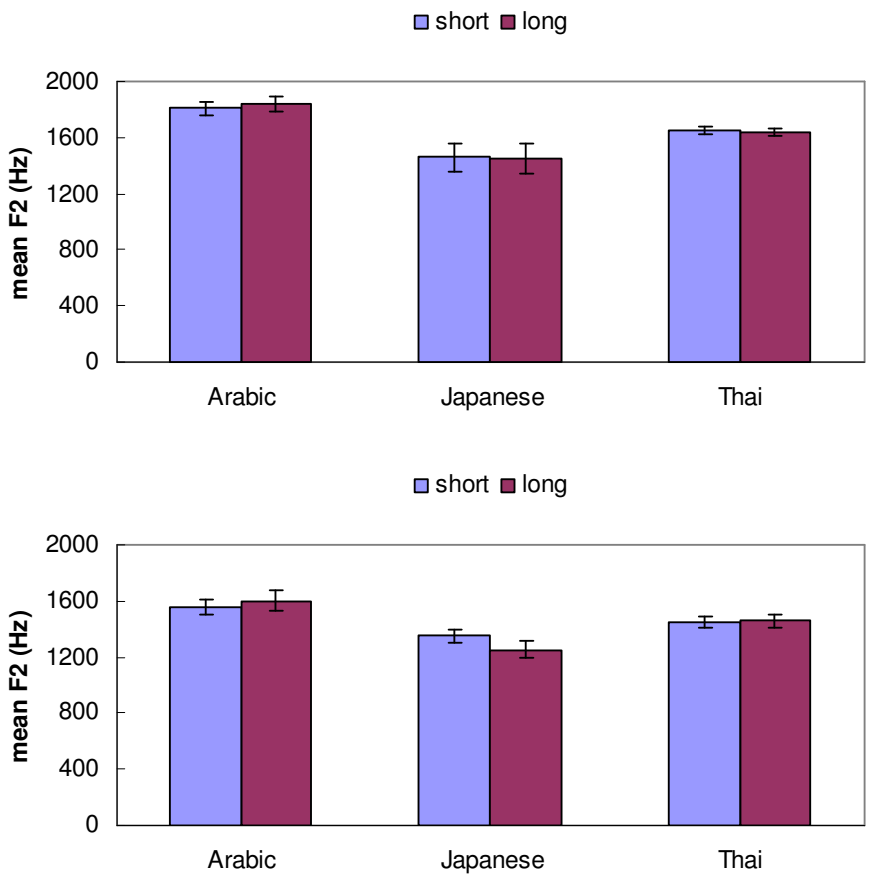

Figure 4. Mean F2 (Hz) at the temporal midpoint for /a/ and /aː/ in Arabic, Japanese and Thai by female (top) and male (bottom) speakers. The brackets enclose ± 1 standard error.

4. Discussion

We have observed that short and long vowels in Arabic, Japanese and Thai are clearly kept distinct in terms of acoustic duration. There were some cross-language differences in the absolute vowel duration and, for both short and long categories, the Thai vowels were longer than the vowels in the other languages. However, the short-to-long ratio did not substantially differ across languages. The effect of vowel length was more limited for the formant data. In other words, vowel quality was quite stable in these languages despite a large durational difference for the short and long vowels. Given these results, one may wonder why the NJ listeners in our earlier perception study did not fully take advantage of their L1 knowledge when they heard Arabic short and long vowels.

Some NJ participants commented that they found it difficult to focus on the target vowel in Arabic words, because the segmental boundary between the target vowel and the following word-final consonant was not clear. Difference in syllable structure (CVCV in Japanese and CVC in Arabic) may have played a role here. In particular, for the NJ listeners who have mora as a basic rhythmic unit, it may be relatively difficult to make short vs. long judgments in monosyllabic (Arabic) words. In monosyllabic words, there is only one vowel and the listeners need to decide whether the target vowel is short or long without referring to the sound outside the syllable. In other words, relative durational differences between the two adjacent syllables may be very important to the NJ listeners when they make short vs. long judgments in a discrimination task.

There was also a remark that it was easier to make perceptual decisions when the stimulus in question contained a long vowel. This may be related to the finding that long vowels stretch to a greater extent than short vowels when the speaking rate varied from fast to slow (Hirata 2004a). It is, thus, possible that the listeners tended to select the long category when in doubt.

Pairs of vowels differing in duration (e.g., /ɪ/ in 'heat' vs. /I/ in 'hit' in English) may be accompanied by vowel quality difference to varying degrees depending on the language (Abramson and Ren 1990; Alghmadi 1998; Hirata and Tsukada 2009; Tsukada and Roengpitya 2008). However, there was very limited acoustic evidence in support of durationally motivated vowel quality difference in the present study. That is, only in the speech of female Japanese and Thai speakers was the effect of length reached statistical significance. There was no effect of vowel length on the quality of Arabic vowels despite the fact that there are less vowel categories in Arabic than in Japanese and Thai. With fewer vowels in the system, one might expect Arabic vowels to show greater variation than Japanese or Thai vowels. These observations may only hold for the /a/-/aˈ/ pair and, in future research, a systematic, cross-linguistic comparison of coarticulation involving more vowel categories should be carried out.

Whether the length contrasts are fragile or robust seems to be influenced by their frequency in the language (Aoyama 2001; Aoyama and Reid 2006). Although lexical frequency was not controlled in this study, it may provide a useful insight in how vowel contrasts are phonetically realized in Arabic, Japanese and Thai.

Given a clear difference between the short and long categories in their acoustic vowel duration in the three languages examined and the well-known difficulty in the production and perception of length contrasts for non-native Japanese learners (Hirata 2004b; Toda 2003), a developmental approach may be very useful. It appears that very few studies have examined how children of varying ages acquire the length contrasts in their L1s (see, however, Aoyama 2001, for an exception).

5. Conclusions

This study compared acoustic characteristics of vowels in Arabic, Japanese and Thai, all of which use vowel duration contrastively. While the effect of vowel length on vowel quality was quite limited, short and long vowels were clearly distinct in duration in all three languages with long vowels being more than twice as long as their short counterparts. The short-to-long ratio did not substantially differ across languages. This suggests that listeners may be attending to more than just acoustic vowel duration in making perceptual judgments on short vs. long vowel contrasts in a discrimination task.

6. Acknowledgment

This study was funded by the Macquarie University New Staff Grant (Reference No. 9200800377) and the University of Western Sydney Seed Grant. Part of this work (entitled "Perception of Arabic and Japanese vowel length contrasts by native vs. non-native Japanese listeners" co-authored with Yukari Hirata) was presented at the Acoustical Society of America (ASA) 2nd Special Workshop on Cross-Language Speech Perception and Variations in Linguistic Experience which was held in Portland, Oregon, USA in May 2009. I thank Professor Haruo Kubozono and Ms. Akiko Takemura of University of Kobe for their help with the recording sessions in Kobe, Japan, Ms. Nejimeh Habib for her help with selecting the Arabic test words and recruiting native Arabic speakers and Mr. Chris Callaghan for his help with the recording session conducted in the Centre for Language Sciences (CLaS), Department of Linguistics, Macquarie University. I also thank participants who represent diverse language backgrounds.

7. References

Abramson, A. S., 1962, *The vowels and tones of standard Thai: Acoustical measurements and experiments.* Indiana University Research Center in Anthropology, Folklore and Linguistics, Pub. 20.

Abramson, A. S., 1974, Experimental phonetics in phonology: Vowel duration in Thai, *Pasaa,* vol. 4, pp. 71-90.

Abramson, A. S., 2001, The stability of distinctive vowel length in Thai, in *Essays in Tai Linguistics,* M. R. K. Tingsabadh and A. Abramson, Eds. Bangkok: Chulalongkorn University Press, pp. 13-26.

Abramson, A. S. and Ren, N. 1990, Distinctive vowel length: duration vs. spectrum in Thai, *Journal of Phonetics,* vol. 18, no. 1, pp. 79-92.

Alghmadi, M. M., 1998, A spectrographic analysis of Arabic vowels: A cross-dialectal study, *Journal of King Saud University,* vol. 10, pp. 3-24.

Aoyama, K. 2000, *A Psycholinguistic Perspective on Finnish and Japanese Prosody: Perception, production and child acquisition of consonantal quantity distinctions.* University of Hawaii at Mānoa.

Aoyama, K. and Reid, L. A., 2006, Cross-linguistic tendencies and durational contrasts in geminate consonants: an examination of Guinaang Bontok geminates, *Journal of the International Phonetic Association,* vol. 36, no. 2, pp. 145-157.

Bohn, O.-S. and Flege, J. E., 1990, Interlingual identification and the role of foreign language experience in L2 vowel perception, *Applied Psycholinguistics,* vol. 11, no. 3, pp. 303-328.

Gandour, J., 1984, Vowel duration in Thai, *Crossroads,* vol. 2, pp. 59-63.

Harnsberger, J. D., 2001, On the relationship between identification and discrimination of non-native nasal consonants, *Journal of the Acoustical Society of America,* vol. 110, no. 1, pp. 489-503.

Hirata, Y., 2004a, Effects of speaking rate on the vowel length distinction in Japanese, *Journal of Phonetics,* vol. 32, no. 4, pp. 565-589.

Hirata, Y., 2004b, Training native English speakers to perceive Japanese length contrasts in word versus sentence contexts, *Journal of the Acoustical Society of America,* vol. 116, no. 4, pp. 2384-2394.

Hirata, Y. and Tsukada, K., 2009, Effects of speaking rate and vowel length on formant frequency displacements in Japanese, *Phonetica,* vol. 66, no. 3, pp. 129-149.

Højen, A. and Flege, J. E., 2006, Early learners' discrimination of second-language vowels, *Journal of the Acoustical Society of America,* vol. 119, no. 5, pp. 3072-3084.

Shibatani, M., 1990, *The languages of Japan.* Cambridge University Press, New York.

Thelwall, R. and Sa'adeddin, M. A., 1990, Arabic, *Journal of the International Phonetic Association,* vol. 20, no. 2, pp. 37-39.

Toda, T., 2003, Acquisition of special morae in Japanese as a second language, *Journal of the Phonetic Society of Japan,* vol. 7, no. 2, pp. 70-83.

Tsukada, K. and Roengpitya, R., 2008, Discrimination of English and Thai words ending with voiceless stops by native Thai listeners differing in English experience, *Journal of the International Phonetic Association,* vol. 38, no. 3, pp. 325-347.

Wayland, R. and Guion, S., 2003, Perceptual discrimination of Thai tones by naïve and experienced learners of Thai, *Applied Psycholinguistics,* vol. 24, no. 1, pp. 113-129.