The perception of Arabic and Japanese short and long vowels by native speakers of Arabic, Japanese, and Persian

Kimiko Tsukada[a)]
Department of International Studies, Macquarie University, North Ryde New South Wales 2109, Australia

(Received 8 February 2010; revised 7 October 2010; accepted 17 November 2010)

This study examines the perception of short and long vowels in Arabic and Japanese by three groups of listeners differing in their first languages (L1): Arabic, Japanese, and Persian. While Persian uses the same alphabet as Arabic and Iranian students learn Arabic in school, the two languages are typologically unrelated. Further, unlike Arabic or Japanese, vowel length may no longer be contrastive in modern Persian. In this study, a question of interest was whether Persian listeners' foreign language learning experience or Japanese listeners' L1 phonological experience might help them to accurately process short and long vowels in Arabic. In Experiment 1, Arabic and Japanese listeners were more accurate than Persian listeners in discriminating vowel length contrasts in their own L1 only. In Experiment 2, Arabic and Japanese listeners were more accurate than Persian listeners in identifying the length categories in the "other" unknown language as well as in their own L1. The difference in the listeners' perceptual performance between the two experiments supports the view that long-term L1 representations may be invoked to a greater extent in the identification than discrimination test. The present results highlight the importance of selecting the appropriate test for assessing cross-language speech perception. © *2011 Acoustical Society of America*.
[DOI: 10.1121/1.3531801]

PACS number(s): 43.71.Hw [MAH] Pages: 989–998

I. INTRODUCTION

Previous linguistic experience including first language (L1) undeniably affects information processing and individual's subsequent language learning. The language to be acquired (or the target language) may include linguistic aspects that are conceptually familiar and/or unfamiliar to the learner. If the objective of second (L2)/foreign language learning is to gain proficiency with limited time and resources, one may naturally wonder which language is "worth learning" especially if the choice is up to each learner rather than a forced one. The decision on which language should be learned may be, to a large extent, driven by learners' interests. However, it may be useful to consider the outcome of language learning from other perspectives such as efficiency of learning.

Two natural human languages are likely to have something in common no matter how linguistically distant or unrelated they may be. Although linguistic similarity and learners' previous linguistic experience may be independent of each other, it may nevertheless be possible to think of the two as complementary. If the two language systems are similar to a large extent, it may be possible to compensate for the lack of experience. On the other hand, it is certainly possible to gain experience with a language that is completely dissimilar to one's own. Would it be the case that the greater the commonality between the two languages, the less learning is required?

In this study, the role of general vs specific experience with vowel length contrasts was examined using Arabic and Japanese as the target languages. The perception of three groups of listeners differing in their L1 was compared. The first two groups, native Arabic and Japanese speakers, were naïve to each other's language. The third group, native Persian speakers, was included to assess how their general knowledge of Arabic and experience of learning it as a foreign language at school on the one hand and the lack or limited experience with vowel length contrasts in L1 on the other hand might influence their discrimination and identification of the short vs long vowels in Arabic (familiar/known language) and Japanese (unfamiliar/unknown language). In general, Persian-speaking people learn Arabic at school for six years on average. Moreover, they are culturally familiar with the language in their daily life. Japanese speakers, on the other hand, typically have no experience with Arabic, but they possess firm vowel length categories in their L1. The difference in listeners' previous linguistic experience outlined above led to the following question: Is it Persian listeners' foreign language learning experience or Japanese listeners' experience with phonemic vowel length contrasts that is more beneficial in processing short and long vowels in Arabic?

Standard Arabic has three vowels /i(ː) a(ː) u(ː)/ and vowel length is phonemic (Nasr, 1960; Mitleb, 1984; Norlin, 1985; Thelwall and Sa'Adeddin, 1990; Alghamdi, 1998; Ryding, 2005; Most *et al.*, 2008). For example, /bir/ "charity" contrasts with /biːr/ "well, fountain" and /sur/ "to be happy with" contrasts with /suːr/ "fence." Long vowels are twice as long as their short counterparts (Norlin, 1985; Alghamdi, 1998; Hajjar, 2005; Ryding, 2005). While some cross-dialectal differences in vowel quality (F1, in particular) have been reported (e.g., Cowan, 1970; Drozdík, 1973; Alghamdi, 1998; de Jong and Zawaydeh, 2002; Ryding, 2005), short and long vowels in different dialects are clearly distinguished durationally with the average short-to-long ratio ranging from 0.4 to

a)Author to whom correspondence should be addressed. Electronic mail: kimiko.tsukada@mq.edu.au

0.45 (see, however, Mitleb, 1984 for a larger ratio in Jordanian Arabic at 0.65). Variations in vowel quality due to vowel length and neighboring sounds have also been noted in several studies (Norlin, 1985; Alghamdi, 1998; de Jong and Zawaydeh, 2002; Ryding, 2005). Phonemic long vowels were shown to occupy a more peripheral portion of the acoustic vowel space than their short counterparts (Norlin, 1985).

Japanese has five vowels /i(ː) e(ː) a(ː) o(ː) u(ː)/ and vowel length is phonemic (e.g., Vance, 1986; Shibatani, 1990; Hirata, 2004a; Hirata and Tsukada, 2009) as in Arabic. For example, /e/ "picture, painting" contrasts with /eː/ "yes" and /sa/ "difference" contrasts with /saː/ "well, let's see." Dialectal variations are primarily realized in vowel quality (Shibatani, 1990). Long Japanese vowels tend to be more than twice as long as their short counterparts when spoken at a fixed speaking rate. Within the same speaking rate, durational overlap between the two length categories is negligible (Hirata, 2004a). Compared to Arabic, vowel quality differences between short and long vowels appear to be small in Japanese (Tsukada, 1999; Hirata and Tsukada, 2009). While longer vowels tend to be more peripheral than shorter vowels in many languages, the extent to which vowel quality difference is involved in encoding a length contrast appears to be language-specific (Engstrand and Krull, 1994; Tsukada and Roengpitya, 2008). Perceptual relevance of length contrasts can be attested by the observation that young Japanese infants discriminate /mana/ and /maːna/ accurately at 9.5 (but not at 7.5) months. This is in contrast with difficulty non-native adult learners face in perceiving Japanese vowel length contrasts (e.g., Hirata, 2004a,b; Hirata *et al.*, 2007; Hirata and Kelly, 2010; Tajima *et al.*, 2008).

Modern Persian has six vowels /i e æ ɒ o u/ and it appears that vowel length is in a transition state and no longer contrastive (Majidi and Ternes, 1991; Toosarvandani, 2004) unlike classical Persian. For example, /dir/ "late" contrasts with /del/ "heart" and /dur/ "far" contrasts with /dor/ "round" (Bijankhan and Nourbakhsh, 2009). Not surprisingly, some of the common pronunciation difficulties encountered by Farsi-speaking English learners include vowel length confusion in word pairs such as "sheep" and "ship," "cart" and "cut," "fool" and "full" among others (Swan and Smith, 2001).

Vowel length categorization which is influenced by various factors (e.g., duration of neighboring segments, number of syllables within a word, speaking rate) has been investigated by various investigators (Nooteboom and Doodeman, 1980; Pind, 1986) and neutralization of phonemic length categories has been reported for many languages of the world (e.g., Lahiri *et al.*, 1987; Myers, 2005; Myers and Hansen, 2007) including Arabic (Cowan, 1970; Norlin, 1985) and Japanese (Kubozono, 2002). However, it appears that this phenomenon is more restricted in Japanese than in Arabic (Kubozono, 2002).

Naturally, one would assume that Persian listeners who do not need to categorize vowels according to their length in L1 pay less attention and are less sensitive to durational variations of vowels than do Arabic and Japanese listeners who need to make lexical decisions based on whether the vowel is short or long. On the other hand, one might question whether existing L1 length categories might interfere with the processing of the length categories in an unknown language especially if the two sets of categories are not perceptually equated with each other and category boundaries need to be re-aligned.

Prominent theories of cross-language and L2 perception such as the perceptual assimilation model (PAM) (Best, 1995) and the speech learning model (SLM) (Flege, 1995, 2003) focus on the perceived relationship between segmentally characterized sounds across the two phonological systems. To give a simple example, the PAM would predict that, if two sounds in one system are perceptually assimilated to two distinct categories in another system, those two sounds would be accurately discriminated.

Two sounds can be minimally contrastive in the prosodic domain (as in tone languages such as Chinese) just as well as they can be a contrast segmentally. If we assume that these two types of contrasts are equivalent in terms of categoricity (Hallé *et al.*, 2004), then it may be possible to hypothesize that, for Arabic and Japanese speakers, listening to vowel length contrasts in each other's language may be analogous to French or German speakers listening to the English /l-r/ contrast or Thai speakers listening to word-final stop place contrasts in the English minimal pairs such as "cap" and "cat" (Tsukada and Roengpitya, 2008). In other words, these are the contrasts that are expected to be discriminated with high accuracy without prior experience with the target language. However, by extension of the SLM, if the Arabic and Japanese vowel length contrasts are similar, but not identical in phonetic realization, more learning or cognitive processing may be required for the Arabic and Japanese listeners compared to the Persian listeners who may enjoy unbiased perception of vowel duration.

We need to be cautious in interpreting the role of prior experience with phonological categories, as there are some studies that suggest that L1-to-L2 mappings at the phonological level may not accurately predict listeners' behavior. One study on French listeners' perception of English approximants /w j r l/ found that they had some perceptual difficulties with English /r/ and tended to hear it as /w/-like despite their L1 having the /r/ sound (Hallé *et al.*, 1999). This perceptual confusion was explained in terms of articulatory-phonetic differences between the English and French /r/ (i.e., central approximant in English and uvular fricative in French). Thus, at the subphonemic level, the /r/ sounds may not be "shared" between the two languages. Furthermore, recent studies on the effect of L1 on the cross-language perception of stop place contrasts (Tsukada, 2006; Tsukada and Roengpitya, 2008) found that despite having experience with unreleased word-final stops in their L1, Australian English speakers were not as accurate as native Thai listeners in discriminating unreleased Thai stop contrasts (e.g., /t/ vs /k/ which is phonetically [t˺] vs [k˺]). Findings analogous to these have been reported in the perception of lexical tones, as well (Wang *et al.*, 2004; Francis *et al.*, 2008). For example, Wang *et al.* (2004) showed that despite their experience with tones in L1, native Norwegian listeners did not process the novel Mandarin lexical tones in the way comparable to native Mandarin listeners and proficient Mandarin-English bilingual listeners. These findings suggest that both phonological learning and

experience with the native variety of acoustic phonetic cues are essential for optimal cross-language speech perception.

At present, it is unclear if the two categories contrasting segmentally are comparable in perceptual saliency to those contrasting prosodically, as the above-mentioned theories do not make predictions about prosodic contrasts. The present study is concerned with Arabic and Japanese, both of which use vowel duration contrastively. Cross-language perception studies focusing on languages other than English are still limited. By assessing typologically unrelated languages such as Arabic and Japanese, it is possible to enhance our understanding of natural spoken language processing by humans. Two experiments, AXB discrimination as described in the procedure and two-alternative forced-choice identification tests, have been conducted to assess the extent to which (1) Arabic and Japanese listeners utilize their L1 knowledge and (2) Persian listeners utilize their foreign language learning experience in cross-language speech perception.

II. EXPERIMENT 1: AXB DISCRIMINATION TEST

A. Method

1. Speakers

Native speakers of Arabic and Japanese participated in the recording sessions lasting between 45 and 60 min. Eight (4 males, 4 females) native Japanese speakers with a mean age of 25.4 years were recorded in a sound-treated room in the Department of Linguistics, University of Kobe, Japan. With the exception of one participant who had lived in the United States for 11 months, none of them had lived overseas for an extended period of time. 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 these speakers. None of these Japanese speakers participated in the study as listeners.

Seven (4 males, 3 females) native Arabic speakers with a mean age of 40 years were recorded in the audio–visual recording studio and the Centre for Language Sciences (CLaS) recording studio at Macquarie University, Sydney. They were all born in Arabic-speaking countries and had lived in Sydney, Australia, for 12 years on average at the time of recording. The Arabic speakers' dialectal backgrounds were Lebanese ($n=4$) and Saudi ($n=3$). Vowel length is phonemic to all these speakers. With the exception of one Lebanese speaker, none of these Arabic speakers participated in the study as listeners. This Lebanese speaker's speech was not included in the stimuli and, furthermore, there were more than three months between the time he was recorded and the time he participated in the perception experiments. All speakers were paid $30 (or the equivalent amount in Japanese yen) for their participation.

2. Stimuli

The speakers read CV_1C words in Arabic and $CV_1CV_2(n)$ words in Japanese (where V_1 was either short or long in each language). The words recorded for this study are shown in Appendixes A and B. While this discrepancy in the syllable structure was not desirable, the occurrence of the CVC sequence is limited in Japanese unlike Arabic. Furthermore, given that vowel length is neutralized in open syllables including the word-final position in Arabic (Thelwall and Sa'Adeddin, 1990), it was not an option to use Arabic CVCV words to match the syllable structure across the two languages. Some of the Japanese word pairs used in this study differed by more than just vowel length, i.e., pitch pattern and the extent of vowel devoicing. This inconsistency needs to be rectified in future research. However, the effect of pitch accent on the production and perception of vowel length is known to be negligible for native speakers of Japanese (Beckman, 1986; Minagawa-Kawai *et al.*, 2002). Availability of relevant information for Arabic vowels in published research appears limited if any. Care was taken to use the words that were likely to be familiar to all the native speakers. Achieving these criteria resulted in some variation in the consonantal context.

All speakers read the target words twice in isolation and once in the short carrier sentence. This decision was made to keep the recording sessions from becoming excessively long, which would introduce fatigue and may result in misreading. The target words produced in a carrier sentence were used as stimuli. This was to include tokens that reflect natural variation in speech sounds. The carrier sentences were انا أكتب ____ أيضاً "I write ___ as well" for Arabic (/ʔana ʔaktub ___ ʔaida/ in broad IPA (International Phonetic Association) transcription) and 次は ____ と言います。"Next I say the word ___" for Japanese (/tsugiwa _____ to iimasu/ in broad IPA transcription). The test words (all real words) were presented visually to each speaker in randomized orders on the computer screen one word at a time. All words were written using appropriate Arabic or Japanese script which explicitly encodes vowel length in the standard orthography. Thus, 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.

Some basic acoustic characteristics of the stimuli used in this study are given in Table I. The EMU speech database system was used for phonetically labelling the speech segments of interest and the formant values were calculated in the R statistical environment. The beginning and end of each vowel token was identified by visual inspection of wideband spectrograms and time domain waveforms. F1 and F2 values were measured and tracked at the vowel midpoint. Crucially, short and long categories are clearly separated in both languages with small standard errors.

The recorded speech materials were digitized at 44.1 kHz using CoolEdit and the target words were segmented and stored in separate files. Prior to the presentation in the perception experiments, each sound file was normalized to 50% of the peak amplitude following the procedures used in previous research (e.g., Guion *et al.*, 2000, Aoyama *et al.*, 2004; Flege and MacKay, 2004). The recorded words were arranged in a triad such that the first and third tokens always represented different words as described below.

3. Listeners

Seven native (3 males, 4 females) Arabic, 15 (2 males, 13 females) Japanese, and 11 (8 males, 3 females) Persian

TABLE I. Mean duration (in ms), F1 and F2 (in Hz) of the Arabic and Japanese vowels used in the stimuli. The standard errors are in parentheses. F1 and F2 were measured at the vowel midpoint.

	Female			Male		
	Duration	F1	F2	Duration	F1	F2
Arabic						
/i/	103 (6)	340 (20)	2278 (69)	87 (4)	394 (13)	1821 (33)
/iː/	235 (8)	296 (10)	2693 (24)	173 (7)	321 (10)	2264 (40)
/a/	114 (4)	836 (38)	1799 (21)	96 (5)	654 (27)	1552 (33)
/aː/	307 (17)	845 (18)	1916 (17)	190 (8)	681 (17)	1558 (35)
/u/	109 (8)	386 (24)	1078 (39)	88 (5)	406 (17)	1168 (84)
/uː/	212 (16)	298 (16)	1056 (91)	183 (9)	329 (9)	848 (18)
Japanese						
/i/	80 (5)	448 (45)	2483 (64)	69 (5)	377 (57)	2269 (68)
/iː/	195 (14)	340 (20)	2740 (55)	176 (14)	302 (10)	2433 (53)
/e/	88 (10)	479 (20)	2557 (86)	79 (6)	382 (14)	2218 (48)
/eː/	182 (14)	510 (16)	2511 (49)	159 (9)	390 (14)	2289 (69)
/a/	67 (3)	798 (45)	1478 (30)	65 (5)	690 (40)	1373 (81)
/aː/	190 (12)	874 (66)	1429 (44)	166 (9)	789 (15)	1283 (29)
/o/	84 (4)	490 (18)	1319 (55)	75 (4)	399 (11)	1221 (47)
/oː/	181 (12)	461 (27)	944 (55)	156 (9)	405 (17)	920 (45)
/u/	59 (8)	375 (19)	1862 (85)	61 (5)	354 (15)	1625 (78)
/uː/	179 (13)	425 (24)	1645 (75)	152 (8)	337 (9)	1491 (64)

listeners participated in the AXB discrimination experiment. Their mean ages were 29.0, 31.9, and 26.6 years, respectively. They had lived in Sydney, Australia, for 10.4, 3.2, and 1.2 years, respectively, when they participated in the study. Six of the Arabic listeners were postgraduate students or staff members at Macquarie University and one was a member of the community in Sydney. She was referred to the study by a friend who saw the advertisement. Their dialectal backgrounds were Egyptian ($n=1$), Iraqi ($n=1$), and Lebanese Arabic ($n=5$). All the Japanese listeners were students or staff members of Department of Linguistics at Macquarie University with the exception of one participant who was a vocational student. They originally came from different parts of Japan including Aichi, Hokkaido, Kanagawa, Miyagi, Tochigi, Tokyo, and Yamaguchi prefectures. The Persian listeners were postgraduate students at Macquarie University or University of Technology, Sydney. All of them confirmed having some knowledge of Arabic from their secondary education.

The participants were tested individually in a session lasting between 30 and 40 min in the Speech Perception Lab in the CLaS or a quiet room at Macquarie University. They heard the stimuli at self-selected comfortable level using the high-quality headphones (Bose Quiet Comfort 2 acoustic noise canceling headphones). According to self-report, all listeners had normal hearing and did not have any language deficiency in their L1. They were paid $20 for their participation.

4. Procedure

The participants' perception of vowel length contrasts in Arabic and Japanese was assessed in a categorical AXB discrimination test, a method frequently employed in previous speech perception research (e.g., Harnsberger, 2001; Wayland and Guion, 2003; Hallé *et al.*, 2004; Højen and Flege, 2006). In the AXB test, the first (A) and third (B) tokens always come from different categories and the listeners need to decide whether the second token (X) belongs to the same category as A (e.g., "sit_2"-"sit_1"-"$seat_3$") or B (e.g., "hit_3"-"$heat_1$"-"$heat_2$") (where the subscripts indicate different speakers). Within a trial, three tokens were always spoken by different speakers of the same sex. Thus, X was never physically identical to either A or B. This was to ensure that the listeners focus on relevant phonetic characteristics that group two tokens as members of the same category without being distracted by audible but phonetically irrelevant within-category variation (e.g., in voice quality). For the half of the trials, the position of X was occupied by one of the short vowels and for the other half, it was occupied by one of the long vowels. All possible AB combinations (i.e., AAB, ABB, BAA, and BBA) were tested.

The presentation of the stimuli and the collection of perception data were controlled by the UAB software (Smith, 1997). The stimuli were blocked by language (Arabic, Japanese) and the order of presentation was counterbalanced across listeners. A total of 100 trials for Arabic and 125 trials for Japanese were presented. The first four and five trials in each language were for practice and were not analyzed. Thus, there were 32 trials for each of the three Arabic vowels and 24 trials for each of the five Japanese vowels.

The listeners were given two options ("1 = 2," "2 = 3") to choose from on the computer screen. They were asked to click on the first option if they thought that the first two tokens in the AXB sequence are the "same" and to click on the second option if they thought that the last two tokens are the "same." The inter-stimulus interval in all trials was set at 0.5 s. They had up to 3 s to respond to each trial which could be played only once. When the listeners did not respond within 3 s, they were prompted to click on a button. No replay was allowed and they needed to click on either button to proceed to the next trial. They were asked to guess if they were unsure and no feedback was provided during the experimental sessions.

B. Results

The raw percent correct scores were log-transformed before being submitted to statistical analyses. Table II shows the mean correct discrimination scores for Arabic and

TABLE II. Mean discrimination scores (log-transformed) for Arabic and Japanese vowel length contrasts as a function of vowel type by three groups of listeners. The standard errors are in parentheses.

Language	Vowel	Arabic	Japanese	Persian
Arabic	/i/-/iː/	1.98 (0.01)	1.91 (0.01)	1.91 (0.02)
	/a/-/aː/	1.97 (0.01)	1.89 (0.01)	1.90 (0.02)
	/u/-/uː/	1.94 (0.01)	1.90 (0.01)	1.92 (0.01)
Japanese	/i/-/iː/	1.89 (0.02)	1.99 (0.00)	1.90 (0.02)
	/e/-/eː/	1.88 (0.02)	2.00 (0.00)	1.84 (0.01)
	/a/-/aː/	1.88 (0.02)	1.98 (0.01)	1.87 (0.02)
	/o/-/oː/	1.87 (0.01)	1.99 (0.00)	1.84 (0.03)
	/u/-/uː/	1.87 (0.01)	1.99 (0.00)	1.89 (0.03)

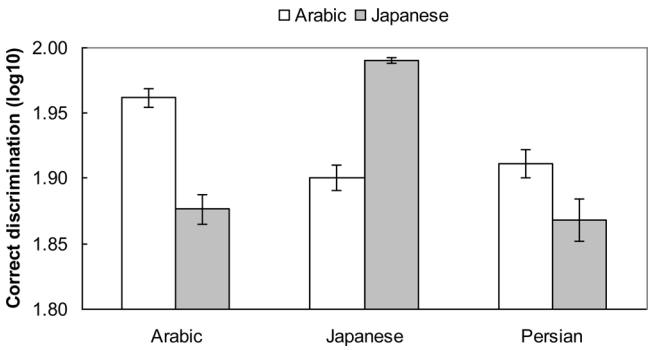

FIG. 1. Mean discrimination scores (log-transformed) for the Arabic and Japanese stimuli averaged across vowel types. The brackets enclose ± 1 standard error.

Japanese vowel length contrasts as a function of vowel type. With the possible exception of the Persian listeners who discriminated the high Japanese vowel contrasts (/i/-/iː/, /u/-/uː/), more accurately than non-high vowel contrasts, the vowel type did not exert appreciable influence on the listeners' discrimination accuracy. As the effect of vowel was not of primary concern in this study, it will not be discussed further.

Figure 1 shows the mean correct discrimination scores averaged across vowels for the three groups of listeners. The Arabic and Japanese listeners showed an opposite discrimination pattern. They discriminated vowel length contrasts in their L1 much more accurately than in the unknown language (1.97 vs 1.88 for the Arabic listeners and 1.99 vs 1.9 for the Japanese listeners). The Persian listeners' discrimination was significantly more accurate for the Arabic (1.91) than Japanese (1.87) vowel length contrasts.

Looking at the individual listeners, all Arabic and Japanese listeners were more accurate in their L1 than in the unknown language. All Persian listeners except for one were more accurate in Arabic than in Japanese, so learning Arabic as a foreign language may have helped them to some extent. However, this learning effect was apparently not enough to differentiate them from the Japanese listeners who have no experience with Arabic, but are familiar with vowel length contrasts in their L1.

The listeners' discrimination scores were submitted to a two-way analysis of variance (ANOVA) in which Group (Arabic, Japanese, Persian) was a between-subjects factor and Stimulus Language (Arabic, Japanese) was a within-subjects factor. Both main effects of Group and Stimulus Language were significant [G: $F(2,30)=11.9$, $p<0.001$, S: $F(1,30)=4.3$, $p<0.05$]. There was a significant two-way interaction [$F(2,30)=67.5$, $p<0.001$], which can be clearly observed in Fig. 1.

The simple effect of Group was significant for both Arabic and Japanese stimuli. For the Arabic vowel length contrast, the Arabic listeners were significantly more accurate than the Japanese and Persian listeners who did not differ from each other [$F(2,30)=9.2$, $p<0.001$]. For the Japanese vowel length contrast, the Japanese listeners were significantly more accurate than the Arabic and Persian listeners who did not differ from each other [$F(2,30)=44.9$, $p<0.001$]. These results are clearly discernible in Fig. 1. Thus, the Arabic and Japanese listeners who have experience with phonological vowel length contrasts in their L1s did not show an advantage over the Persian listeners whose L1 does not utilize vowel duration phonologically. The simple effect of Stimulus Language was significant for all three groups of listeners [$F(1,30)=13.8-83.2$, $p<0.001$]. As was mentioneded above, the Arabic and Persian listeners were more accurate in Arabic than in Japanese and the Japanese listeners showed the reverse pattern of results. This suggests that familiarity with length categories in their L1 did not transfer optimally in the processing of the unknown language.

C. Discussion

All listeners including the Persian listeners who do not utilize vowel length contrastively in their L1 performed the vowel length discrimination task well above chance (log-transformed score: 1.7). However, despite the expectation that the Arabic and Japanese listeners may generalize their L1 phonological experience to the unknown language with vowel length contrasts, their discrimination accuracy was optimal only in their L1 in the AXB discrimination experiment. Neither Arabic nor Japanese listeners showed any advantage over the Persian listeners, for whom vowel length is no longer phonemic, in discriminating the vowel length contrasts in the unknown language.

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 in their categorization of the English vowel contrast /i/-/ɪ/ (as in "heat" vs "hit") (Bohn and Flege, 1990). Further, Ingram and Park (1998) who tested the perception of the difficult English /l/-/r/ contrast by Korean and Japanese learners of English reported the overall discrimination scores of 82% and 75%, respectively, which is well above chance. Of course, the study by Ingram and Park and the present study differ in various ways. Particularly, the two studies crucially differ in the kind of phonemic contrasts and the listeners involved (novel/unknown vs non-novel/known contrasts and listeners with vs without knowledge of the target language). Thus, a strict comparison is not intended. Nevertheless, the Korean and Japanese listeners' discrimination scores are very similar to the Arabic and Japanese listeners' discrimination scores when they listened to the unknown language [a raw score of 75% (log-transformed score of 1.88) for the Arabic listeners responding to Japanese and a raw score of 80% (log-transformed score of 1.9) for the Japanese listeners responding to Arabic]. Considering that the /l/-/r/ contrast is notoriously difficult to produce and perceive, particularly, for Japanese learners of English, the listeners in the Ingram and Park (1998) study performed the task relatively well.

Why is it that the Arabic and Japanese listeners who are familiar with vowel length contrasts did not take advantage of their L1 knowledge to process vowel length contrasts in the unknown language? Could this be related to the task they performed? In Experiment 2, the same three groups of listeners participated in a forced-choice identification test.

III. EXPERIMENT 2: TWO-ALTERNATIVE FORCED-CHOICE IDENTIFICATION TEST

A. Method

1. Stimuli

A subset of the Arabic ($n = 144$) and Japanese ($n = 180$) tokens presented in Experiment 1 was used in the identification test. These materials were presented with distractors such that the listeners would hear 200 tokens per language. The stimuli were blocked by language and the order of presentation was counterbalanced across listeners.

2. Listeners

Five (3 males, 2 females) Arabic and 11 (1 male, 10 females) Japanese listeners who participated in Experiment 1 identified the length category of the vowels in the Arabic and Japanese words in Experiment 2. All 11 (8 males, 3 females) Persian listeners participated in this identification test. Their mean ages were 29.7, 32.4, and 26.6 years for the Arabic, Japanese, and Persian groups, respectively. They had lived in Sydney, Australia, for 8.6, 2.9, and 1.2 years, respectively, when they participated in the study.

Typically, the listeners participated in the two experiments on the same day with a short break in between. All the listeners completed the discrimination test first, then the identification test which lasted between 30 and 40 min. Thus, the entire experimental session took approximately 1.5 h when the listeners participated in both discrimination and identification tests. They were tested individually in the Speech Perception Lab in the CLaS at Macquarie University using the same equipment as in the discrimination test. All of them reported normal hearing and had no history of language problems in their L1s. They were paid $20 for their participation.

3. Procedure

The listeners were given two options ["XXX (X)," "X (X)"] to choose from on the computer screen. They were asked to click on the first option if they thought that the "word" they heard had a long vowel and to click on the second option if they thought that the "word" they heard had a short vowel. The presentation of the stimuli and the collection of perception data were controlled by the UAB software (Smith, 1997). The stimuli were blocked by language (Arabic, Japanese) and the order of presentation was counterbalanced across listeners.

For the identification task, the listeners were given 10 (5 Arabic, 5 Japanese) practice trials with correct answers (i.e., as was intended by the native speakers who produced the words). This was intended to give the listeners an opportunity to "calibrate" the two length categories. The listeners had up to 3 s to respond to each token. When the listeners did not respond within 3 s, the message "Would you like to hear the previous stimulus again?" appeared on the screen. Thus, unlike Experiment 1, they were allowed to listen to each token as many times as they wished. However, they were cautioned not to use this option excessively, because listening many times would not always help and that they might get confused. They were asked to guess if they were unsure and no feedback was provided during the experimental sessions.

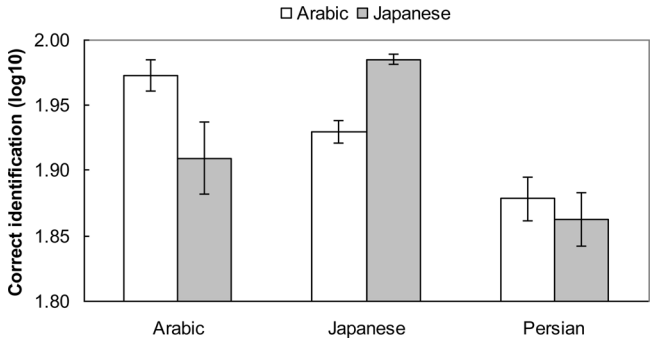

FIG. 2. Mean identification scores (log-transformed) for the Arabic and Japanese stimuli averaged across vowel types. The brackets enclose ± 1 standard error.

B. Results

The raw percent correct scores were log-transformed before being submitted to statistical analyses. Figure 2 shows the mean correct identification scores averaged across vowels for the three groups of listeners. At first glance, the overall pattern of results looks similar to the one in the discrimination test in Experiment 1 (Fig. 1). The Arabic and Japanese listeners were highly accurate in identifying the short or long vowels in their L1 but less so in the unknown language (1.97 vs 1.91 for the Arabic listeners and 1.99 vs 1.93 for the Japanese listeners). However, in this task, the difference in their perception accuracy between the L1 and unknown language was smaller than in the discrimination test. Although this trend did not reach statistical significance ($p = 0.09$), this held true for 3 out of 5 Arabic and 7 out of 11 Japanese listeners (with one Japanese listener showing no difference). The Persian listeners' identification scores did not significantly differ for Arabic (1.88) and Japanese (1.86). As for the individual listeners, all Arabic and Japanese listeners performed better in their L1 than in the unknown language just as in the AXB discrimination test. Nine out of 11 Persian listeners performed better in Arabic than in Japanese.

To gain further insight into the listeners' response patterns, the results were rearranged and plotted in Fig. 3 according to the length category (i.e., short vs long) of the vowel in each stimulus token presented. All three groups of listeners made more identification errors when they heard the long Arabic vowels than when they heard the short Arabic vowels (Arabic: 1.99 vs 1.95, Japanese: 1.97 vs 1.86, Persian: 1.91 vs 1.84). This was particularly noticeable in the non-native groups. There was no such bias according to vowel length for the Japanese stimuli. This may mean that, even when the vowel was intended to be long by the native Arabic speakers, it was not perceived as such by the non-native listeners for whom the lexical status of the stimuli would be ambiguous.

Alternatively, it may be relatively difficult to make short vs long judgments on the vowel in monosyllabic words presented in isolation, because the listeners need to decide whether the target vowel is short or long without referring to the sound outside the syllable. In the case of the Japanese

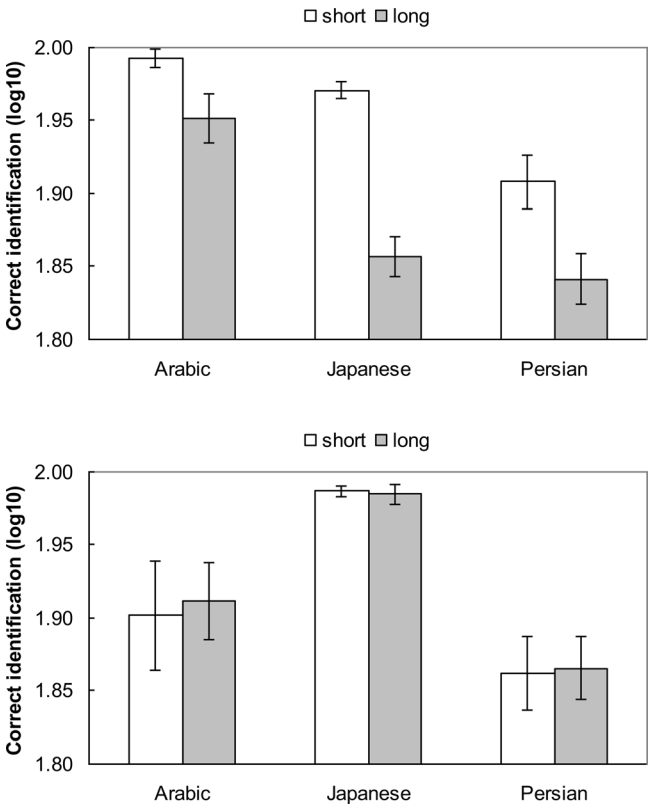

FIG. 3. Mean identification scores (log-transformed) for the Arabic (top) and Japanese (bottom) stimuli according to the length category. The brackets enclose ± 1 standard error.

stimuli, the listeners (non-native listeners, in particular) may have been aided by the presence of the second vowel in the $CV_1CV_2(n)$ words which always belonged to the short category with the exception of the "fukee" vs "fuukee" pair. In other words, relative durational differences between the two adjacent syllables may be very useful to the listeners especially when they are responding to "non-lexical" items.

The listeners' identification scores were submitted to a three-way ANOVA with Group (Arabic, Japanese, Persian) as a between-subjects factor and Stimulus Language (Arabic, Japanese) and Length (short, long) as within-subjects factors. The main effects of Group and Length, but not Stimulus Language, were significant [G: $F(2,24)=12.3$, $p<0.001$, L: $F(1,24)=11.8$, $p<0.01$]. There were significant Group × Stimulus Language [$F(2,24)=25.8$, $p<0.001$] and Stimulus Language × Length interactions [$F(1,24)=32.7$, $p<0.001$]. A three-way interaction did not reach significance.

The simple effect of Group can be observed for both Arabic and Japanese in Fig. 3, but the pattern of between-group difference clearly depended on the length category in Arabic. For the identification of the length category of the Arabic vowels, the Arabic listeners were most accurate for both short (1.99) and long (1.95) vowels. However, the Japanese listeners (1.97) did not significantly differ from the Arabic listeners when they listened to the short Arabic vowels and, crucially, both Arabic and Japanese listeners were significantly more accurate than the Persian listeners (1.91) [$F(2,24)=9.3$, $p<0.001$]. As for the long Arabic vowels, the Arabic listeners were significantly more accurate than the Japanese (1.86) and Persian (1.84) listeners who did not differ from each other [$F(2,24)=9.1$, $p<0.01$]. For the identification of the length category of the Japanese vowels, the Japanese listeners were most accurate for both short (1.99) and long (1.99) vowels. However, only the difference between the Japanese and Persian listeners (1.86) reached significance when the listeners responded to the short Japanese vowels [$F(2,24)=10.3$, $p<0.001$]. In other words, the Arabic listeners (1.9) did not significantly differ from either Japanese or Persian listeners. As for the long Japanese vowels, the Japanese listeners were significantly more accurate than the Arabic (1.91) and Persian (1.87) listeners who did not differ from each other [$F(2,24)=14.0$, $p<0.001$]. These results may be due to the small number of the Arabic listeners ($n=5$) who took part in the identification test and their relatively large standard errors in identifying the Japanese short vs long vowels.

C. Discussion

We observed that, in general, the Arabic and Japanese listeners identified the length category of vowels in both their L1 and unknown language better than did the Persian listeners. Particularly, their advantage over the Persian listeners was measurable when they listened to the short vowels in the unknown language. Crucially, the difference between the Japanese and Persian groups reached statistical significance when they listened to the short Arabic vowels. Furthermore, the difference between Arabic and Japanese listeners did not reach statistical significance when they listened to the short Japanese vowels. These findings are in contrast with the results obtained in the AXB discrimination test where neither the Arabic nor Japanese listeners showed an advantage over the Persian listeners in discriminating vowel length contrasts in the unknown language. Thus, the Arabic and Japanese listeners differed to a greater extent from the Persian listeners in the identification than in the discrimination test.

One needs to be cautious in comparing the results from Experiments 1 and 2, because the observed results may be attributable to the methodological differences in the two experiments such as the response time and conditions allowed to the listeners. While this is not desirable, it is unlikely to be the determining factor of the listeners' performance. Recall that the cutoff time was 3 s in both Experiments 1 and 2. In Experiment 1, when 3 s were up, the listeners were prompted to click on a button without any chance of replay. They simply had to guess at that point. In Experiment 2, on the other hand, when 3 s were up, the listeners were allowed to replay the previous stimulus. Although this may seem like a substantial difference, when the percentages of trials which were not responded to within 3 s are compared across the experiments, the difference is negligible as shown in Table III. The Arabic and Japanese listeners' distributional pattern of responses is practically identical across the experiments. As such, different results of the two experiments are unlikely to have been caused by the presence or absence of the replay option beyond 3 s.

Furthermore, when the results from the two experiments were directly compared in a three-way ANOVA (Group, Language, Task), only the main effect Group [$F(2,24)=12.5$, $p<0.001$] and two- and three-way interactions involving the Group factor were significant [G × L:

TABLE III. Number of trials that was not responded to within 3 s as a function of three groups of listeners. The percentages are in parentheses.

	Language	Arabic	Japanese	Persian
Experiment 1	Arabic (96 triads)	21/672 (3%)	32/1440 (2%)	82/1056 (8%)
(Discrimination)	Japanese (120 triads)	62/840 (7%)	5/1800 (0%)	162/1320 (2%)
Experiment 2	Arabic (200 tokens)	34/1000 (3%)	61/2200 (3%)	194/2200 (9%)
(Identification)	Japanese (200 tokens)	69/1000 (7%)	16/2200 (1%)	163/2200 (7%)

$F(2,24)=54.9$, $p<0.001$, G × T: $F(2,24)=4.0$, $p<0.05$, G × L × T: $F(2,24)=6.4$, $p<0.01$]. Neither the main effect of Task nor the Language × Task two-way interaction was significant. While the Group × Task interaction reached significance, this was apparently due to the fact that the Arabic and Japanese listeners' scores were higher in the identification task and the Persian listeners did the opposite and scored higher in the discrimination task. These results suggest that it was the *pattern* of Group × Language interaction that was differentially affected by the task, resulting in the significant three-way interaction.

Taken together, the results of the two experiments are more consistent with the idea that different task demands are imposed in the discrimination and identification tests (Ingram and Park, 1998). It is possible that having firmly established L1 length categories does not predict listeners' discrimination accurately in the AXB discrimination task when the language is unknown. Listeners may simply employ online processing without referring to their L1 categories. In the identification task, the Japanese (and, to a less extent, Arabic) listeners possibly referred to the internal, long-term L1 categories, which helped them to decide the length category of the incoming stimulus of the unknown language.

IV. GENERAL DISCUSSION

The primary objective of this study was to investigate the role of different types of linguistic experience in cross-language speech perception. Three groups of listeners differing in their L1s (Arabic, Japanese, and Persian) listened to Arabic and Japanese words minimally contrasting in vowel length. It was predicted that the listeners would differ in how they discriminate and identify short vs long vowels in Arabic and Japanese according to their L1 backgrounds.

For the Arabic and Japanese listeners listening to each other's language, it would be inappropriate to characterize vowel length contrasts as "novel," as both of these languages utilize vowel duration phonologically. Of particular interest was the perception of Arabic vowel length contrasts by the native speakers of Japanese and Persian, two groups for whom the Arabic contrast was not "novel," but for different reasons. The Japanese listeners had experience categorizing vowel length phonemically in their L1 but had no knowledge of Arabic. The Persian listeners, on the other hand, had experience with the Arabic language in their culture and education, but their L1 may be losing or possibly has lost vowel length contrasts. As it turned out, both groups of listeners perceived the Arabic short and long vowels less accurately than the native Arabic listeners. However, the extent to which the non-native listeners diverged from the native Arabic listeners differed in the two experiments and the Japanese listeners outperformed the Persian listeners in the identification (but not discrimination) of the short Arabic vowels. As for the Japanese length contrasts, the Japanese listeners outperformed the two non-native groups in both their discrimination and identification with one exception which is that the Arabic listeners were not significantly less accurate than the Japanese listeners in identifying the short Japanese vowels.

As briefly mentioned above, it was observed that the extent to which the listeners benefited from their L1 phonological experience in perceiving vowels in the unknown language depended on the experimental task which they performed. The Japanese and, to a less extent, Arabic listeners identified the length category of the vowels in each other's language better than did the Persian listeners. The present results provide support to Ingram and Park (1998) who observed different performance profiles for the identification and discrimination tests by Japanese and Korean learners of English. The results from their study and the present one converge to support the view that long-term L1 representations may be invoked to a greater extent in the identification test, which may engage listeners to be in the linguistic mode of processing, than in the discrimination test.

It is unclear whether two sounds contrasting segmentally are equivalent in categoricity or perceptual salience to two sounds contrasting prosodically. For example, a distinction between short and long Japanese vowels is susceptible to speaking rate (Hirata, 2004a) and length contrasts may be more easily manipulated than segmental contrasts. It is also important to point out that the number of length categories would be undoubtedly less than the number of vowel and/or consonant categories. Both Arabic and Japanese have only two length categories, short and long. There are languages with more length categories, but such languages are rare (McRobbie-Utasi, 2007; Remijsen and Gilley, 2008; Remijsen and Manyang, 2009). Assuming, as suggested by Remijsen and Gilley (2008), that the phonetic space for vowel length contrasts is fixed and does not stretch or compress, perhaps the human auditory system cannot cope with many levels of differentiation for length categories that uni-dimensionally (or predominantly) rely on the temporal domain.

In order to assess the validity of the current models of cross-language and L2 speech perception (e.g., PAM, SLM) with respect to the results obtained in this study, how listeners identify the length category in the unknown languages must be established empirically. As mentioned in the Introduction, neither the PAM nor SLM makes predictions about non-segmental contrasts. While the extent of perceived similarity between the L1 and the target sounds is a useful and important concept in these models, with increasing evidence that just having distinct

length (or tone) categories in one's L1 is not sufficient for optimal cross-language speech perception (Wang *et al*., 2004; Francis *et al*., 2008), a different approach may be needed to account for the processing of the contrasts with limited "degrees of freedom." Given that neither Arabic nor Japanese listeners were as accurate as native listeners in perceiving vowel length in the unknown language (i.e., mileage from L1 was limited), experience with specific (i.e., phonemic vowel length) AND native phonetic realization of sounds may be essential in accurate cross-language speech perception. The importance of native experience was also highlighted in a study that examined stop place contrasts by Australian English and Thai listeners (Tsukada and Roengpitya, 2008). The present results demonstrate that this applies to the perception of vowels as well as consonants and also to prosodic as well as segmental contrasts.

Although the benefit of L1 was not fully exhibited in the cross-language perception of the Arabic and Japanese listeners in this study, it might still be informative to examine the perception of listeners who have no experience with Arabic or Japanese, but who are familiar with length contrasts in their L1 (e.g., Finnish or Thai). It may be the case that the unbalanced combination of L1 and unknown language was disruptive to the Arabic and Japanese listeners and they were unduly disadvantaged in the unknown language in this study. If they repeat the same tests with a different set of languages, both of which unknown to them, they may utilize different processing strategies and show better performance that might reflect their L1 phonology.

Further, it would be necessary to understand acoustic characteristics of the stimuli that yielded high vs low discrimination/identification accuracy rates, i.e., "easy/clear" vs "hard/unclear" trials to understand if and to what extent listeners' L1 is reflected in the way they organize their perceptual space. It is also possible that the degree of coarticulation with surrounding consonants may be different for Arabic and Japanese vowels which, in turn, may interact with native and non-native listeners' perception of length contrasts.

V. CONCLUSION

The Arabic and Japanese listeners' response patterns for vowel length contrasts in each other's language suggest that experience with L1 phonological categories may not be the best predictor for the discrimination accuracy in an unknown language. In the forced-choice identification task, on the other hand, the listeners appear to utilize their long-term L1 categories to make short vs long judgments on the vowel of the incoming stimulus. The different pattern of results across the two experiments became even more insightful in comparison with the third group, the Persian listeners, who do not use vowel duration contrastively in their L1. This comparison gave us more confidence in the view that the Arabic and Japanese listeners may refer to their long-term L1 representations to a greater extent in the identification than discrimination test. This view was expressed by Ingram and Park (1998) who tested the different contrast (English /l/-/r/) and listener groups (Korean and Japanese L2 learners) from the present study. Taken together, the present results highlight the importance of selecting the appropriate test for assessing cross-language speech perception and approaching the questions by including listeners from diverse language backgrounds.

ACKNOWLEDGMENTS

This study was funded by the Macquarie University New Staff Grant (Reference No. 9200800377) and the Macquarie University Research Development Grant (Reference No. 9200900585). The author thanks Professor Haruo Kubozono (currently the National Institute for Japanese Language and Linguistics) 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 CLaS, Department of Linguistics, Macquarie University. The author also thanks participants who represent diverse language backgrounds and the associate editor and three anonymous reviewers for their helpful comments on the previous versions of this manuscript.

APPENDIX A: ARABIC WORDS USED IN THIS STUDY

Vowel	Short	Long
/i/-/iː/	دِن din "large jug"	دين diin "religion"
	رِق riq "slavery"	ريق riiq "saliva"
	سِب sib "swear"	سيب siib "leave"
	زِر zir "button"	زير ziir "large jar"
/a/-/aː/	بَن ban "coffee (beans)"	بان baan "to appear"
	دَم dam "blood"	دام daam "to keep on"
	سَب sab "to curse"	ساب saab "to leave"
	شَب shab "young man (colloquial)"	شاب shaab "young man (standard)"
/u/-/uː/	دُب dub "bear"	دوب duub "melt"
	حُر hur "free"	حور huur "women with beautiful eyes"
	رُح ruh "go"	روح ruuh "spirit"
	سُم sum "poison"	سوم suum "negotiate the price"

APPENDIX B: JAPANESE WORDS USED IN THIS STUDY

Vowel	Short	Long
/i/-/iː/	汁 shiru "soup"	シール shiiru "sticker"
	来た kita "came"	聞いた kiita "listened"
	ビル biru "building"	ビール biiru "beer"
/e/-/eː/	席 seki "seat"	世紀 seeki "century"
	世間 seken "world"	政権 seeken "government"
	駅 eki "station"	鋭気 eeki "spirit"
/a/-/aː/	下部 kabu "lower part"	カーブ kaabu "curve"
	角 kado "corner"	カード kaado "card"
	後 ato "later"	アート aato "art"
/o/-/oː/	横 yoko "side"	洋子 yooko "girl's name"
	女子 joshi "girl"	上司 jooshi "boss"
	のど nodo "throat"	濃度 noodo "density"
/u/-/uː/	主人 shujin "(my) husband"	囚人 shuujin "prisoner"
	父兄 fukee "parents"	風景 fuukee "scenery"
	手記 shuki "memoirs"	周期 shuuki "cycle"

Alghamdi, M. M. (**1998**). "A spectrographic analysis of Arabic vowels: A cross-dialectal study," J. King Saud Univ. Eng. Sci. **10**, 3–24.
Aoyama, K., Flege, J. E., Guion, S. G., Akahane-Yamada, R., and Yamada, T. (**2004**). "Perceived phonetic dissimilarity and L2 speech learning: The case of Japanese /r/ and English /l/ and /r/," J. Phonetics **32**, 233–250.
Beckman, M. E. (**1986**). *Stress and Non-Stress Accent* (Foris Publications, Dordrecht), pp. 145–200.
Best, C. T. (**1995**). "A direct realist view of cross-language speech perception," in *Speech Perception and Linguistic Experience: Issues in Cross-language Research*, edited by W. Strange (York, Baltimore), pp. 171–203.
Bijankhan, M., and Nourbakhsh, M. (**2009**). "Voice onset time in Persian initial and intervocalic stop production," J. Int. Phonetic Assoc. **39**, 335–364.
Bohn, O.-S., and Flege, J. E. (**1990**). "Interlingual identification and the role of foreign language experience in L2 vowel perception," Appl. Psycholinguist. **11**, 303–328.
Cowan, W. (**1970**). "The vowels of Egyptian Arabic," Word **26**, 94–100.
de Jong, K. and Zawaydeh, B. (**2002**). "Comparing stress, lexical focus, and segmental focus: Patterns of variation in Arabic vowel duration," J. Phonetics **30**, 53–75.
Drozdík, L. (**1973**). "The vowel system of Egyptian colloquial Arabic," Asian African Stud. **9**, 121–127.
EMU speech database system, http://emu.sourceforge.net/ (Last viewed August 5, 2010).
Engstrand, O., and Krull, D. (**1994**). "Durational correlates of quantity in Swedish, Finnish and Estonian: Cross-language evidence for a theory of adaptive dispersion," Phonetica **51**, 80–91.
Flege, J. E. (**1995**). "Second language speech learning: Theory, findings, and problems," in *Speech Perception and Linguistic Experience: Issues in Cross-language Research*, edited by W. Strange (York, Baltimore), pp. 233–277.
Flege, J. E. (**2003**). "Assessing constraints on second-language segmental production and perception," in *Phonetics and Phonology in Language Comprehension and Production, Differences and Similarities*, edited by A. Meyer and N. Schiller (Mouton de Gruyter, Berlin), pp. 319–355.
Flege, J. E., and MacKay, I. R. A. (**2004**). "Perceiving vowels in a second language," Stud. Second Lang. Acquis. **26**, 1–34.
Francis, A. L., Ciocca, V., Ma, L., and Fenn, K. (**2008**). "Perceptual learning of Cantonese lexical tones by tone and non-tone language speakers," J. Phonetics **36**, 268–294.
Guion, S. G., Flege, J. E., Akahane-Yamada, R., and Pruitt, J. C. (**2000**). "An investigation of current models of second language speech perception: The case of Japanese adults' perception of English consonants," J. Acoust. Soc. Am. **107**, 2711–2724.
Hajjar, N. (**2005**). *Living Arabic in Context: An Introductory Course* (Bisan, Bayrut), pp. 20.
Hallé, P., Best, C. T., and Levitt, A. (**1999**). "Phonetic vs. phonological influences on French listeners' perception of American English approximants," J. Phonetics **27**, 281–306.
Hallé, P., Chang, Y.-C., and Best, C. T. (**2004**). "Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners," J. Phonetics **32**, 395–421.
Harnsberger, J. D. (**2001**). "On the relationship between identification and discrimination of non-native nasal consonants," J. Acoust. Soc. Am. **110**, 489–503.
Hirata, Y. (**2004a**). "Effects of speaking rate on the vowel length distinction in Japanese," J. Phonetics **32**, 565–589.
Hirata, Y. (**2004b**). "Training native English speakers to perceive Japanese length contrasts in word versus sentence contexts," J. Acoust. Soc. Am. **116**, 2384–2394.
Hirata, Y., and Kelly, S. (**2010**). "Effects of lips and hands on auditory learning of second-language speech sounds," J. Speech Lang. Hear. Res. **53**, 298–310.
Hirata, Y., and Tsukada, K. (**2009**). "Effects of speaking rate and vowel length on formant frequency displacement in Japanese," Phonetica **66**, 129–149.
Hirata, Y., Whitehurst, E., and Cullings, E. (**2007**). "Training native English speakers to identify Japanese vowel length contrast with sentences at varied speaking rates," J. Acoust. Soc. Am. **121**, 3837–3845.
Højen, A., and Flege, J. (**2006**). "Early learners' discrimination of second-language vowels," J. Acoust. Soc. Am. **119**, 3072–3084.
Ingram, J. C. L., and Park, S.-G. (**1998**). "Language, context, and speaker effects in the identification and discrimination of English /r/ and /l/ by Japanese and Korean learners," J. Acoust. Soc. Am. **103**, 1161–1174.
Kubozono, H. (**2002**). "Temporal neutralization in Japanese," in *Laboratory Phonology VII*, edited by C. Gussenhoven and N. Warner (Mouton de Gruyter, Berlin), pp. 171–201.
Lahiri, A., Schriefers, H., and Kuijpers, C. (**1987**). "Contextual neutralization of vowel length: Evidence from Dutch," Phonetica **44**, 91–102.
Majidi, M.-R., and Ternes, E. (**1991**). "Persian (Farsi)," J. Int. Phonetic Assoc. **21**, 96–98.
McRobbie-Utasi, Z. (**2007**). "The instability of systems with ternary length distinctions: The Skolt Saami evidence," in *Saami Linguistics*, edited by I. Toivonen and D. Nelson (John Benjamins, Amsterdam), pp. 167–205.
Minagawa-Kawai, Y., Maekawa, K., and Kiritani, S. (**2002**). "Effects of pitch accent and syllable position in identifying Japanese long and short vowels: Comparison of English and Korean speakers," J. Phonetic Soc. Jpn. **6**, 88–97.
Mitleb, F. (**1984**). "Vowel length contrast in Arabic and English: A spectrographic test," J. Phonetics **12**, 229–235.
Most, T., Levin, I., and Sarsour, M. (**2008**). "The effect of modern standard Arabic orthography on speech production by Arab children with hearing loss," J. Deaf Stud. Deaf Educ. **13**, 417–431.
Myers, S. (**2005**). "Vowel duration and neutralization of vowel length contrasts in Kinyarwanda," J. Phonetics **33**, 427–446.
Myers, S., and Hansen, B. B. (**2007**). "The origin of vowel length neutralization in final position: Evidence from Finnish speakers," Natural Lang. Ling. Theory **25**, 157–193.
Nasr, R. T. (**1960**). "Phonemic length in Lebanese Arabic," Phonetica **5**, 209–211.
Nooteboom, S. G., and Doodeman, G. J. N. (**1980**). "Production and perception of vowel length in spoken sentences," J. Acoust. Soc. Am. **67**, 276–287.
Norlin, K. (**1985**). "Acoustic analysis of vowels and diphthongs in Cairo Arabic," Stud. African Ling. **9**, 238–244.
Pind, J. (**1986**). "The perception of quantity in Icelandic," Phonetica **43**, 116–139.
R statistical environment, http://cran.r-project.org/ (Last viewed June 28, 2008).
Remijsen, B., and Gilley, L. (**2008**). "Why are three-level vowel length systems rare? Insights from Dinka (Luanyjang dialect)," J. Phonetics **36**, 318–344.
Remijsen, B., and Manyang, C. A. (**2009**). "Illustrations of the IPA: Luanyjang Dinka," J. Int. Phonetic Assoc. **39**, 113–124.
Ryding, K. C. (**2005**). *A Reference Grammar of Modern Standard Arabic* (Cambridge University Press, Cambridge), pp. 25–26.
Shibatani, M. (**1990**). *The Languages of Japan* (Cambridge University Press, Cambridge), pp. 160–163.
Smith, S. (**1997**). *User Manual for UAB Software* (Department of Rehabilitation Sciences, University of Alabama at Birmingham (UAB)).
Swan, M., and Smith, B. (**2001**). *Learner English: A Teacher's Guide to Interference and Other Problems* (Cambridge University Press, New York), pp. 179–180.
Tajima, K., Kato, H., Rothwell, A., Akahane-Yamada, R., and Munhall, K. (**2008**). "Training English listeners to perceive phonemic length contrasts in Japanese," J. Acoust. Soc. Am. **123**, 397–413.
Thelwall, R., and Sa'Adeddin, M. A. (**1990**). "Arabic," J. Int. Phonetic Assoc. **20**, 37–39.
Toosarvandani, M. D. (**2004**). "Vowel Length in Modern Farsi," J. Royal Asiatic Soc. **14**, 241–251.
Tsukada, K. (**1999**). *An Acoustic Phonetic Analysis of Japanese-Accented English* (Ph.D. thesis, Macquarie University, Sydney, Australia).
Tsukada, K. (**2006**). "Cross-language perception of word-final stops in Thai and English," Bilingualism: Lang. Cognit. **9**, 309–318.
Tsukada, K. (**2009**). "An acoustic comparison of vowel length contrasts in Arabic, Japanese and Thai: Durational and spectral data," Int. J. Asian Lang. Process. **19**, 127–138.
Tsukada, K., and Roengpitya, R. (**2008**). "Discriminatioin of English and Thai words ending with voiceless stops by native Thai listeners differing in English experience," J. Int. Phonetic Assoc. **38**, 325–347.
Vance, T. J. (**1986**). *An Introduction to Japanese Phonology* (State University of New York Press, Albany), pp. 9–16.
Wang, Y., Behne, D. M., Jongman, A., and Sereno, J. (**2004**). "The role of linguistic experience in the hemispheric processing of lexical tone," Appl. Psycholinguist. **25**, 449–466.
Wayland, R., and Guion, S. (**2003**). "Perceptual discrimination of Thai tones by naïve and experienced learners of Thai," Appl. Psycholinguist. **24**, 113–129.