



# Cross-language vowel perception and production by Japanese and Korean learners of English

**John C. L. Ingram and See-Gyoon Park**

*Department of English, University of Queensland, Brisbane, 4072, Australia*

*Received 26th January 1996, and in revised form 4th April 1997*

---

This paper investigates the roles of language-specific phonological learning and inherent phonetic contrastiveness in the perception of non-native vowels. Native speakers of Korean and Japanese, at two levels of English language experience, were assessed on the perception and production of Australian English monophthongal non-back vowels: /i: ɪ e æ a:/. Prototypicality ratings, or perceived similarities of the foreign vowels to their nearest native (L1) phonemic targets, were also examined, to assess models of cross-language vowel perception. Korean is of interest because of a recent phonological merger of two front vowels (/e/ and /ɛ/), which has produced a generation split among speakers of Seoul dialect above and below 45–50 years of age (Hong, 1991). The present study is the first reported case of how a phonemic merger, resulting in cross-generation differences within a speech community, can influence speakers' perception and production of non-native vowels. The effects of L1 phonological learning on vowel perception were also observed in the tendency of the Japanese, but not the Korean listeners, to normalize tokens of non-native vowels for speaker-dependent durational variation, consistent with the respective phonological roles of vowel length in Japanese and Korean.

© 1997 Academic Press Limited

---

## 1. Introduction

Current theories of non-native vowel perception in second language learning have built upon the generally accepted premise, often referred to as the Contrastive Analysis Hypothesis (CAH: Lado, 1957), that phonological contrasts in the native language exert a strong influence over L2 vowel perception, at least in the initial stages of second language learning (Flege, 1992; Best, 1995). The complementary question of how first language (L1) phonological learning affects the learner's accommodation to non-phonological sources of phonetic variation in second language (L2) speech has only recently begun to be systematically explored (Strange, Bohn, Trent, McNair & Bielec, 1996; Magnuson & Akahane-Yamada, 1996).

It is generally accepted that "attunement" of the auditory system (Best, 1995) to phonological contrasts of the ambient language in first language learning entails the

acquisition of perceptual strategies for adapting to speaker and situation-dependent sources phonetic variation, which might otherwise mask phonological contrasts in speech. Non-native listeners are known to have difficulty accommodating to speaker variation in perceiving non-native phonological targets, and the perceptual strategies that they adopt in the face of speaker variation depend, in part, upon how well they have established phonetically robust categorization criteria for the phonological distinction in question. Thus, Magnuson & Akahane-Yamada (1996) found that only Japanese listeners with native-English like accuracy in /r-1/ discrimination successfully accommodated to speaker-related phonetic variation in a "mixed" listening condition, where the speaker varied unpredictably from item to item within a block of trials.

Complementary to the above, it has also been found that exposure to tokens from a variety of speakers enhances learning of non-native phonemic contrasts (Lively, Logan & Pisoni, 1993), possibly by focusing learners' attention more to the level of phonological contrasts, and less upon characteristics of the speaker or other contingent auditory properties of the stimuli. It is not known to what extent perceptual normalization strategies developed in first language learning are transferable to second language learning. One aim of the present study is to provide exploratory data on this question, by comparing Japanese and Korean listeners' perception of Australian English front vowels, with particular attention to the /e-æ/ contrast (*beg-bag*), which both learner groups find relatively difficult.

Japanese and contemporary Korean present contrasting situations with respect to the L1 status of the features of vowel length and vowel quality, relevant to the perception of Australian English front vowels. Comparisons among appropriately selected groups of Japanese and Korean learners, at different levels of prior exposure to Australian English, should provide useful information on how prior L1 learning of these features and their attendant mechanisms of perceptual attunement affect non-native vowel perception.

Two influential models of cross-language vowel perception (Best, 1995; Flege, 1995) are driven by consideration of how the foreign vowels are assimilated to native phoneme categories. (Both models deal more generally with speech perception, but we restrict consideration here to vowel perception.) Flege (1987) proposed the notion of a *perceptual equivalence* class to account for the purported effect that some non-native vowels are more readily accommodated than others by second language learners. Certain L2 sounds are sufficiently phonetically different from their nearest L1 targets to be perceived as "new" or "foreign", whereas others are sufficiently close to L1 targets to be classified as "similar", though not identical to some L1 phonemic target. Flege (1987) was not very explicit about the phonetic criteria which determine an L2 vowel's classification as "similar to an L1 target" or as "new", but subsequently (Flege & Munro, 1994) assumed that phonetic distance between vowels could be related directly to distances between point targets in the Bark-scaled  $F_1$ - $f_0$ / $F_2$ - $F_1$  plane (Syrdal & Gopal, 1986).

Whereas Flege's model deals only with the case where one vowel falls within the boundaries of an L1 target, as opposed to one that does not, Best's model generates predictions of relative perceptual discriminability on a range of L1 assimilation possibilities. Discrimination of two non-native sounds will be maximized where each is assimilated to a different native phoneme category (a two-category contrast). Where both sounds are assimilated to a single phonemic category, but where one

sound constitutes a closer phonetic match to that category than the other, moderate discriminability is predicted. However, where two foreign sounds are equally good candidates for a single category, discrimination will be poor. Uncategorizable contrasts in which neither of the sounds can be readily assimilated to a native category will also be poorly discriminated.

Best's perceptual assimilation model (PAM) also differs from Flege's model in explicitly asserting a *gestural* basis for perceived similarities and discrepancies between native and non-native speech sounds, consistent with the Direct Realist view, in which the gestural components of distal speech objects are extracted from the speech signal, without intervening inferential processing from an auditory/acoustic base representation. Studies of cross-modal speech perception have shown that both adults and infants integrate auditory and visual cues into a gesturally coherent percept (see Best, 1995 and Vihman, 1996 for reviews). While these findings support the Direct Realist model, the precise mechanism by which such integration is achieved remains to be clarified.

In this paper, we present new data on the role of prior L1 phonological learning on the integration of temporal and vowel quality features in the perception of "foreign vowels". The vowel contrasts under consideration (high to low front vowels, and the low central vowel /a/ in Australian English) did not yield differential predictions for an auditory *vs.* gestural basis for perceptual similarities, such as might, for example, be generated by examining the reactions of learners whose L1 possesses both high back rounded and unrounded vowels (/u/ and /ʊ/) to the Australian English high central rounded /ɜ/ (see Ingram & Pittam, 1987<sup>1</sup>). Thus, one important source of differential predictions between Flege and Best's models could not be assessed in the data presented here (however, preliminary analysis of Korean and Japanese learners' perceptual responses to Australian English back vowels appears to favor a gestural over a psycho-acoustic basis for categorization of non-native vowels). More generally, although both Flege's and Best's models yield predictions concerning Japanese and Korean listeners' perceptual responses to Australian English vowels, the predictions are similar for the vowels investigated here (see Section 2.1.).

The plan of this paper is as follows. Section 2 provides background information on the vowel systems of Australian English, Japanese, and Korean, necessary to interpret the three experiments reported here. Section 3 reports an experiment in which Japanese and Korean listeners, varying in age and levels of experience in English, made forced choice identifications of five Australian English vowels /i, ɪ, e, æ, a:/ embedded in /h\_d/ carrier words, recorded by two male native speakers of Australian English. The identification task required listeners to adapt to any speaker-dependent phonetic variation that might affect identification of the foreign vowels. Acoustic variation the stimulus items, crucial for understanding listeners' responses is also described. Section 4 reports measurements on listeners' productions of the foreign vowels tested in the previous experiment, in order to assess whether perceptual differences between the language groups were carried over into production. Section 5 reports a further experiment which examined

<sup>1</sup> For example, formant measurements of their vowel productions indicated that immigrant Vietnamese schoolchildren initially assimilated Australian English /ɜ/ to Vietnamese /u/, rather than to Vietnamese /ʊ/, thereby suggesting that the lip rounding feature was perceptually salient, despite the consideration that [u] is acoustically closer to [ɜ].

TABLE I. Monophthongal vowels of Australian English. Vowels used in the Identification experiment appear in **bold**

Tense			Lax		
Phonemic	IPA	Word	Phonemic	IPA	Word
Symbol	Symbol		Symbol	Symbol	
<b>/i/</b>	<b>[i:]</b>	<b>heed</b>	/ɪ/	<b>[ɪ]</b>	<b>hid</b>
(/eɪ/)	[eɪ]	hay) <sup>1</sup>	/e/	<b>[e]</b>	<b>head</b>
			/æ/	<b>[æ]</b>	<b>had</b>
<b>/a/</b>	<b>[a:]</b>	<b>hard</b>	/ʌ/	[a]	Hudd
/ɜ:/	[ɜ:]	herd	/ə/	[ə]	ahead
/ɔ:/	[o:]	hoard	/ɒ/	[ɒ]	hod
/u/	[u:]	who'd	/ʊ/	[ʊ]	hood

<sup>1</sup> Clearly a diphthong in Australian English.

native-phoneme categorization judgements and goodness ratings for the stimuli used in the identification experiment. Taken together, the results of the three experiments provide strong evidence for the differential impact of L1 phonological experience on the perception of non-native vowels, and reveal how the transfer of normalization strategies for phonological feature detection in L1 can affect perceptual processing in an L2 listening situation. Section 6 summarizes the main findings and offers a perspective on how phonological and phonetic levels of speech processing may interact in non-native vowel perception.

## 2. Australian, Japanese, and Korean vowels

Australian English has 12 monophthongal vowels, five of which are conventionally labelled “tense” and seven as “lax”. They are shown in Table I, in phonemic transcription for ease of comparison with other English dialects, and in IPA symbols, to indicate their approximate vowel quality and quantity. Some of the tense vowels (/i/ and /u/) are often quite diphthongal in pronunciation and the tense counterpart of /e/ (*let*), /eɪ/ (*late*) is regarded as a diphthong. Vowel length is arguably the most prominent phonetic correlate of the tense-lax contrast, though, as the IPA transcriptions in Table I indicate, there are significant vowel quality differences for all but one of the tense-lax pairs /a-ʌ/ (*card-cud*), where the phonetic contrast is almost solely one of length. The low front vowel [æ] (*cad*) is problematical in this classification. Its inherent duration places it among the long vowels (Fletcher & McVeigh, 1992), but its exclusion from open syllable position places it phonologically among the “lax” vowels. Australian English /æ/ is generally regarded as being the only lax vowel not to have a “tense” counterpart, though this has been a topic of intermittent debate for many years (see Durie & Hajek, 1994; Ingram, 1995).

Japanese has two sets of five monophthongal vowels (/i/, /e/, /a/, /o/, and /u/) which are contrastively short or long (mono-moraic or bi-moraic). Unlike Australian English, Japanese long and short vowels do not differ substantially in quality,

though the geminate forms are slightly more peripheral than their non-geminate counterparts (Keating & Huffman, 1984). Hence, Japanese may be said to have a phonological contrast of vowel length, whereas Australian English has a tense-lax contrast, of which vowel length is a major phonetic component. The notable feature of the Japanese vowel system with respect to the five front vowels examined in the present study is that Japanese lacks a phonemic target in the low-front region of the vowel chart (i.e., [æ]).

The vowel system of present day Korean (Seoul dialect) is complicated by two on-going sound changes, one being the loss of contrastive vowel length, and the other the merger of a mid-high front vowel /e/ and a mid-low front vowel, usually transcribed /ɛ/. Korean is traditionally described as possessing phonologically short and long vowels. However, the length contrast seems to be disappearing in the Seoul dialect; it is preserved only in the speech of older speakers and only in the most formal speech style (citation forms). Magen & Blumstein (1993) found that, of 59 long/short Korean minimal pairs, only four were consistently produced with a statistically significant difference in vowel duration by all four of their speakers, three of whom were between 23 and 26 years of age. They concluded that vowel length in Korean “is no longer a productive contrast”. This is consistent with concerns expressed by educators that the younger generation does not know when to pronounce vowels long or short. Unlike Japanese Kana, vowel length is not represented orthographically in the Korean Hangeul script, although hyphens or colons for long vowels may be found in dictionary entries as a guide to pronunciation.

Apparently independently of the loss of phonological vowel length, the Seoul dialect has also undergone a merger of the mid-high front vowel /e/ and the low-mid front vowel /ɛ/. This change has occurred post World War II, under the influence of the Southern Kyungsang and Cholla dialects and has produced an age-related dialect cline. From formant measurements of vowel productions, Hong (1991) found that the amount of differentiation between tokens of /e/ and /ɛ/ varied directly with the age of her Seoul dialect speakers. The results of perceptual tests showed that the ability to discriminate minimal pairs differing in these two front vowels also varied directly with the age of the listener.<sup>2</sup>

### 2.1. Predicted transfer effects

It is possible to predict how listeners will respond on the basis of L1 experience to specific dimensions of phonetic contrast. Consider firstly the dimension of quantity of vowel length. As vowel length forms a basic and stable phonological contrast in the Japanese vowel system, it is to be expected that Japanese listeners will have acquired effective attunement or “normalization” strategies for extracting this feature from the speech signal, in the face of speaker-related speech rate variation—strategies which they may or may not deploy in the perception of (tense-lax) non-native targets. The status of vowel length is equivocal in Korean. For younger speakers at least, it no longer appears to be an active phonological contrast

<sup>2</sup> But interestingly, discrimination scores ranged between 0% correct and 72% correct in the oldest listener. Further sociolinguistic investigation showed that the merger constitutes a variable, in the Labovian sense, showing stratification by speech style, social class, and sex of speaker.

and consequently mechanisms for its extraction as a perceptual target in L1 may not be well developed and thus not available for transfer to L2.

Turning to the dimension of vowel quality, and setting aside for the moment the question of whether quality distinctions should be treated in gestural or auditory-acoustic terms, we note that Japanese lacks a native vowel quality target that could serve as a model for Australian English [æ]. Older Korean speakers of the conservative Seoul dialect, who preserve the /e-ε/ contrast in other speech, possess a native phonemic target with which to assimilate the Australian [æ] (namely, Korean /ε/). But younger Koreans may be less favorably situated in this respect than even the Japanese, because their merged native /e-ε/ phonological target will presumably encompass a wide range of variation in the vowel space (sufficient to span the range of tokens of [e] and [ε] that conservative speakers keep phonologically distinct).

In terms of Flege's model of non-native vowel perception, Japanese listeners may be expected to treat Australian English [æ] as a "new" or novel vowel that falls outside the phonetic boundaries of native Japanese /e/ or /a/, whereas younger Koreans will incorporate Australian [e] and [æ] tokens into a single merged /e-ε/ phoneme category. Flege's model, based upon vowel quality considerations alone, establishes the following cline of perceivability for the critical case of the Australian English /e-æ/ contrast: older Koreans ><sup>3</sup>Japanese > younger Koreans. Best's model yields the same ranking of perceivability of the contrast for the three groups of listeners. According to her model, the Australian English /e-æ/ contrast will be a two-category assimilation for older Koreans (and therefore most highly discriminable), an uncategorized-*vs.*-categorized assimilation for Japanese listeners (and therefore quite highly discriminable), and a single-category or category-goodness difference type assimilation for younger Koreans (and therefore of moderate to poor discriminability).

Flege's model makes the additional prediction with respect to the learnability of the foreign /e-æ/ contrast that the Japanese should benefit more from equivalent L2 experience with the contrast than the Koreans.

However, these predictions are of limited value because they take no account of likely group differences in response to the vowel length dimension, discussed previously. Moreover, the problematical status of Australian English /æ/ as a long-short vowel makes predictions based on its duration hazardous. Consequently, the perceptual assimilation of non-native vowels to native phonemic categories can only be established *a posteriori*, from phoneme category goodness ratings, obtained from the perspective of the listeners' first language.

### 3. Experiment 1: Non-native vowel identification

#### 3.1. Subjects

Five groups of non-native listeners and a native-English control group participated in Experiment 1. Two of the non-native groups were relatively inexperienced in English. The Japanese and the Korean inexperienced subjects (JI and KI groups,

<sup>3</sup> ">" means "find the contrast more perceivable"

TABLE II. Subject groups and characteristics

Group	Description	Age (years)	<i>n</i>	Sex	Occupation	English experience
KI	Korean: less English experience	24–28 $\bar{X} = 26$	4	M	University students	Less than 1 year in English environment
KE	Korean: more English experience	25–38 $\bar{X} = 32$	4	M	3 Teachers of Korean	More than 5 years in English environment
KO	Korean: older monolingual	51–58 $\bar{X} = 55$	4	M	2 Businessmen 1 Professor 3 Announcers	Monolingual. No English experience
JI	Japanese: less English experience	21–23 $\bar{X} = 22$	4	M	University students	Less than 1 year in English environment
JE	Japanese: more English experience	33–44 $\bar{X} = 38$	4	M	Teachers of Japanese	More than 5 years in English environment

see Table II) were all students, aged in their twenties, and had been resident in Australia less than 12 months at the time of testing. Two groups with more English experience (JE and KE groups) were somewhat older, aged in their thirties; all had been resident in Australia for more than 5 years and all but two were teachers of Korean or Japanese in Australian universities (see Table II). Subjects in these four groups used English on a daily basis. The difference in English fluency between the inexperienced (JI and KI) and the experienced (JE and KE) subjects was considerable.

A fifth group (KO) comprised four non-English speaking Koreans aged in their fifties, who spoke a conservative (pre-merger) variety of the Seoul dialect. Three were radio announcers and one a professor of History. This group was tested in Korea. Although these subjects would inevitably have had some contact with English, none had resided in an English-speaking environment or had any conversational fluency in English. A sixth, control group of eight native speakers of Australian English, six females and two males, ranging from 22 to 49 years of age, also took the forced-choice vowel identification test.

### 3.2. Method and materials

#### 3.2.1. Speakers

Two phonetically trained Australian male speakers, in their forties, recorded three tokens each of the words: “heed, hid, head, had, hard”. Items were elicited in citation form on the understanding that they would be used in a listening test for non-native speakers. The tape recorded items were subsequently digitized at a 20 kHz sampling rate and 16 bit quantization for measurement of their acoustic characteristics, and for preparation of a listening tape to be used in the identification experiment. Both speakers spoke a similar variety of Standard Australian English,

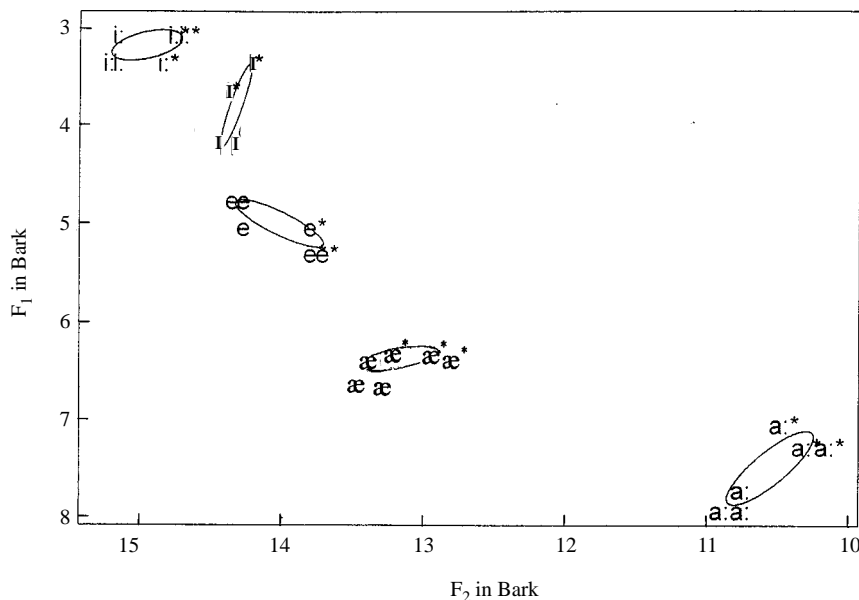
though their personal voice characteristics were readily distinguishable, at least to native listeners.

### 3.2.2. Stimulus characteristics

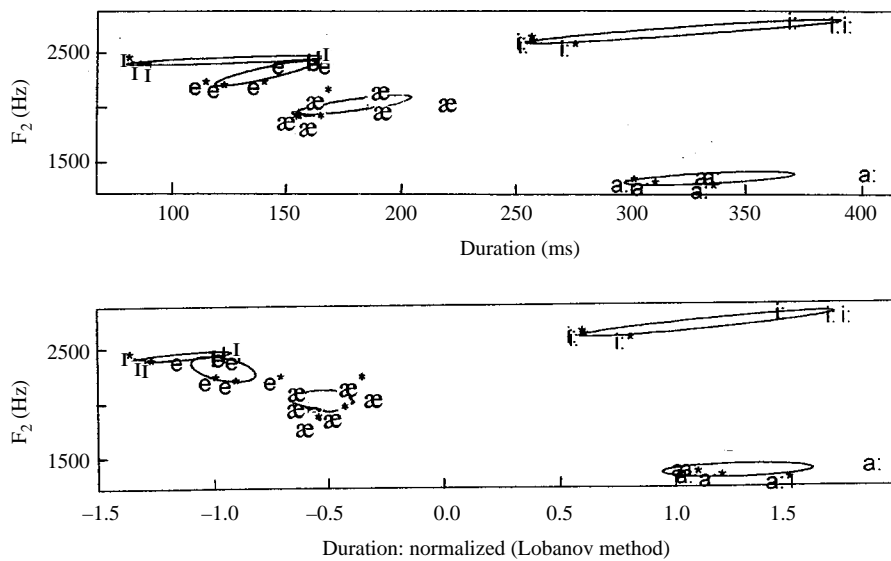
First and second formant frequencies ( $F_1$  and  $F_2$ ) and vowel duration measurements were obtained from the 30 digitized master tokens using a PC speech workstation (Sensimetrics, Speechstation<sup>TM</sup>). Formant measurements were performed by locating a cursor at the mid-point of the vowel, in the center of a formant band on the spectrogram. The FFT spectrum of the signal, located at the cursor, was simultaneously displayed, and the LPC smoothed spectrum of the same signal frame could be superimposed on the FFT spectrum. The smoothed LPC spectrum was used as a check only on the formant frequencies estimated by eye from the spectrogram and FFT spectrum.

Fig. 1 shows a plot of the  $F_1$  and  $F_2$  frequencies (in Bark) for the 30 master tokens produced by the two Australian English speakers. The tokens of Speaker B have been superscripted with an asterisk (\*) to distinguish the two speakers' tokens; one standard deviation ellipses have been drawn for each vowel. Although lower  $F_2$  values for Speaker B suggest that his vowel productions are more back, the formant values for individual vowels clustered into separate regions of the  $F_1$ - $F_2$  space, with no overlap.

The temporal characteristics of the Australian English tokens are shown in Fig. 2. The upper portion of the graph shows the duration of vowel tokens plotted against their  $F_2$  values (to provide spatial separation of the vowel classes; Speaker A produced consistently longer tokens than speaker B(\*) particularly on the vowels [i:]



**Figure 1.** Mid-vowel first and second formant frequencies of test vowels produced by two speakers of Australian English (Speakers A and B). Each speaker produced three tokens of each vowel; starred tokens were produced by Speaker B, unstarred by Speaker A.



**Figure 2.** Temporal characteristics of target Australian vowels plotted against F<sub>2</sub> frequency (in Hz). Upper plot = duration unnormalized; lower plot = duration normalized (by Lobanov method). Starred tokens (\*) from Speaker B; unstarred, from Speaker A.

and [ɪ]). A repeated measures two-way Analysis of Variance (ANOVA), with the factors of Vowel type and Speaker, yielded highly significant main and interaction effects (Vowel:  $F(4, 20) = 219.2$ ,  $p < 0.001$ ; Speaker  $F(1, 20) = 114.2$ ,  $p < 0.001$ ; Vowel  $\times$  Speaker:  $F(4, 20) = 7.45$ ,  $p < 0.001$ ).

The lower portion of Fig. 2 shows the effects of normalizing the stimuli for speaker-dependent differences in vowel duration using the Lobanov (1971) method, which converts raw duration measurements to z-scores calculated about speaker means and standard deviations. As expected, the normalization enhanced separation of the vowels by intrinsic duration, particularly between the lax (short) and tense (long) classes. However, because the normalization method does not take into account the interaction of speaker differences with vowel type, some speaker-dependent temporal variation remains.

In summary, the stimulus items carried little speaker-dependent vowel quality variation, as assessed by F<sub>1</sub>-F<sub>2</sub> frequency measurements, but considerable speaker-dependent variation in vowel duration; this variation was reduced, but not eliminated, by a standard statistical normalization procedure.

### 3.2.3. Identification test procedure

A stimulus tape was prepared from the master tokens, taking what was judged to be the best exemplar of each word from each of the two speakers. The listening tape contained 50 trials in which the 5 test words spoken by the 2 speakers were each repeated 5 times in a randomized order of presentation, subject to the restriction that no token followed itself on successive trials. On a given trial, each item was presented 3 times, over headphones, with an inter-stimulus interval of 1.5 s and an inter-trial interval of 5 s. For the Australian resident subjects, the test was

group-administered in the Language Laboratory of the University of Queensland, using a Tandberg IS9 master station controlling Tandberg 5600 listening stations.

Subjects responded with forced-choice identification of the stimulus by circling one of five alternatives on a response sheet, with the target words written in conventional English orthography. The Korean and Japanese learners of English were quite familiar with the stimulus words and English orthography. The older Korean (KO) group, who lacked familiarity with spoken English, were provided an initial training session to familiarize them with the test words and their written forms. All of the KO group were highly literate individuals and had no difficulty learning the task. They were tested individually, using a portable Sony cassette recorder and Sennheiser HD414SL headphones.

### 3.3. Results

The Australian control group performed close to 100% correct, as expected. Five subjects made no errors in 50 trials. One subject made a single error (an /ɪ/ heard as /e/), and two others made 2 errors each, on the same token (an /æ/) heard as an /e/ [repeated]). Table III summarizes the vowel identification rates with a

TABLE III. Vowel identification confusion matrices for the 5 groups of subjects; target (stimulus) vowels on the rows and subjects' responses on the columns

		Korean groups					Japanese groups				
		KI <i>n</i> = 5 Inexperienced response →					JI <i>n</i> = 5 Inexperienced				
		i	ɪ	e	æ	a	i	ɪ	e	æ	a
i	i	50					i	49	1		
t	ɪ	8	41	1			ɪ	49	1		
e	e	1		25	24		e	1	48	1	
m	æ			23	27		æ		4	46	
	a			1		49	a				50
		KE <i>n</i> = 5 Experienced					JE <i>n</i> = 5 Experienced				
		i	ɪ	e	æ	a	i	ɪ	e	æ	a
i	i	50					i	50			
ɪ	ɪ	14	36				ɪ	50			
e	e			45	5		e		50		
æ	æ			12	38		æ		1	49	
a	a					50	a				50
		KO <i>n</i> = 4* Old non-English									
		i	ɪ	e	æ	a					
i	i	49		1							
ɪ	ɪ	10	38	1							
e	e		9	40	1						
æ	æ			15	35						
a	a					50					

\* Raw frequencies for the KO group were adjusted to facilitate comparisons with other groups, where *n* = 5.

TABLE IV. Kruskal-Wallis rank sum tests for main effects and selected comparisons between groups. Overall identification score across the 5 vowels and the 5 groups is tested first, followed by separate tests for each vowel, with selected comparisons between groups nested under each significant main effect

Dependent variable	$\chi^2$	df	<i>p</i> level	Comparisons
Overall Identification Score	17.08	4	0.002	
Korean <i>vs.</i> Japanese	12.51	1	0.000	JI,JE > KE,KO*
Between Korean groups	3.31	1	0.06	KE,KO > KI
Identification Score [i:]		4	n.s.	
Identification Score [ɪ]	11.06	4	0.026	
Korean <i>vs.</i> Japanese	9.58	1	0.002	JE,JI > KE,KO,KI
Identification Score [e]	15.34	4	0.004	
Korean <i>vs.</i> Japanese	4.34	1	0.037	JE,JI > KE,KO
Between Korean groups	0.42	1	n.s.	KE,KO <i>vs.</i> KI
Identification Score [æ]	10.20	4	0.037	
Korean <i>vs.</i> Japanese	6.79	1	0.009	JE,JI > KE,KO
Between Korean groups	0.13	1	n.s.	KE,KO <i>vs.</i> KI
Identification Score [a:]		4	n.s.	

\*“JI,JE > KE,KO” indicates that the combined Japanese (JI + JE) group obtained a significantly greater score on the Identification index than a combined Korean group, comprising the KE and KO sub-groups.

confusion matrix for each group of non-native listeners, indicating the stimulus items (on the rows) and the response distributions (across the columns), summed over subjects in the group. Table IV presents a statistical analysis of the errors (the off-diagonal entries in Table III). For the statistical analysis of the subjects' responses, an Identification index (Id) was calculated:

$$\text{Id.} = \frac{\text{number of correct responses} - \text{number of incorrect responses}}{\text{number of responses}}$$

The Identification index yields scores that range over possible values of +1 to -1. The distribution of the Identification scores was found to be non-normal, resembling a “J” curve. Therefore, a non-parametric test, the Kruskal–Wallis rank sum test was used in place of a series of one-way ANOVA and selected comparisons between means. The results of the Kruskal–Wallis tests are shown in Table IV.

A significant main effect of listener group (KE, KI, KO, JI, JE) was found ( $\chi^2 = 17.08$ ,  $df = 4$ ,  $p = 0.0019$ ) on the Identification scores, calculated over all test items. Selected pair-wise comparisons were then made between sub-groupings of the subjects, as indicated in Table IV, again using the Kruskal–Wallis test. The Identification scores fell into three significantly different bands. The Japanese listeners (JE and JI groups) performed better than the English-experienced Koreans and the older Koreans (the KE and KO groups;  $\chi^2 = 12.51$ ,  $df = 1$ ,  $p = 0.001$ ), who in turn marginally out-performed the English-inexperienced Koreans (the KI group;  $\chi^2 = 3.31$ ,  $df = 1$ ,  $p = 0.06$ ).

Main effects for listener group were then tested on the Identification scores for each of the five vowels considered separately. The long vowels [i:] and [a:] were correctly identified in nearly all instances and made no contribution to the differences between listener groups. The Japanese performed more accurately than

the Koreans on the /ɪ/ vowel (JE, JI > KE, KO, KI;  $\chi^2 = 9.58$ ,  $df = 1$ ,  $p = 0.002$ ), never confusing it with its tense (long) counterpart [i:]. However, the major source of variability between groups lay in perception of the vowels [e] and [æ], where the pattern of performance was basically that described above for the Identification scores calculated over all items. The Japanese listeners were significantly better than the Koreans in discriminating [e] and [æ], and within the Korean group, there was a non-significant trend for the English-inexperienced group (KI) to perform worse than the English-experienced and the older Korean group (KO).

### 3.4. Discussion

Most errors of identification for Korean and Japanese listeners stemmed from confusion over the Australian English /e-æ/ contrast. The pattern of performance across the non-native listener groups indicated a complex interplay between the effects of L1 transfer and L2 experience.

Though in the expected direction, the effect of English experience was not statistically significant for the two groups of Japanese listeners (the JE *vs.* JI comparison). This may have been caused by ceiling effects, as error rates were low for both groups (with only 4 errors from 50 presentations of /e/, and 2 errors from 50 presentations of /æ/ in the JI group, and essentially error-less performance in the JE group). For the Korean KE and KI groups, the differences were statistically significant. However, the differences in performance cannot be securely attributed to differences in English experience alone. While both the KE and the KI groups belong to the post-war generation of speakers, for whom the /e-ɛ/ contrast has been neutralized, the KE group were, on average, 6 years older than the KI listeners. Thus, the possibility that age-related differences in the residual status of the L1 /e-ɛ/ contrast influenced the KE and the KI groups' performance cannot be discounted.

Also, it is notable that the KO listeners' possession of an L1 phonemic model for the foreign /e-æ/ contrast was not sufficient to confer an advantage over the Japanese listeners, who performed better despite the absence of any native target between /e/ and /a/. This result may not seem in accord with predictions of Best's and Flege's models (see above). However, it must be borne in mind that the older Koreans were not commensurate with the Japanese in terms of English experience. The comparisons that can be validly drawn between the younger Koreans and the Japanese, matched for English experience, yielded results in accord with the models' predictions. The Japanese performed better on the /e-æ/ contrast than did the Koreans, though the latter showed significant improvement with greater exposure to English (KE *vs.* KI comparison).

What was the basis of the superior performance of the Japanese listeners on the /e-æ/ contrast? Greater dispersion of phonetic targets in the L1 vowel space for "merged" Korean /e-ɛ/, which permitted assimilation of L2 [e] and [æ] tokens to a single native category, was one likely explanation. Alternatively (or additionally), it was possible that the Japanese listeners were making more effective use of durational cues relevant for the Australian English /e-æ/ contrast than the Koreans. To investigate these questions further, two distinct approaches were taken: an acoustic analysis of listeners' productions of the Australian English targets (Exp. 2), and an analysis of perceptual judgements made from the viewpoint of the

listener's native language (Exp. 3). The latter took the form of eliciting perceptual classifications using the native orthography (Japanese Kana and Korean Hangul) and of obtaining Prototypicality ratings of how closely the foreign vowels matched native targets.

#### 4. Experiment 2: Acoustic analysis of vowel productions

Although the relationship between speech perception and production is complex and controversial (e.g., Mattingly & Studdert-Kennedy, 1991; Pittam & Ingram, 1992), it may be expected that cross-language differences in the perception of a set of foreign vowels will be mirrored in L2 learners' productions of those same vowels, and listeners who discriminate better among L2 vowels would be expected to produce them more accurately. This assumption is supported by the recent finding that perceptual training enhanced the intelligibility of Japanese learners' productions of the English /r-l/ contrast (Akahane-Yamada, Tohkura, Bradlow, & Pisoni, 1996).

Assuming that vowel perception and production are linked through a common set of articulatory/auditory parameters, then measurement of non-native productions may provide some indication of which features are used in recognition. Specifically, it was hoped that acoustic analysis of their productions would shed light on the source of the Japanese superiority in /e-æ/ discrimination, whether it could be attributed to more effective use of vowel quality or timing features, or some combination of the two.

First and second formant frequency and vowel duration measurements were made to assess vowel quality and quantity (length), as previously used to characterize native-speaker differences in the test stimuli (Section 3.2.2.). After appropriate normalization for speaker differences, these acoustic measurements were then used to test how well the non-native production tokens could be classified into their intended Australian English phoneme categories using a Gaussian statistical classifier.

##### 4.1. Methods and procedures

The same Japanese and Korean speakers as in Experiment 1 were asked to read aloud the /h\_d/ words that they had previously attempted to identify. Items were presented in a word list using standard English spelling. Each word was read five times in succession and the middle three tokens were analyzed acoustically. Production data were obtained only from the Korean and Japanese groups resident in Australia, and not from the KO group who were not users or learners of English.

Subjects' productions were digitized at a 10 kHz sampling rate. The duration of each vowel was measured by hand from dual spectrogram and waveform displays; mid-vowel  $F_1$ ,  $F_2$ , and  $F_3$  formant frequency measurements were taken from the spectrogram, supplemented by an LPC smoothed spectral display, as described previously.

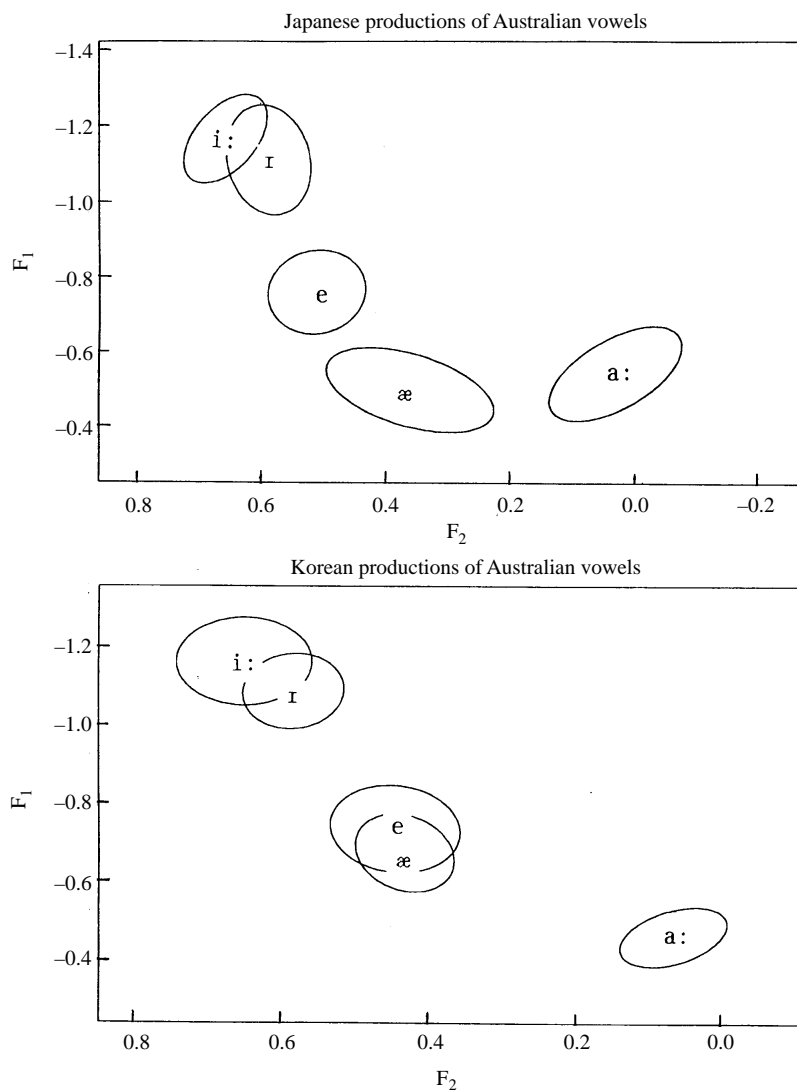
Because the experiment groups contained male and female speakers, formant measurements were speaker-normalized by Nearey's log-normalization method (Nearey, 1978). This method was chosen as it has been shown to perform well in a

cross-dialectal context (Hindle, 1978). Similarly, the vowel duration measurements were normalized to minimize speaker-dependent variation in speech rate using the Lobanov (1971) method, which is more appropriate for duration normalization.

#### 4.2. Results

##### 4.2.1. Vowel quality

Mid-vowel  $F_1$  and  $F_2$  plots of the Japanese and Korean productions of Australian English front vowels are shown in Fig. 3, with phonetic symbols indicating the



**Figure 3.**  $F_1/F_2$  plots (centroids and 1 standard deviation ellipses) of Japanese and Korean productions of Australian English front vowels. Normalized by Nearey method.

TABLE V. Confusion matrices for Japanese and Korean vowel productions, generated by a Gaussian classifier, operating on (Nearey) normalized, center vowel, first and second formant frequency measurements ( $F_1$  and  $F_2$ )

	Korean (KI + KE groups)					Japanese (JI + JE groups)				
	i:	ɪ	e	æ	a:	i:	ɪ	e	æ	a:
i:	24	6	0	0	0	i:	18	12	0	0
ɪ	11	19	0	0	0	ɪ	8	19	3	0
e	0	3	23	4	0	e	0	1	25	4
æ	0	0	19	11	0	æ	0	0	3	24
a:	0	0	0	0	30	a:	0	0	0	2

centroids of the target vowel categories and the ellipses set at one standard deviation. There was clear separation of [e] and [æ] tokens into distinct regions of the vowel space in the case of the Japanese productions, but a considerable degree of overlap for the Korean productions. It appeared that the Japanese were producing acoustically distinct vowel quality targets for the /e-æ/ contrast, whereas the Koreans were not.

The discriminatory power of the formant measurements for retrieving the target vowel categories was tested using a Gaussian classifier, operating a closed test, on the basis of Mahalanobis distances (a facility of the  $Mu^+$  speech database system; Harrington, Cassidy, Fletcher & McVeigh, 1993). The confusion matrices generated for the Japanese and Korean tokens by the classifier are shown in Table V.

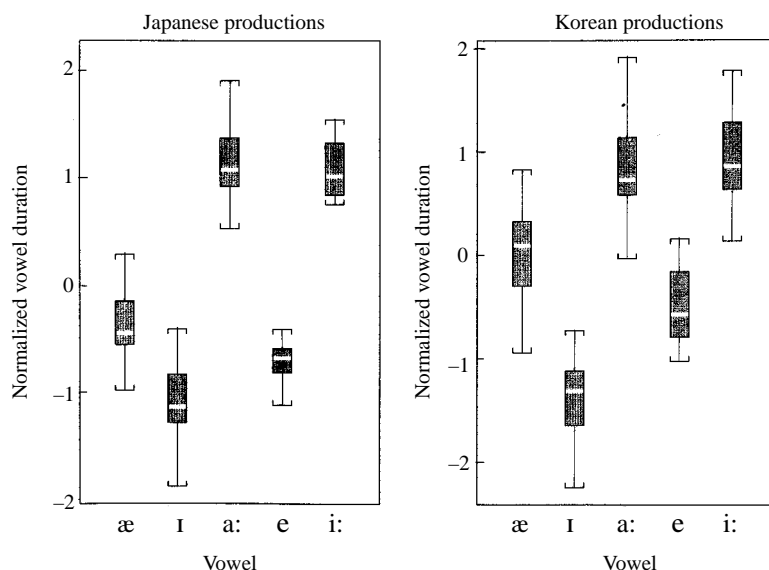
Comparison of the confusion matrices confirmed that the /e/ and /æ/ tokens produced by the Koreans were less acoustically discriminable on the basis of normalized  $F_1$ - $F_2$  values than were those of the Japanese subjects. Approximately 63% of the Korean /æ/ productions were misclassified as /e/, whereas 80% of the Japanese /æ/ productions were correctly classified, with errors equally distributed between /e/ and /a:/.

These group differences were also examined at the level of individual subjects. Appendix A shows formant plots for each of the twenty speakers. For the Japanese speakers, the target formant values for /e/ and /æ/ were acoustically separated in all cases. However, for three of the English-experienced Japanese subjects, /æ/ clustered with /a/. Three of the English experienced Koreans (KE group) produced non-overlapping formant targets for /e/ and /æ/ (subjects no. 3, 4, and 5), but the rest produced tokens which appeared to cluster about a *single* vowel target.

Thus, in terms of vowel quality, the production data reinforced the perceptual findings reported previously (Tables III and IV) in suggesting that the Japanese speakers responded more effectively to vowel quality cues for the /e-æ/ contrast than did the Korean speakers.

#### 4.2.2. Duration

In their production of Australian vowels, the Japanese speakers showed a tendency to categorically distinguish long and short vowels, whereas the Korean speakers' productions were more graded in their temporal characteristics, monotonically reflecting the phonetic gradation of vowel duration that was present in the



**Figure 4.** Box plots of durations of Japanese and Korean speakers' productions of Australian vowels; Lobanov normalized.

Australian targets. This can be seen graphically in Fig. 4, where the speaker-normalized distributions of the Korean and Japanese speakers' vowel durations have been plotted. For the Japanese productions, a straight-line partition of the tense (long) and lax (short) vowels can be made, separating the distributions of individual vowel durations. The whiskers on the box plots extend 1.5 times the inter-quartile range above and below the median, encompassing approximately 99% of observations for a normal distribution.

A series of *t*-tests was conducted on the five pairs of vowels that differed minimally from one another in duration. The results are shown in Table VI, which also reports the mean (un-normalized) duration ratio ( $V1/V2$ ) for each pair of vowels.

For the Japanese speakers' productions, the duration ratios and *t*-test results indicate that the (L2) cross-category contrast between (tense/long) /a:/ and (lax/short) /æ/ was more prominent than the other three (within category) duration contrasts. For the Korean speakers' vowels, duration contrasts within the lax vowels were not appreciably less prominent than the /a:-æ/ cross-category

TABLE VI. Duration ratios and *t*-tests for selected pairs of vowels produced by Japanese & Korean speakers

Vowel pair V1-V2	Japanese productions			Korean productions		
	Duration V1/V2	<i>t</i> -test value	<i>p</i>	Duration V1/V2	<i>t</i> -test value	<i>p</i>
[i:-a:]	0.92	-0.79	0.430	1.02	0.54	0.590
[a:-æ]	1.63	-7.77	0.000	1.27	-5.51	0.000
[æ-e]	1.14	-2.05	0.044	1.20	-3.08	0.003
[e-ɪ]	1.20	-3.18	0.002	1.49	-5.50	0.000

contrast. However, the statistical trends are not clear-cut and the strongest claim that may be made here is that Japanese productions of Australian vowels showed a *tendency* to categorically distinguish tense and lax L2 phonological targets.

In summary, the Japanese speakers' superiority in perception of the /e-æ/ contrast (Section 3.3.) was also reflected in their production of F<sub>1</sub>-F<sub>2</sub> target values for these vowels, which were found to cluster into distinct regions of the formant space, compared with the Koreans, whose targets for /e/ and /æ/ were not acoustically separable by a Gaussian classifier. Thus, more accurate perception and production of the vowel quality contrast appeared to underlie the Japanese performance, rather than simply more effective use of temporal cues, as was initially suspected. The significance of this finding for models of L2 vowel perception is discussed in Section 6 below.

However, there were also differences in the way Japanese and Korean speakers reproduced the temporal characteristics of the L2 vowels, which may be relevant for their perceptual accommodation to Australian English. The temporal characteristics of the Japanese vowel productions did not provide better statistical separation among the 5 target vowels than did the Korean productions. Rather, they suggested that the Japanese speakers were implementing a categorical distinction that they probably also employed in perception. Direct evidence for this was obtained from a second perception experiment in which listeners were induced to perceive the target vowels from the viewpoint of their native language.

### 5. Experiment 3: Prototypicality ratings

In the third experiment, perceptual judgements of the same five vowels were obtained from the standpoint of the listeners' native language. The aim was to assess the impact of prior phonological learning on vowel perception, by examining how the foreign vowels were classified in terms of native phonological categories and through perceptual ratings of their relative "goodness" as exemplars of native vowel types.

#### 5.1. Method

Preceding the foreign vowel identification task, which provided the data for Experiment 1, a short native-language identification and prototype rating test was conducted, in which Japanese and Korean listeners were asked to transcribe the stimuli in their native orthography (Kana and Hangul, respectively) and to rate how closely each stimulus vowel matched the native vowel chosen in their transcription.

The stimuli were the same as those presented for identification as English vowels in Experiment 1. Tokens from the two Australian speakers were randomly distributed to trials. Each token was presented over headphones and rated a total of 5 times in the course of the experiment. On any given trial, listeners were allowed 7 s to label the stimulus, by transcribing it in the native orthography, and rate its proximity to the native target on a similarity scale ranging from "0 = identical", to "10 = very different". Listening conditions and procedures were otherwise the same as those of Experiment 1.

Korean orthography provides distinct symbols for 4 of the 5 vowel quality targets represented in the experiment, including the "merged" Korean /e/ and /ɛ/.

Korean speakers were instructed to use a hyphen to indicate “long vowels”, following a special convention familiar to Korean readers but not part of Hangeul orthography. In the case of Japanese, only 3 orthographic symbols were available for transcription of the front and low-central vowels, but moraic length is represented orthographically in Kana.

## 5.2. Results

Table VII shows how listeners identified the Australian English vowels in terms of native phonological categories and shows the mean prototype ratings obtained for each vowel. Raw frequencies are shown for the identifications, from a total of 50 responses per item for the Japanese (JI + JE) and Korean (KI + KE) learners of English and 20 per item for the older Korean (KO) group. Response patterns within language groups, between the more and less experienced learners of English, did not differ significantly so the responses from the JI and JE groups and the KI and KE groups were combined.

Responses to items from Speakers A and B were tabulated separately and revealed striking differences between the two language groups in how they responded to speaker-related variation in duration of the stimulus items. The Japanese listeners consistently classified vowels as long or short, without regard to the substantial speaker-related differences in duration between tokens of the same phonological type. The Korean listeners, on the other hand, were strongly influenced by the absolute duration of individual tokens and consequently produced classifications that were inconsistent across Speakers A and B. For example, the Japanese were consistent in their classification of phonologically long and short high front vowels (/i:/ and /ɪ/) across speakers. All tokens of Australian English /i:/ were appropriately identified with the long or geminate Japanese /i:/, regardless of the speaker. Similarly, the majority of short /ɪ/ tokens were consistently classified as “short”, regardless of the speaker (45/50 for Speaker A; 46/50 for Speaker B). However, the Korean listeners produced quite different native-language classifications of these vowels, depending on the speaker. All of Speaker B’s /ɪ/ tokens were classified as “short” by the combined KI + KE group, but only 70% of Speaker A’s /ɪ/ tokens were so classified. Inspection of Fig. 2 reveals that the Koreans’ classifications correlate strongly with the pattern of speaker differences in the absolute duration of the stimulus tokens. In all cases where there were discrepant duration classifications across speakers by Korean listeners, the direction of the discrepancy accorded with differences in absolute duration between tokens for Speaker A and Speaker B. The Japanese, on the other hand, were evidently normalizing the stimuli for speaker-related duration differences in their native-categorization of the tokens as “long” or “short”.

Interestingly, the KO group also responded to the absolute duration of tokens and failed to show evidence of normalizing them for speaker-dependent rate variation, as might have been expected if the phonological contrast of vowel length were an active perceptual target for older, but not younger Korean speakers, as suggested earlier (Section 2.1.). There were no speaker normalization effects discernible in the native-language categorizations of vowel quality, and none were expected given the lack of inter-speaker variation in  $F_1$ - $F_2$  targets for the stimuli (Fig. 2). However, expected differences between older and younger Koreans in native categorization of

TABLE VII. Native language identifications and prototype ratings. Vowel categorization based upon native orthography (Japanese: Kana, Korean: Hangul). Prototype ratings assigned on a 10-point scale of proximity to native vowel targets. Ratings for tokens from the two Australian speakers (A and B) reported separately. (See text for details.)

		Japanese listeners (JI + JE groups)									
		Responses: native language referents									
Target English	Speaker	i:	i	e:	e	a:	a	Other	Proto-type ratings		
i:	A	50							4.82		
	B*	50							4.64		
ɪ	A	5	45						5.64		
	B*	4	46						5.06		
e	A				50				4.00		
	B*				50				4.18		
æ	A			2	15	4	29		7.34		
	B*				14	1	35		7.26		
a:	A			1		49			4.62		
	B*					50			5.32		
		Korean listeners (KI + KE groups)									
		Responses: native language referents									
Target English	Speaker	i:	i	e:	e	ɛ:	ɛ	a:	a	Other	Proto-type Ratings
i:	A	47	3								3.98
	B*	38	12								6.60
ɪ	A	14	35		1						5.00
	B*	50									5.02
e	A		1	7	30	2	10				3.76
	B*			35		15					4.34
æ	A			6	6	22	16				4.56
	B*			2	13	1	34				3.90
a:	A							47	2	1	4.28
	B*							45	4	1	3.88
		Korean monolinguals (KO group)									
		Responses: native language referents									
Target English	Speaker	i:	i	e:	e	ɛ:	ɛ	a:	a	Other	Proto-type Ratings
i:	A	20									2.20
	B*	20									1.95
ɪ	A	13	7								2.85
	B*	1	17		2						1.65
e	A			4	12	1	3				3.25
	B*		1		17		2				1.75
æ	A			1		18	1				2.40
	B*			1		5	14				1.55
a:	A							20			1.35
	B*							20			0.70

vowel quality differences for /e/ and /æ/ were observed. While there was no difference between younger and older Koreans in their rate of “correct” assimilation of Australian English /e/ to Korean /e/ (KE + KI = 72/100, KO = 33/40,  $\chi^2 = 0.008$ ,  $df = 1$ ,  $p = 0.926$ ), the older Koreans were more likely to assimilate the vowel quality of Australian English /æ/ to Korean /e/ and thus make a vowel quality distinction that was evident from their responses in Experiment 1 (KE + KI = 71/100, KO = 38/40,  $\chi^2 = 4.184$ ,  $df = 1$ ,  $p = 0.041$ ).

We speculated from the results of Experiment 1 that the Japanese listeners’ superior perception of the /e-æ/ contrast was due to their more effective use of duration cues. However, Experiment 2 provided indirect evidence from their productions of the Australian English targets that they also perceived the /e-æ/ quality difference more clearly than did younger Koreans (production data from the KO group was not obtained). The Japanese vowel system does not provide separate vowel quality categories for /e/-/æ/-/a/, so Japanese listeners were obliged to make use of duration categories provided by the native phonology for separately categorizing these sounds. The favored strategy (Table VII) was to assimilate Australian English /æ/ to short Japanese /a/, with a less favored option being to assimilate it to short Japanese /e/. In either case, Japanese listeners ignored speaker-dependent temporal variation in making their categorization.

Prototypicality ratings differed significantly between listener groups. The Australian English vowels were perceived as closer to native L1 targets by the Korean monolingual listeners than by either of the two English learner groups (one-way ANOVA  $F(2, 23) = 34.44$ ,  $p < 0.001$ ). It seems likely that, as monolinguals with limited exposure to English, they had not developed distinct perceptual referents for the foreign targets and consequently located them entirely within the perceptual framework of their native vowel space.

The Koreans’ Prototypicality ratings did not differ significantly across vowel targets. The Japanese, on the other hand, gave significantly higher (less prototypical) ratings to Australian English /æ/ than to any other target.

## 6. General discussion

In this section we summarize the major findings from the three experiments and draw implications for recent models of cross-language vowel perception in second language learning.

Firstly, Experiment 1 provided further evidence of the importance of native-category transfer effects in the perception of non-native vowels. Comparisons between younger and older Korean listeners’ perceptions of the /e-æ/ contrast in Australian English showed how an on-going phonemic merger in L1, resulting in cross-generation differences within a speech community, can differentially impact upon listeners’ perception of a similar vowel contrast in L2. While this is a novel finding, it might have been predicted from any of the current models of cross-language vowel perception (e.g., Flege, 1995; Best, 1995) that trace their lineage to the contrastive analysis hypothesis.

Some indication of the relative strength of transfer effects, through prior possession of an appropriate L1 phonetic model for the L2 vowel contrast, versus experience in L2, can be gained from a three-way comparison of the KE, KI, and KO groups’ performance on the /e-æ/ contrast. The older Koreans, with no

intensive exposure to English (KO group), performed better than young Koreans with 1 year or less intensive exposure to Australian English (KI group), and approximately as well as Koreans with more than 5 years' residence in Australia and a high level of English fluency (KE group). Caution is required in these comparisons because the factors of L2 experience and Age (the main indicator of possession of the pre-merger /e-æ/ contrast) are confounded in our subjects.

Secondly, the results of the three experiments, taken collectively, provide new insight into the operation of perceptual mechanisms underlying transfer effects. Again, the mechanism has been anticipated by theoretical accounts (see, for example, that of Best, 1995) but has not, to the authors' knowledge, been previously demonstrated empirically. Phonological transfer effects in L2 vowel perception invoke normalization strategies (or "attunement mechanisms") required for extracting particular phonological features from non-phonological sources of phonetic variation in the speech signal. The operation of this mechanism was revealed by comparisons between the responses of Japanese and Korean subjects to tense (long) and lax (short) vowels in the face of speaker-dependent temporal variations in vowel production.

In Experiment 2, acoustic measurements of vowel duration from L2 productions indicated a tendency towards categorical representation of vowel length contrasts by the Japanese subjects, consistent with the central role of quantity contrasts in Japanese phonology. The Korean subjects, in whose native language quantity contrasts play a marginal phonological role and appear to be no longer productive, showed a phonetically graded response to temporal contrasts among the L2 vowels.

But clear evidence that Japanese listeners normalized the stimuli for speaker-dependent vowel duration, whereas the Koreans failed to do so, came from Experiment 3, where listeners classified the foreign vowels into native language categories. The Japanese listeners were consistent across speakers in their native-categorizations of the vowel tokens, whereas the Koreans varied in ways predictable from speaker-dependent differences in the duration of stimulus tokens.<sup>4</sup>

The greater accuracy of the Japanese speakers' perception and production of the Australian English front vowels and the /e-æ/ contrast in particular, compared with that of Koreans with matched English experience, appears to involve processing of both vowel quality and quantity features. On the one hand, Experiment 2 showed that the Japanese produced /e/ and /æ/ tokens with acoustically separable vowel qualities, whereas the Koreans' targets overlapped in  $F_1$ - $F_2$  space. We take this as indirect evidence that the Japanese perceived the /e-æ/ quality distinction more clearly. On the other hand, evidence from Experiment 3 showed that Japanese listeners consistently managed to assimilate the difficult-to-classify Australian English /æ/ to a short vowel native category (predominantly, Japanese /a/) in the face of substantial speaker variation. This ability to separate phonological from non-phonological sources of variation on a given phonetic dimension goes to the heart of the speech perception task.

The significance of these findings on the comparative performance of Japanese and Korean learners of Australian English for current theories of cross-language vowel

<sup>4</sup> A reviewer suggested that the number of native categories for classifying the English vowels (6: 3 qualities  $\times$  2 lengths for Japanese; 8: 4  $\times$  2 for Korean) may have adversely affected the consistency of the latter's classifications. Perhaps so, but this would not explain why the Korean listeners' responses consistently patterned with the temporal differences between the two speakers' vowels.

perception (and speech perception more generally) appears to lie in reinforcing the role of language-specific phonological learning in the operation of processes that hitherto have been attributed to universal phonetic capabilities (e.g., the Motor theory of speech perception and its close relatives; Liberman & Mattingly, 1985). While we cannot offer specific suggestions as to underlying mechanisms, the evidence from this study indicates that basic processes, such as speaker normalization, and the integration of temporal and spectral properties of speech into unitary percepts that underlie phonetic similarity judgements, are contingent upon phonological features and training in L1. Our results complement recent findings on infant speech perception capabilities, involving magnet effects in vowel perception (e.g., Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992), language-specific prosodic parsing strategies as aids to word recognition (e.g., Cutler, 1994; Jusczyk, Cutler & Redanz, 1993), and other kinds of preferential sensitivity to native language characteristics, well summarized by Vihman (1996).

While competitors to the Motor theory of speech perception, such as the Direct Realist theory of Best (1995), accommodate language-specific learning effects in the development of speaker-normalization capabilities, and in the integration of cross-modal or cross-dimensional sources of sensory information in speech perception, an explicit account of how this integration is achieved remains beyond the reach of current models. The findings and methods of this study are deficient for the task of developing such a model, but are suggestive of the way ahead.

For example, it was clear from the native-language categorizations that the Australian /æ/-/e/ tokens were perceptually differentiated by the Japanese listeners, but not by the younger Korean listeners. Clearly, the relevant parameter here is the breadth of clustering of acceptable phonetic tokens about a phonemic type. In a situation of phonological merger, such as the merger of /e/ and /ɛ/ in contemporary Seoul Korean, the boundaries of the emergent /e~ɛ/ phoneme would need to be quite broad to accommodate tokens from conservative speakers who will keep the /e/ and /ɛ/ targets “deliberately” distinct, at least in formal speech. Discrimination testing among tokens ranging across the [e-ɛ-æ] vowel quality space is needed to confirm that the differences in L1 categorization of the stimuli by the Japanese and Korean subjects were, in fact, responsible for their differential performance on the same tokens for identifications as L2 targets. Although it seems most likely that the L2 perception differences were expressions of L1 perceptual learning effects, it remains to be ascertained, through systematic discrimination testing, precisely how such learning effects operate to warp perceptual vowel spaces of the two listener groups. Kuhl’s model of Perceptual Magnets is suggestive in this regard. However, the warping effects of tight versus diffusely clustered phoneme targets remain to be investigated.

Similarly, although the present study succeeds in establishing an empirical link between phonological transfer effects and speaker-normalization processes required for extracting phonological targets from speech, a model of cross-language vowel perception should specify more precisely the phonetic conditions under which phonological transfer will take place (or fail to occur) and the consequences that ensue for speech processing. The differential responses of the Japanese and the Korean listeners to the temporal characteristics of the stimuli in this study provide suggestions on this point for further research.

In particular, the prototype ratings of Experiment 3 showed that the Japanese listeners were perceptually aware of the non-native character of the Australian English /æ/. It is interesting that although /æ/ was clearly perceived as “non-native” by Japanese listeners, it was also distinguished from its near neighbor /a:/, in part at least by the application of an L1-based speaker normalization strategy for the identification of phonological length in Japanese. Thus, it appears that an impressionistically “foreign” vowel may nevertheless access native-language perceptual processing strategies in the course of its identification, provided that the vowel in question is assimilable to a native target on at least one phonological feature.

In general, our findings suggest that L2 learners confronted with identification of foreign vowels seek to apply or modify perceptual strategies for phonological feature identification that were acquired in the course of first language learning. Where L1 processing strategies fail, perhaps because error tolerances for matching the input stimulus against some internal representation are not met, listeners respond in a graded fashion to phonetic properties of the stimulus, unmediated by phonological expectations. One of the challenges for models of foreign vowel perception is to yield more explicit predictions as to the phonetic conditions under which L1 phonological processing strategies will fail to apply, given a particular set of input speech stimuli.

The authors wish to thank Dr Patrice Speeter Beddor and two anonymous reviewers for their many helpful and incisive comments on earlier drafts of this paper. The authors are also deeply indebted to Catherine Best for her thorough and very insightful review of the thesis (Park, S.-G., 1997) research from which this report derives.

### References

- Akahane-Yamada, R., Tohkura, Y., Bradlow, A. R. & Pisoni, D. B. (1996) Does training in speech modify speech perception? In *Proceedings of the fourth International Congress on Spoken Language Processing (ICSLP-96)*, University of Delaware and A. I. du Pont Institute
- Best, C. T. (1995) A direct realist view of cross-language speech perception. In *Speech perception and linguistic experience: issues in cross-language research*, (W. Strange, editor), pp. 171–203. Baltimore: York Press
- Cutler, A. (1994) Segmentation problems, rhythmic solutions, *Lingua*, **92**, 81–104
- Durie, M. & Hajek, J. (1994) A revised standard phonemic orthography for Australian English vowels, *Australian Journal of Linguistics*, **14**, 93–107
- Flege, J. E. (1987) The production of “new” and “similar” phones in a foreign language: evidence for the effect of equivalence classification, *Journal of Phonetics*, **15**, 47–65
- Flege, J. E. (1992) Talker and listener effects on degree of perceived foreign accent, *Journal of the Acoustical Society of America*, **91**(1), 370–389
- Flege, J. E. (1995) Second language speech learning theory, findings, and problems. In *Speech perception and linguistic experience: issues in cross-language research*, (W. Strange, editor) pp. 233–277. Baltimore: York Press
- Flege, J. E. & Munro, M. J. (1994) Auditory and categorical effects on cross-language vowel perception, *Journal of the Acoustical Society of America*, **95**(6), 3623–3641
- Fletcher, J. E. & McVeigh, A. (1992) Towards a model of segment and syllable duration in Australian English. *Proceedings of the fourth Australian International Conference on Speech Science and Technology* (J. Pittam, editor), pp. 28–33. Brisbane: A.S.S.T.A.
- Harrington, J., Cassidy, S., Fletcher, J. & McVeigh, A. (1993) The Mu<sup>+</sup> system for corpus based speech research, *Computer Speech and Language*, **7**, 305–331
- Hindle, D. (1978) Approaches to vowel normalisation in the study of natural speech. In *Linguistic variation: models and methods*, (D. Sankoff, editor), pp. 161–170. New York: Academic Press
- Hong, Y.-S. (1991) A sociolinguistic study of Seoul Korean. *Korean Unification Studies, Series 12*, Seoul: Research Centre for Peace & Unification of Korea

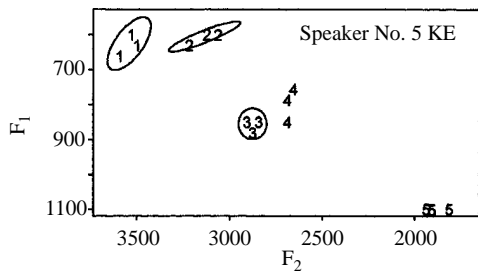
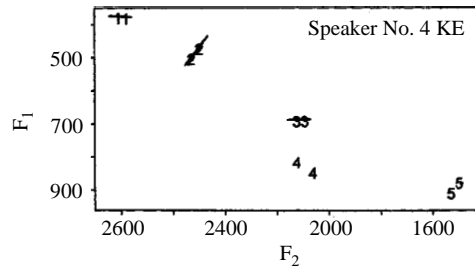
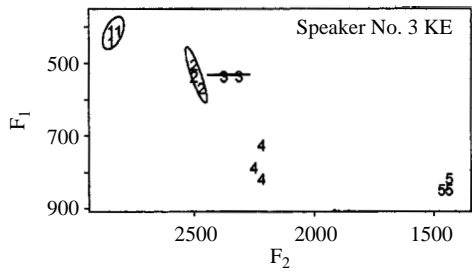
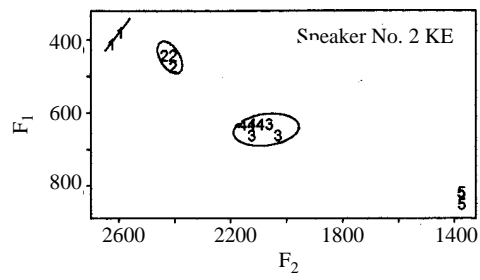
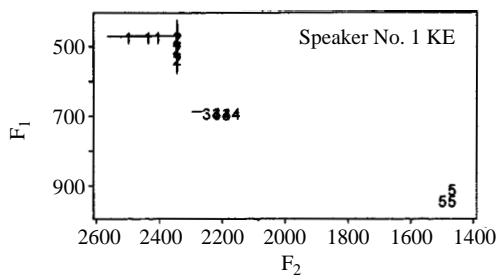
- Ingram, J. C. (1995) One step forward and two backwards: reflections on Durie and Hajek's proposed revisions to the phonemic transcription of Australian English vowels, *Australian Journal of Linguistics*, **15**, 215–239
- Ingram, J. C. & Pittam, J. (1987) Auditory and acoustic correlates of perceived accent change: Vietnamese schoolchildren acquiring Australian English, *Journal of Phonetics*, **15**, 127–143
- Jusczyk, P. W., Cutter, A. & Redanz, N. J. (1993) Infants' preference for the predominant stress patterns of English words. *Child Development*, **64**, 675–687
- Keating, P. A. & Huffman, M. K. (1984) Vowel variation in Japanese, *Phonetica*, **41**, 191–207
- Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N. & Lindblom, B. (1992) Linguistic experience alters phonetic perception in infants by 6 months of age, *Science*, **255**, 606–608
- Lado, R. (1957) *Linguistics across cultures*. Ann Arbor, MI: University of Michigan Press
- Lieberman, A. & Mattingly, I. (1985) The Motor theory of speech perception revised, *Cognition*, **21**, 1–36
- Lively, S. E., Logan, J. S. & Pisoni, D. B. (1993) Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories, *Journal of the Acoustical Society of America*, **94**(3), 1242–1255
- Lobanov, B. (1971) Classification of Russian vowels spoken by different speakers, *Journal of the Acoustical Society of America*, **49**, 606–608
- Magen, H. S. & Blumstein, S. E. (1993) Effects of speaking rate on the vowel length distinction in Korean, *Journal of Phonetics*, **21**, 387–409
- Magnuson, J. S. & Akahane-Yamada, R. (1996) Acoustic correlates to the effects of talker variability on the perception of English /r/ and /l/ by Japanese listeners. In *Proceedings of the fourth International Congress on Spoken Language Processing (ICSLP-96)*, University of Delaware and A. I. du Pont Institute
- Mattingly, I. & Studdert-Kennedy, M. (1991) *Modularity and the motor theory of speech perception*. Erlbaum
- Nearey, T. (1978) Phonetic feature systems for vowels. PhD dissertation, University of Connecticut. Reproduced by Indiana University Linguistics Club. Bloomington: I.U.L.C.
- Park, S.-G. (1997) Australian English pronunciation acquisition by Korean and Japanese learners of English. Unpublished Ph.D. dissertation, Department of English, University of Queensland
- Pittam, J. & Ingram, J. C. (1992) Accuracy of perception and production of compound and phrasal stress by Vietnamese Australians, *Applied Psycholinguistics*, **13**, 1–12
- Strange, W., Bohn, O.-S., Trent, S. A., McNair, M. C. & Bielec, K. C. (1996) Context and speaker effects in the perceptual assimilation of German vowels by American listeners. In *Proceedings of the fourth International Congress on Spoken Language Processing (ICSLP-96)*, University of Delaware and A. I. du Pont Institute
- Syrdal, A. K. & Gopal, H. S. (1986) A perceptual model of vowel recognition based on the auditory representation of American English vowels, *J. Acoust. Soc. Am.*, **79**, 1086–1100
- Vihman, M. V. (1996) *Phonological development: the origins of language in the young child*. Cambridge, Mass.:Blackwell

**Appendix**

Format frequency plots for Korean and Japanese productions of Australian English vowels.

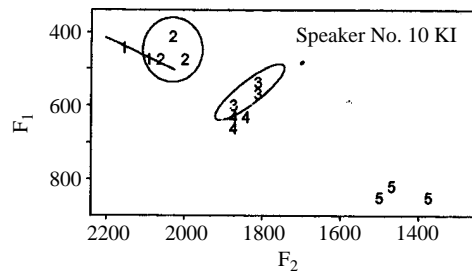
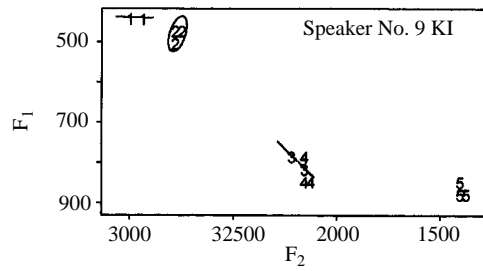
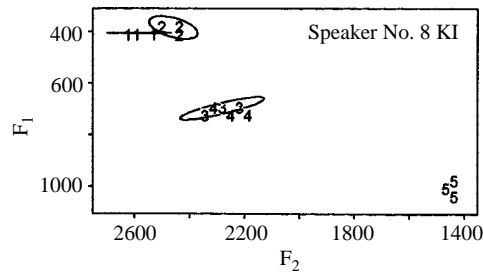
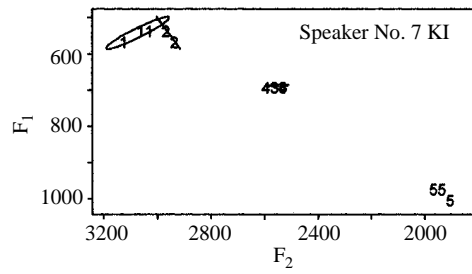
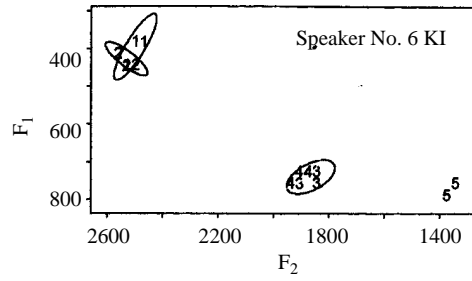
Legend: 1 = [i:], 2 = [ɪ], 3 = [e], 4 = [æ], 5 = [a:]

*Korean: More English Experience*

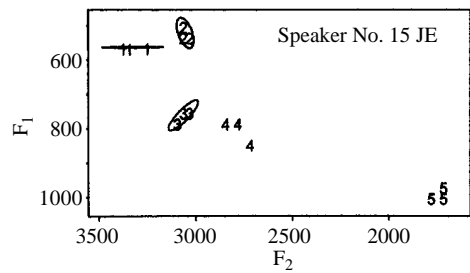
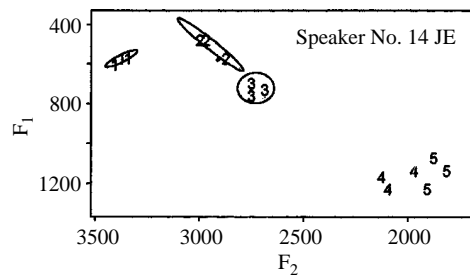
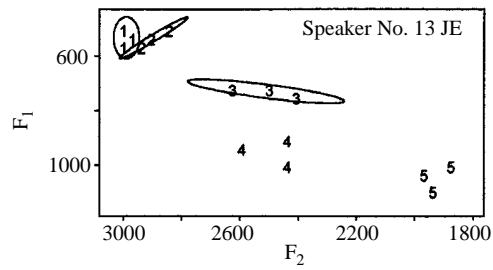
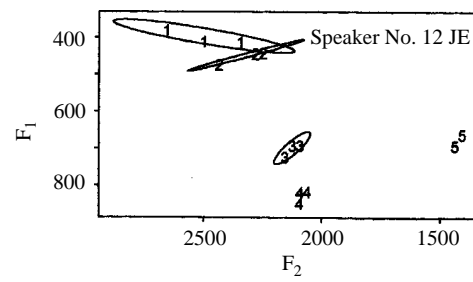
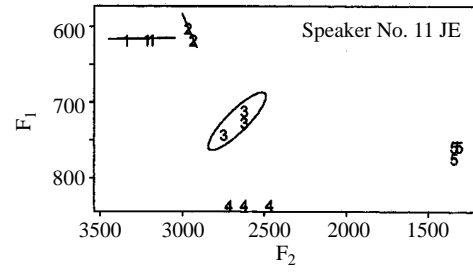


Appendix, cont.

Korean: Less English Experience



Japanese: More English Experience



Appendix, cont.

Japanese: Less English Experience

