

# Vocabulary size matters: The assimilation of second-language Australian English vowels to first-language Japanese vowel categories

RIKKE L. BUNDGAARD-NIELSEN  
*University of Western Sydney*

CATHERINE T. BEST  
*University of Western Sydney and Haskins Laboratories, New Haven, CT*

MICHAEL D. TYLER  
*University of Western Sydney*

Received: August 7, 2008      Accepted for publication: August 2, 2009

## ADDRESS FOR CORRESPONDENCE

Rikke L. Bundgaard-Nielsen, MARCS Auditory Laboratories, University of Western Sydney, Locked Bag 1797, Penrith South DC, NSW 1797, Australia. E-mail: r.bundgaardnielsen@uws.edu.au

## ABSTRACT

Adult second-language (L2) learners' perception of L2 phonetic segments is influenced by first-language phonological and phonetic properties. It was recently proposed that L2 vocabulary size in adult learners is related to changes in L2 perception (perceptual assimilation model), analogous to the emergence of first-language phonological function (i.e., attunement to the phonological identity of words) associated with the "vocabulary explosion" at 18 months. In a preliminary investigation of the relationship between L2 perception and vocabulary size, Japanese learners of Australian English identified Australian English vowels, provided goodness of fit ratings, and completed a vocabulary size questionnaire. We adopted a "whole-system" approach, allowing learners to apply all native vowel system possibilities to the full L2 vowel system. Learners with a larger L2 vocabulary were more consistent in their vowel assimilation patterns, compatible with the L2 perceptual assimilation model.

Second-language (L2) learners have foreign accents that are influenced by their particular native (first) language (L1; Flege, 2002; Flege, Munro, & MacKay, 1995; Munro, Flege, & MacKay, 1996). It is less obvious, but equally important, that learners of an L2 also have an "accent" in their perception of the new language (Jenkins, Strange, & Polka, 1995), which is systematically related to the perceived

similarities between the phonological segments of their L1 and L2 (Best & Strange, 1992; Flege, 1987). Although learners may find some nonnative contrasts easy to discriminate, it is common for two or more L2 phones to be perceived as identical or similar to just one native phoneme (Bohn, 1995; Bohn & Flege, 1992; Flege, Bohn, & Jang, 1997; Flege & MacKay, 2004; Goto, 1971; Guion, Flege, Akahane-Yamada, & Pruitt, 2000; Werker & Tees, 1984). However, difficulties in perceiving nonnative phones do not always persist as L2 proficiency increases (Flege et al., 1995; Ingram & Park, 1997; Tsukada et al., 2005), although it is unclear what drives this change and results in an increased nativelike perception or production. The purpose of this paper is to investigate the role of vocabulary size in how L2 learners learn to perceive these nonnative phones.

In particular, the paper investigates the role of vocabulary size on L2 *vowel* perception, taking into consideration the entire L1 and L2 vowel systems, rather than a subset of nonnative and native consonants or vowels. We focus on a whole vowel system for a number of reasons. As discussed in detail below, there is ample evidence that a learner's L1 vowel inventory (size and organization) influences how L2 vowels are perceived, and vowels are less discretely perceived (and articulated) than consonants (see, e.g., Strange, 1998a, 1998b), likely resulting in vowels being inherently more interconnected as a system. If this is the case, then failure to include the entire vowel system in a perceptual experiment could result in an ecologically invalid estimation of the perceptual flexibility L2 learners.

There is abundant evidence that the size and organization of the L1 vowel inventory influences how L2 learners perceive the vowel contrasts in their new language. For example, native speakers of Spanish, a language with no temporal or tense-lax spectral contrasts, struggle to discriminate between British English /i/ and /ɪ/, because both are perceived as instances of Spanish /i/ and their native phonology is not attuned to durational differences as being significant for vowel identification (Escudero & Boersma, 2004). In contrast, native speakers of Serbian, a language that also lacks this vowel contrast but does have a long and short version of /i/, discriminate between these vowels very well on the basis of duration because they are attuned to durational differences as a marker of phonological identity (Krebs-Lazendic & Best, 2008). Moreover, native speakers of German, a language with an /i/-/ɪ/ contrast, discriminate /i/ and /ɪ/ quite well on the basis of spectral differences, although the German realization of these two vowels is slightly different than the English pronunciations (Bohn & Flege, 1992; Iverson & Evans, 2007).

In addition, the number of vowels in learners' L1 influences their L2 vowel perception. The perceptual difficulty experienced by an L2 learner is partly determined by the size of the L1 vowel inventory relative to the L2 vowel inventory. Thus, it is harder for speakers of L1s with smaller vowel inventories (such as Spanish) to acquire a rich L2 vowel inventory relative to speakers of L1s with larger vowel inventories (such as German and Norwegian). This is because several L2 vowels may be perceived as similar to just one L1 vowel category and consequently will be hard to discriminate (Iverson & Evans, 2007).

Perceptual sensitivity to L2 vowel duration has been shown to be a more highly salient cue to vowel identity than spectral information. For example, Japanese learners of English systematically distinguish American English tense (/i:/, eɪ, æ,

ɑ, ɔ, oʊ, u:/) and lax (/ɪ, ε, ʌ, ʊ/) vowels (Strange et al., 1998), Australian English (AusE) vowels /æ/ and /ɛ:/ (Ingram & Park, 1997), and Canadian English /ɪ/ and /i/ (Morrison, 2002) on the basis of duration, even though the *durational* differences per se are not phonemic in any of these three dialects of English, for which the relevant dimensions are tense and lax, leading English listeners to rely on spectral rather than purely temporal differences. The high saliency of durational differences is thus particularly strong for speakers of languages where duration is phonemic (including vowel length contrasts), such as Estonian (McAllister, Flege, & Piske, 2002), although it has also been documented to a lesser extent for speakers of L1s where vowel duration is not phonemic but who have a tense–lax distinction, such as English and German (Bohn, 1995; Bohn & Flege, 1992; Cebrian, 2006; Flege et al., 1997; Gottfried & Beddor, 1988; Goudbeek, Cutler, & Smits, 2008), and even less so for speakers of languages without length (gemination) and tense–lax distinctions, such as Spanish.

Nevertheless, differences and interactions between the L1 and L2 phonological systems of a learner does not mean that an L2 learner's perception will never improve or come to more closely resemble that of a native speaker of the L2. A large body of literature documents that a learner's perceptual difficulties are partly determined by their level of familiarity with and use of the L2, and that increased L2 exposure and use typically leads to improved perception and production in the L2 (Flege et al., 1995; Ingram & Park, 1997; Tsukada et al., 2005).

Unfortunately, it is not clear how this change in perception (and production) occurs, and the underlying mechanisms have not been adequately addressed by the two primary theoretical models of L2 production and perception: the speech learning model (SLM; Flege, 1995), which focuses on experienced L2 speakers, and the perceptual assimilation model (PAM; Best, 1995), which focuses on naive listeners. Recently, an extension of PAM, PAM-L2 (Best & Tyler, 2007), which focuses on L2 development from within the general PAM framework, has attempted to fill this theoretical gap as well as provide testable hypotheses for how perceptual changes come about.

PAM (Best, 1994, 1995) assumes that L1 acquisition is essentially the fine-tuning of the perceptual systems to those articulatory gestures in the L1 that are meaningful and that naive perception of L2 phones will reflect this L1 tuning in terms of the phonetic realizations and the phonological organization of the speaker's native language. PAM-L2 (Best & Tyler, 2007) also assumes that perceptual learning is possible at all ages but will be influenced by the entire language learning history of the individual.

PAM-L2 further posits that an increased L2 vocabulary may “exert forceful linguistic pressure” (Best & Tyler, 2007, p. 32) on the learner to attune to articulatory, phonetic, and phonological differences in the L2 that have previously been ignored in the L1. That is, for successful L2 comprehension the learner must differentiate between an increasing number of contrasting L2 words that initially sound identical through the L1 phonological system. We propose that this consequently causes the learner to rephonologize, that is, establish an L2 phonology, by modification of or addition to the learner's existing L1 phonological system. This vocabulary-driven rephonologization is proposed to work in a similar fashion to the emergence of L1 phonological function that appears to be driven by the

“vocabulary explosion” in children at around 18 months (e.g., Metsala & Walley, 1998). We use the term phonological function to mean the ability of adults and even word-learning 19-month-old toddlers to recognize words on the basis of their phonological (abstract) identity, rather than on the basis of familiarity with the surface, phonetic patterns of the specific utterances they have experienced (see Best, Tyler, Gooding, Orlando, & Quann, 2009).

The vocabulary-driven linguistic pressure may be particularly strong in densely populated lexical neighborhoods with many minimal pairs, especially where several contrastive L2 phones are assimilated into the same L1 category. It is important to note that we do not imply that reattunement and rephonologization in L2 learners necessarily leads to the formation of a complete *new* set of L2 phonological categories, but rather that the learner settles on a phonological system for the L2, perhaps by stretching the L1 inventory as far as possible, and perhaps by forming new L2 categories.

Given that most theories of L2 acquisition (such as PAM/PAM-L2 and SLM) assume that L2 acquisition is based on the same processes of acquisition as is L1 acquisition (although the language history of an individual most often results in L2 acquisition being affected by the L1 of that learner), the suggestion that L2 reattunement and rephonologization is closely associated with L2 vocabulary development does have some support. Research into L1 acquisition in young infants and children suggests a circular or cyclic relationship between segmental speech perception, vocabulary size, and speech production abilities (see, e.g., Smith, McGregor, & Demille, 2006). Early vocabulary size has been found to predict the speech perception abilities of children between the ages of 2 and 6 (Metsala, 1999), highlighting the interdependency of vocabulary development and speech perception. Segmental perception (of *phonetic* contrasts) in the first year of life has also been found to predict vocabulary size in the second year of life (Tsao, Liu, & Kuhl, 2004), suggesting that successful phonetic attunement facilitates word learning, which in turn supports the onset of phonologization at the time of the “vocabulary expansion” (see, e.g., Stager & Werker, 1997; Swingley, 2003).

The present study tests the PAM-L2 based hypothesis that a larger L2 vocabulary drives a process of rephonologization for adult L2 learners. We do so by examining the similarities and differences in L2 speech perception in two groups of speakers (differing in the size of their L2 vocabulary, but not on any other central measure such as age of acquisition, L2 immersion duration, and number of years of foreign language acquisition) who speak a language with a limited vowel inventory and duration-based phonological vowel distinctions (Japanese), and who are acquiring an L2 with a larger vowel inventory but no pure duration-based phonological distinctions (AusE).

This L1–L2 combination has been traditionally seen as posing a significant challenge to Japanese learners, who will have to adjust their limited vowel inventory of just five unique, and relatively pure, vowels as these occur in five spectrally similar short–long pairs (/i, i:, e, e:, a, a:, u, u:, o, o:/; International Phonetic Association, 1999). However, when the possible Japanese bimoraic combinations (/ie, ia, iu, io, ei, ea, eu, eo, ai, ae, au, ao, ui, ue, ua, uo, oi, oe, oa, ou/) are added to the L1 Japanese learners’ category inventory, the task of mapping the 18 AusE vowels

(monophthongs /i:, ɪ, e, ɛ:, ɜ:, ə, æ, ɐ, ɞ:, ɔ:, ʊ, ʉ:, ɔ/ and diphthongs /ɪə, æɪ, əɛ, ɔɪ, əʉ, æɔ:/; Cox, 2006) may be less taxing. However, it complicates the picture that all AusE monophthongs are characterized by some degree of spectral change during production and are generally differentiated on the basis of Frequency 1 (F1) and F2, with the exception of /i:/-/ɪ/ and /ɐ/-/ɞ:/, which are differentiated mainly on the basis of duration (Cox, 2006).

In a cross-language mapping experiment, we adopted a “whole-system” approach and presented the learners with the entire AusE vowel inventory and all Japanese monomoraic or bimoraic categories possible. This approach allows a more complete assessment of the perceptual flexibility of the learners than would the presentation of a preselected subset of the L2 vowels because it provides the Japanese learners with the full range of L2 input as well as the opportunity to exploit all native sensitivities as well as L2 vowel differences, spectral and durational.

In line with the PAM-L2 hypothesis that L2 vocabulary size contributes centrally to L2 phonological reattunement (Best & Tyler, 2007), we expected L2 learners with a larger L2 vocabulary to more consistently identify L2 vowels in terms of their L1 vowel categories than learners with a smaller L2 vocabulary. This is based on the reasoning that the more advanced acquisition of L2 vocabulary by the former subgroup would have already driven them to perceptually attune to nonnative phonetic and phonological differences and to begin to reorganize their native phoneme inventory to accommodate the L2 vowel system, at least to some degree (and/or to begin to establish new L2 vowel categories for certain L1-L2 differences, as discussed below). This is compatible with both PAM and SLM, which propose (a) that learners are likely to perceive (and produce) nonnative phones on the basis of their similarity to or dissimilarity from existing L1 phones (on a scale from *new* to *similar* to *identical*), and (b) that more advanced learners may be more successful in integrating the L2 phones into their existing phonological system (or establishing new L2 phonemes), because their increased experience with the (vocabulary of their) L2 may have prompted them to better attune to the meaningful (i.e., phonological) differences in the L2 and more successfully use their L1 categories and sensitivities to differentiate between (and possibly create novel categories for) these L2 phones.

In the case of L2 vowel acquisition, we further suggest that it is unlikely that an L2 vowel will be perceived as an entirely new category (i.e., as unrelated to any L1 vowel category) in early acquisition, but rather that “difficult” L2 vowels will be initially be perceived as somewhat similar to a number of L1 vowels, in line with research indicating that vowels are produced in a manner less discrete than consonants and rather continuously perceived. We also suggest that such vowels will be more consistently identified as belonging to just one L1 category through further perceptual reattunement and rephonologization, which may allow the learner to identify those aspects of the uncategorized L2 phone that most systematically relate to just *one* L1 category (facilitating L2 comprehension), or realize that the phone does not systematically relate to *any* L1 category, leading perhaps to the formation of a new L2 category.

Further, in line with PAM and PAM-L2, we expected that the Japanese learners would be highly sensitive to durational differences as well as to some spectral differences in the AusE vowel system. Specifically, we expected that short AusE

monophthongs would be identified as Japanese monomoraic vowels, long AusE monophthongs would be identified as bimoraic identical (“long”) Japanese vowels, and diphthongs would be identified as instances of bimoraic Japanese vowel combinations.

In agreement with PAM, our expectation was that the perceptual assimilations of AusE vowels by the learners would reflect not only gradient phonetic similarities (acoustic or articulatory) of their native vowel inventory with the nonnative phones but also the perceived L1/L2 similarities on an abstract, lexically distinctive phonological level.

## METHOD

### *Stimuli*

Eight male L1 Western Sydney Australian English (AusE<sub>WS</sub>) speakers ( $M_{\text{age}} = 22.7$  years) produced five randomized repetitions of the 11 AusE stressed monophthongs and 7 diphthongs in a first syllable stressed /hVbɐ/ context in citation form and in a carrier sentence context (“I say /hVbɐ/ for fun”). The /hVbɐ/ context was chosen to minimize consonant coarticulation effects on the vowels (Strange et al., 2007). Although we tentatively suggest that lexical neighborhood density may generally play a role in rephonologization by increasing pressure on the learner to attune to the phonological organization of the L2, the present study did not directly assess the effect of lexical neighborhood. The focus was instead on the role of vocabulary size in perceptual reattunement and rephonologization. The fact that the selected /hVbɐ/ context yields but one AusE real word (harbor) may be a prudent approach as any effects that the nonword nature of the disyllables may have on the performance of the learners is likely to be in the direction of *reducing* the effect of the learners’ vocabulary on their L2 perception, that is, we took a conservative approach that mitigates against our hypothesis.

Speakers were instructed to talk as if to a friend, at a normal conversational speaking rate. Three speakers were selected on the basis of subjective judgments of similar voice quality, general reading style, and speed by two phonetically trained experimenters. Three tokens of each vowel were selected from each of those three speakers on the basis of subjective judgements of similar intonation pattern and speaking rate.

The recording took place in a sound-attenuated room at the University of Western Sydney, using a Shure SM10A headset microphone, an LG laptop computer, and an external soundcard (Edirol UA-25). The recordings were high-pass filtered in Cool Edit with a cutoff frequency of 75 Hz to eliminate any possibility of low-frequency rumble or electrical noise from the recording. This did not impair the intonation of the utterances, as that none of the f0 contours for the selected tokens dipped below 75 Hz. The intensity of all words was adjusted in Praat so that the root mean square intensity of the target vowel was equal to 70 dB (Boersma & Weenink, 2008).

Vowel onset was defined as the beginning of the first regular and recognizable pitch pulse, and vowel offset was defined as the cessation of regular pitch pulsing. Voicing during the closure for /b/ was not considered part of the vowel. Following

Table 1. Citation-context vowel duration and frequency 1, 2, and 3 values at 25%, 50%, and 75% of the target vowels and diphthongs of Western Sydney Australian English

Vowel	Duration (ms)	F1 (Hz)			F2 (Hz)			F3 (Hz)		
		25%	50%	75%	25%	50%	75%	25%	50%	75%
/i:/	128	335	317	310	2169	2216	2219	3102	3148	3032
/ɪ/	59	333	342	344	2119	2085	1910	2983	2706	2424
/ɪə/	129	327	325	341	2172	2167	2142	3132	3118	2405
/e/	74	547	558	528	1873	1829	1697	2565	2474	2388
/e:/	143	512	523	526	1833	1869	1789	2622	2552	2543
/æ/	85	711	695	646	1535	1487	1412	2618	2570	2538
/eɪ/	130	540	489	409	1910	2006	2020	2673	2634	2479
/æɔ/	142	708	731	681	1524	1367	1226	2477	2417	2540
/aɛ/	147	708	702	603	1185	1374	1575	2589	2533	2562
/ɐ/	73	698	678	628	1198	1171	1102	2719	2734	2703
/ɐ:/	141	672	682	657	1262	1265	1227	2636	2640	2543
/ɔ/	70	618	608	552	933	930	904	2760	2771	2745
/o:/	119	454	497	491	698	694	944	2825	2838	2859
/oi/	142	499	474	387	1079	1549	1935	2530	2486	2537
/ʊ/	76	378	385	361	929	951	986	2528	2532	2531
/u:/	110	376	378	371	1579	1544	1449	2197	2201	2184
/əʊ/	116	570	522	467	1240	1310	1321	2411	2365	2302
/ɜ:/	144	487	504	490	1462	1468	1425	2409	2421	2405

previous research, such as Strange et al. (1998), vowel duration as well as F1, F2, and F3 values at 25%, 50%, and 75% of the target vowel of the first syllable were estimated. The formant values were estimated using the Praat command “To Formant (burg).” The time step was 2.5 ms, the maximum number of formants was 5, the maximum formant frequency was 5 kHz, and the window length was 25 ms. Preemphasis was from 50 Hz. The obtained measurements are shown in Table 1 and Table 2.

Six dependent variables (F1 and F2 at 25%, 50%, and 75% of the vowel) were analyzed in an  $18 \times 2$  multivariate analysis of variance with the independent variables of vowel type and presentation context (sentence vs. citation presentation), which revealed that there was no main effect of presentation context,  $F(6, 282) = 1.207$ , *ns*. There was, however, a significant interaction between presentation context and vowel type,  $F(102, 1,722) = 1.693$ ,  $p < .001$ . Exploring the effect of presentation context on vowel type, we analyzed nine dependent variables (F1, F2, and F3 at 25%, 50%, and 75% of each of the vowels) with the independent variable of presentation context. There was a significant effect of context only for four diphthongs, /eɪ/,  $F(9, 8) = 4.511$ ,  $p = .023$ ; /ɪə/,  $F(9, 8) = 3.706$ ,  $p = .039$ ; /aɛ/,  $F(9, 8) = 3.589$ ,  $p = .043$ ; and /oɛ/,  $F(9, 8) = 12.148$ ,  $p = .001$ , possibly reflecting slight differences in the timing of the transition from one vowel target to the other in the two contexts. We did not include vowel duration in our analysis, as any difference in duration would most likely have reflected

Table 2. *Sentence-context vowel duration and frequency 1, 2, and 3 values at 25%, 50%, and 75% of the vowels and diphthongs of Western Sydney Australian English*

Vowel	Duration (ms)	F1 (Hz)			F2 (Hz)			F3 (Hz)		
		25%	50%	75%	25%	50%	75%	25%	50%	75%
/i:/	140	330	304	308	2198	2116	2250	3052	3008	3031
/ɪ/	66	330	347	364	2182	2166	2039	2989	2955	2437
/ɪə/	145	305	316	339	2241	2210	2215	2831	2946	2827
/e/	92	541	559	534	1793	1830	1750	2434	2459	2353
/e:/	162	488	506	512	2021	1945	1908	2611	2426	2397
/æ/	91	806	763	678	1560	1479	1450	2381	2464	2487
/eɪ/	149	607	489	399	1840	1942	2096	2415	2441	2498
/æɔ/	160	752	752	684	1627	1462	1290	2298	2472	2335
/æe/	159	710	738	630	1157	1344	1638	2534	2449	2465
/ɐ/	69	736	697	628	1208	1171	1112	2644	2624	2630
/ɛ:/	149	726	725	699	1309	1278	1266	2613	2631	2567
/ɔ/	69	611	586	527	976	969	931	2520	2534	2540
/o:/	127	441	440	441	751	734	756	2533	2526	2525
/oɪ/	151	469	477	403	975	1375	1936	2411	2263	2289
/ʊ/	67	342	359	348	905	897	876	2444	2435	2468
/u:/	141	513	494	429	1344	1380	1381	2232	2176	2212
/əu/	141	513	494	429	1344	1380	1381	2232	2176	2212
/ɜ:/	165	487	519	511	1510	1486	1449	2328	2332	2284

speaking rate rather than intrinsic differences in the vowels between the sentence and citation conditions.

The formant values reported in Tables 1 and 2 are consistent with the vowel data reported by Cox (2006), with the exception of the diphthong /æɪ/. In our data, this diphthong is produced with a much higher (and also slightly more fronted) first target than /æ/ and should probably be transcribed as [eɪ].

### Participants

Eleven native speakers of Japanese (8 female, mean age = 26.4 years) participated in the study. All had studied English in middle and high school mainly with native Japanese teachers, as well as some native English guest teachers (mean years of study = 7, mean age of onset = 11.6 years). Although two participants had only Japanese teachers, the others had also been exposed to either AusE or British English, although to a lesser extent than their exposure to Japanese accented English. Five participants had been exposed to more than one native English dialect (with two reporting exposure to four different dialects of English). A tally of the countries of origins of the teachers showed that nine participants had teachers who spoke AusE; seven had British English-speaking teachers, three had American English-speaking teachers, two had New Zealand English-speaking teachers, and two had Irish English-speaking teachers. Of course, all had contact with various dialects



of English through music and films and other mass media. The highly dialectally varied L2 input that these students had received most likely contributed to some degree of target confusion (as reported for native Danish learners of English, Bohn & Bundgaard-Nielsen, 2009) instead of a specific dialectal bias. All participants reported that they did not yet feel “confident” in speaking English. They were all students at English colleges in Sydney, their goal was to learn spoken English, and all had spent less than 12 weeks in Australia (*mean* stay = 7 weeks). None had previously lived outside of Japan.

### *Procedure*

All participants were tested in a quiet, sound-attenuated experiment room at MARCS Auditory Laboratories at University of the Western Sydney. Stimuli were presented over studio headphones from a Mac Book using PsyScope. The participants first heard a randomized presentation of all AusE<sub>WS</sub> vowels over headphones ( $N = 324$ ) in citation and complete carrier sentence contexts (presentation order of the two contexts was counterbalanced across participants) and identified them in terms of their L1 vowels on a computer screen, using a grid of Japanese katakana symbols representing all short, long, and combination vowels possible: /i, i:, e, e:, a, a:, u, u:, o, o:, ie, ia, iu, io, ei, ea, eu, eo, ai, ae, au, ao, ui, ue, ua, uo, oi, oe, oa, ou/ preceded by /h/. In addition, the participants rated goodness of fit of each vowel on a scale from 1 (*poor*) to 7 (*excellent*). Finally, they completed a multiple-choice L2–English vocabulary size test (Nation & Beglar, 2007). The Nation and Beglar vocabulary size test was selected for two reasons. The first and most important reason is that, rather than assessing vocabulary acquisition on different “levels” of vocabulary (see, e.g., Laufer & Nation, 1999; Nation, 1993), it offers an estimate of a learner’s recognition vocabulary irrespective of which level of vocabulary the learner has focused on in his or her L2 acquisition process. The second reason is that, although tests of general proficiency will likely also be associated with segmental perception (perhaps as a *result* of improvements to L2 vocabulary size and the proposed related improvements to segmental perception), they include other measures of L2 proficiency that have not previously been directly posited to contribute centrally to the process of (re)phonologization, as has vocabulary development in L1 learning infants and children. The possibility of an analogous vocabulary growth effect on L2 speech perception was originally proposed in PAM-L2 (Best & Tyler, 2007), and was examined for the first time in the present report.

## RESULTS

### *General assimilation pattern*

An L2 vowel was defined as “categorized” if it was identified as an instance of one L1 category in more than 50% of presentations (following Best, Faber, & Levitt, 1996). Thirteen of the 18 AusE vowels were consistently assimilated to Japanese vowel categories in both the citation and sentence contexts. Five were uncategorized, that is, not identified as any one Japanese vowel category for 50% or more of tokens (see Table 3).

Table 3. *The assimilation of citation Australian English to first language Japanese vowel categories by all listeners*

Categorized to L1							Uncategorized
Single Mora		Double Mora		Bimoraic Combinations			
ID	GR	ID	GR	ID	GR		
/ɪ/ → /i/	4.90	/i:, Iə/ → /i:/	4.83/4.65	/æ/ → /ai/	5.10	/ɜ:/	
/e/ → /e/	5.12	/e:/ → /e:/	4.84	/æi/ → /ei/	5.29	/æ/	
/ɐ/ → /a/	5.13	/ɛ:/ → /a:/	4.87			/o:/	
/ɔ/ → /o/	4.86			/oi/ → /oi/	5.27	/æɔ/	
/ʊ/ → /u/	5.05	/u:/ → /u:/	4.88			/əu/	

Note: The Australian English and first language Japanese are the left and right items in the ID column, respectively. GR, mean goodness rating on a 7-point scale (1 = *poor fit*, 7 = *excellent fit*).

The pattern of assimilation did not differ for citation and sentence presentations, such that an L2 phone, which was identified as belonging to a particular L1 category in one presentation context, was also identified as belonging to that L1 category in the other context. Results from both contexts thus reflected the Japanese learners' sensitivity to spectral as well as durational information in similar ways. Most L2 vowels were assimilated into unique L1 categories, although both AusE /i:/ and /Iə/ were assimilated to Japanese /i:/ (i.e., tense). AusE<sub>WS</sub> vowels were typically identified as bimoraic Japanese vowels, and short (i.e., lax) AusE<sub>WS</sub> vowels were typically identified as monomoraic Japanese vowels. The spectrally similar (but durationally different) vowel pairs /e/-/e:/ and /ɐ/-/ɛ:/ were successfully assimilated to four separate Japanese vowel categories, most likely on the basis of systematic durational differences, although the slight spectral differences may also have contributed.

The average goodness ratings differed for vowels in citation ( $M = 4.96$ ) and sentence contexts ( $M = 5.25$ ),  $F(1, 24) = 8.691$ ,  $p < .01$ . All 30 possible Japanese categories were used at least once by at least one participant in both citation and sentence contexts. The identification of AusE<sub>WS</sub> /ei/ as Japanese /ei/ seems to reflect the different realization of this diphthong between our data and that of previous work on AusE (Cox, 2006). The identification of AusE<sub>WS</sub> /ei/ and AusE<sub>WS</sub> /e:/ as two different Japanese categories /ei/ and /e:/; thus, reflects both phonetic similarities between the L2 vowels and the chosen L1 categories, and phonological as both of these two Japanese categories are actually realized as [e:] (see Tsujimura, 1996, p. 32).

*Analysis by vocabulary size*

The mean estimated vocabulary size (from the Nation & Begar test) for the 11 listeners was 6,009 words (range = 4,100–7,800 words). To test the hypothesis that

L2 vocabulary development is a core contributing factor in L2 vowel phonological reorganization, they were divided into high (HV) and low vocabulary (LV) groups using a median split (HV:  $n = 5$ ,  $M_{\text{vocabulary}} = 7,200$ ; LV:  $n = 6$ ,  $M_{\text{vocabulary}} = 5,017$ ). A series of two-tailed paired  $t$  tests revealed that the HV and LV group differed in terms of their L2 vocabulary size,  $t(9) = 5.40$ ,  $p = .001$ , but they did not differ in terms of years of English study in Japan,  $M_{\text{HV}} = 7.8$  years,  $M_{\text{LV}} = 6.3$  years,  $t(9) = .140$ ,  $p = .89$ , or in their length of residence in Australia on the day of testing,  $M_{\text{HV}} = 7.2$  weeks,  $M_{\text{LV}} = 6.8$  weeks,  $t(9) = 1.976$ ,  $p = .10$ .

As predicted, differences in the identification patterns of the learners were related to their vocabulary size. The general identification pattern for the HV and LV groups was almost identical (see Table 3 for that common pattern), given that only one L2 vowel, AusE<sub>WS</sub> /ae/, reached the “categorized” criterion for the HV group (94% perceived as /ai/) but not for the LV group (49% perceived as /ai/) in the C context. However, the two groups differed systematically in two other aspects of their performance on the assimilation task. First, the HV group selected a significantly smaller number of L1 categories for the AusE<sub>WS</sub> vowels ( $M = 25$ ) than the LV group ( $M = 27.67$ ). Paired two-tailed  $t$  tests revealed that this difference was significant: HV–LV for citation context:  $t(17) = 2.77$ ,  $p < .001$ ; HV–LV for sentence context:  $t(17) = 2.40$ ,  $p < .001$ . Second, paired two-tailed  $t$  tests revealed that the HV group was more consistent in its selection of L2 phone to L1 category matching than the LV group: the mean identification score of the categorized vowels for the HV group was 85% (sentence) and 80% (citation), whereas the LV group’s mean was lower at 73% (sentence) and 74% (citation), citation context:  $t(12) = 1.26$ ,  $p < .23$ ; sentence context:  $t(12) = 2.38$ ,  $p < .04$ .

### *Uncategorized vowels*

There was considerable variation in the overall identification pattern for the five uncategorized AusE<sub>WS</sub> vowels. For example, in the case of AusE<sub>WS</sub> /o:/, the three most frequently selected Japanese categories included 91% of the tokens. In contrast, for AusE<sub>WS</sub> /ɜ:/, the three most frequently selected Japanese categories included only an average of 53% of the tokens (see Figure 1).

The HV and LV groups differed in the consistency of their pattern of identification for the uncategorized vowels, with the top three selections for the HV group averaging 74% of responses, and the LV group averaging only 65% of responses,  $t(9) = 14.11$ ,  $p = .001$  (see Table 4). However, the difference in identification consistency between the two presentation contexts was not significant,  $t(9) = 2.17$ .

## DISCUSSION

This preliminary study examined the relationship between L2 vocabulary size and L2 vowel perception in two groups of native Japanese speakers who differed in the size of their L2 English vocabulary. The purpose was to test the PAM-L2 proposal that L2 vocabulary acquisition is central to a phonological reorganization in the learner (Best & Tyler, 2007). We hypothesized that a larger L2 vocabulary would be positively associated with more consistent L2 vowel identification. We expected that all Japanese learners, irrespective of vocabulary, would make use of

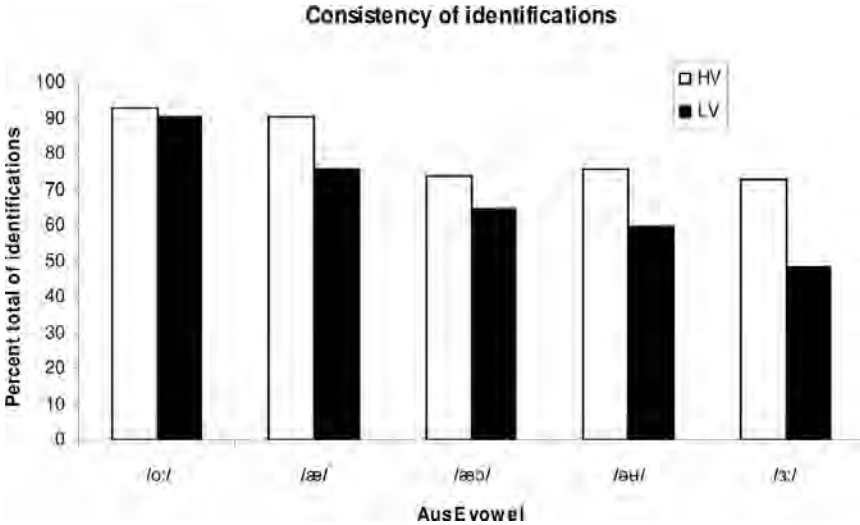


Figure 1. The total percentages of the three most selected Japanese categories for the five uncategorized Western Sydney Australian English vowels for the high vocabulary group (HV) and the low vocabulary group (LV).

spectral and temporal phonetic information in the L2 target items to identify the AusE<sub>WS</sub> vowels with reference to their native Japanese vowel categories. We also expected that the assimilation pattern would reflect not just phonetic similarity, but also phonological similarities.

Our results support the hypothesis that HV and LV L2 learner groups differ in a manner consistent with PAM-L2 predictions. Although the general patterns of identification were similar for the HV and LV groups, the HV learners were not only more consistent in their identification of the L2 phones in terms of L1 assimilation scores, but also in terms of how many alternative L1 categories they selected for each L2 vowel. Further support for a link between vocabulary size and L2 phonological reattunement was found in the consistency of the HV group in their identification of the uncategorized vowels. We interpret this greater consistency as a sign of the HV group being more advanced in integrating the L2 phones into their existing phonological system. Likewise, we interpret the fact that the HV group (but not the LV group) reached identification criterion when identifying AusE<sub>WS</sub> /æi/ as consistent with the PAM-L2 hypothesis about L2 vocabulary effects on L2 vowel perception.

Although we acknowledge that these results do not preclude the possibility that influences other than vocabulary size, such as general proficiency, may underlie the observed differences, our explanation is consistent with the PAM-L2 theoretical framework. A general proficiency explanation would be less parsimonious because it does not provide a clear theoretical link to L2 segmental perception. A prior report on Japanese learners of English differing in L2 fluency alone (vocabulary size was not reported to have been assessed independently in that study)

Table 4. *Assimilation patterns for the top three selected first language Japanese vowel categories (vertical) for the five uncategorized Western Sydney Australian English vowels (horizontal)*

Vowel	Context	Group	1st	%	2nd	%	3rd	%	Total (%)
/o:/	Citation	HV	/ou/	42	/o/	32	/o:/	21	96
		LV	/o:/	47	/ou/	28	/o/	15	90
	Sentence	HV	/o:/	42	/ou/	37	/o/	16	95
		LV	/o:/	44	/ou/	33	/o/	9	86
/æ/	Citation	HV	/a/	39	/a:/	18	/ea/	13	70
		LV	/e/	36	/a/	26	/a:/	16	77
	Sentence	HV	/a/	33	/ea/	28	/e/	20	81
		LV	/a/	42	/e/	20	/a:/	9	70
/æɔ/	Citation	HV	/ea/	42	/a:/	20	/au/	18	80
		LV	/a:/	23	/au/	24	/ea/	19	65
	Sentence	HV	/ea/	45	/e:/	18	/a:/	11	74
		LV	/a:/	19	/ea/	19	/ea/	17	55
/əʊ/	Citation	HV	/ou/	48	/u:/	13	/a:/	4	66
		LV	/ou/	26	/u:/	19	/oi/	9	54
	Sentence	HV	/u:/	41	/ou/	19	/au/	9	69
		LV	/u:/	31	/a:/	10	/au/	9	51
/ɜ:/	Citation	HV	/u:/	31	/ua/	16	/au/	12	59
		LV	/u:/	39	/e:/	9	/ua/	9	56
	Sentence	HV	/u:/	18	/eu/	18	/ua/	14	50
		LV	/u:/	17	/e:/	16	/ue/	15	47

*Note:* Group refers to high vocabulary learners (HV) and low vocabulary learners (LV) and percentage (%) indicates the total percentage of each of the uncategorized Australian English vowels that were identified as each of the top three first language categories.

found that high- and low-proficiency learners did not differ in their identification of AusE vowels /i:, ɪ, e, æ, ɜ:/ using the appropriate AusE labels (Ingram & Park, 1997), indicating that L2 perception is not straightforwardly related to L2 fluency.

The results are also consistent with PAM-L2 claims that L2 learners will use both temporal and spectral information in their identification of the L2 vowels, irrespective of their L2 vocabulary, as the Japanese learners consistently used both mono- and bimoraic L1 categories in a pattern that reflected both native phonetic and native phonological properties. The systematic use of single and bimoraic Japanese categories to identify the short (i.e., lax) and long (i.e., tense) AusE vowels, respectively, is also consistent with the claims that durational information plays an important role in the discrimination of nonnative phones, in particular when the spectral information is ambiguous to the listener (Bohn, 1995; Bohn & Flege, 1992; Cebrian, 2006; Flege et al., 1997; Gottfried & Beddor, 1988). The higher mean goodness rating in the sentence ( $M = 5.25$ ) versus citation context ( $M = 4.96$ ) is likely caused by durational information provided in the carrier sentence that helps learners disambiguate phonological duration information to

which these native Japanese listeners are highly sensitive in their L1. This additional information in the sentential context may also be the reason for the Japanese listeners' use of a smaller number of L1 categories in the sentence context, relative to the citation context.

The insights provided by taking the whole-system approach to L2–L1 vowel perception adopted here (in addition to insights gained from subsystem-based approaches; see, e.g., Strange et al., 1998) are evident in the fact that Japanese L2 learners, irrespective of their L2 vocabulary size, identified the majority of the AusE vowels as members of Japanese vowel categories (mono- as well as bimoraic) in a systematic manner. This pattern of successful assimilation would not have been clear from study of a subset of AusE vowels, whether based on a contrastive analysis of the relatively limited Japanese singleton-vowel inventory and the much larger AusE vowel inventory, or the previous literature on English vowel acquisition by Japanese speakers. Those earlier studies did not make the full L1 inventory of tools to differentiate vowels available to the learners. However, the relative success of the Japanese learners in exploiting both spectral and durational vowel settings of the L1 is consistent with PAM claims that L2 phones are classified in terms of L1 phonemes when they are not perceived as too discrepant from the L1 phonemes, and learners will make use of whatever information they can to make sense of the nonnative system.

## CONCLUSION

We propose that future research continues to explore the role of L2 vowel acquisition in L2 segmental perception and that the “whole-system” approach to perceived L1/L2 vowel similarities, pioneered here, could fruitfully be adopted to gain a better understanding of the flexibility of a learner's perceptual system, in particular, with respect to studies of L2 vowel perception. Furthermore, in light of research suggesting that a learner's L2 phonological development plateaus within the first year of acquisition in immersion settings, and that the most significant changes in nonnative vowel perception are apparent after only 6–12 months of L2 immersion for adult learners (Aoyama, Flege, Guion, Akahane-Yamada, & Yamada, 2004; Flege & Liu, 2001; Tsukada et al., 2005), we encourage a meticulous examination of the relationship between L2 vocabulary growth and L2 phonological acquisition during the first 12 months of immersion. Such studies would help determine exactly when during the first 12 months of immersion the learner reaches this plateau, when and how the phonological reorganization begins or peaks, and the extent to which it asymptotes. In conclusion, our findings are consistent with the PAM-L2 prediction that L2 vocabulary acquisition drives a learner to rephonologize early in immersion-based conversational acquisition.

## ACKNOWLEDGMENTS

This work was supported by US National Institutes of Health (NIH) Grant DC00403 (to C.B.) and by UWS Research Training Scheme funds (to R.L.B.-N.) from MARCS Auditory Laboratories.

## REFERENCES

- Aoyama, K., Flege, J. E., Guion, S. G., Akahane-Yamada, R., & Yamada, T. (2004). Perceived phonetic dissimilarity and L2 speech learning: The case of Japanese /r/ and English /l/ and /r/. *Journal of Phonetics*, 32, 233–250.
- Best, C. T. (1994). The emergence of native-language phonological influences in infants: A perceptual assimilation model. In J. C. Goodman & H. Nusbaum (Eds.), in *The development of speech perception: The transition from speech sounds to spoken words* (pp. 167–224). Cambridge, MA: MIT Press.
- Best, C. T. (1995). A direct-realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 171–204). Timonium, MD: York Press.
- Best, C. T., Faber, A., & Levitt, A. (1996). *Perceptual assimilation of non-native vowel contrasts to the American English Vowel System*. Paper presented at the Meeting of the Acoustical Society of America, Indianapolis, IN.
- Best, C. T., & Strange, W. (1992). Effects of language-specific phonological and phonetic factors on cross-language perception of approximants. *Journal of Phonetics*, 20, 305–330.
- Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In J. Munro & O.-S. Bohn (Eds.), *Second language speech learning: The role of language experience in speech perception and production* (pp. 13–34). Amsterdam: John Benjamins.
- Best, C. T., Tyler, M. D., Gooding, T., Orlando, C., & Quann, C. (2009). Toddlers' spoken word familiarity preferences in native versus nonnative dialects. *Psychological Science*, 20, 539–542.
- Boersma, P., & Weenink, D. (2008). Praat: Doing phonetics by computer (Version 5.0.27) [Computer program]. Retrieved June 28, 2008, from <http://www.praat.org/>
- Bohn, O.-S. (1995). *What determines the perceptual difficulty encountered in the acquisition of non-native contrasts?* Paper presented at the 13th International Conference of Phonetic Sciences, Stockholm.
- Bohn, O.-S., & Bundgaard-Nielsen, R. L. (2009). Second language speech learning with diverse input. In M. Young-Scholten & T. Piske (Eds.), *Input matters in SLA* (pp. 207–218). Clevedon: Multilingual Matters.
- Bohn, O.-S., & Flege, J. E. (1992). The production of new and similar vowels by adult German learners of English. *Studies in Second Language Acquisition*, 14, 131–158.
- Cebrian, J. (2006). Experience and the use of non-native duration in L2 vowel categorization. *Journal of Phonetics*, 34, 372–387.
- Cox, F. (2006). The acoustic characteristics of /hVd/ vowels in the speech of some Australian teenagers. *Australian Journal of Linguistics*, 26, 147–179.
- Escudero, P., & Boersma, P. (2004). Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition*, 26, 551–585.
- 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. (1995). Second language speech learning—Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233–277). Baltimore, MD: York Press.
- Flege, J. E. (2002). Interactions between the native and second-language phonetic systems. In P. Burmeister, T. Piske, & A. Rohde (Eds.), *An integrated view of language development—Papers in honor of Henning Wode* (pp. 217–244). Trier, Germany: Wissenschaftlicher Verlag Trier.
- Flege, J. E., Bohn, O.-S., & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics*, 25, 437–470.
- Flege, J. E., & Liu, S. (2001). The effect of experience on adults' acquisition of a second language. *Studies in Second Language Acquisition*, 23, 527–552.
- Flege, J. E., & MacKay, I. R. A. (2004). Perceiving vowels in a second language. *Studies of Second Language Acquisition*, 26, 1–34.
- Flege, J. E., Munro, M., & MacKay, I. R. A. (1995). Effects of age of second language learning on the production of English consonants. *Speech Communication*, 16, 1–26.
- Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds “L” and “R.” *Neuropsychologia*, 9, 317–323.

- Gottfried, T. L., & Beddor, P. S. (1988). Perception of temporal and spectral information in French vowels. *Language and Speech, 31*, 57–75.
- Goudbeek, M., Cutler, A., & Smits, R. (2008). Supervised and unsupervised learning of multidimensionally varying nonnative speech categories. *Speech Communication, 50*, 109–125.
- Guion, S. G., Flege, J. E., Akahane-Yamada, R., & Pruitt, J. C. (2000). An investigation of current models of second language speech perception: The case of Japanese adults' perception of English consonants. *Journal of the Acoustical Society of America, 107*, 2711–2724.
- International Phonetics Association. (1999). *Handbook of the International Phonetic Association*. Cambridge, MA: Author.
- Ingram, J. C. L., & Park, S.-G. (1997). Cross-language vowel perception and production by Japanese and Korean learners of English. *Journal of Phonetics, 25*, 343–370.
- Iverson, P., & Evans, G. E. (2007). Learning English vowels with different first-language vowel systems: Perception of formant targets, formant movement, and duration. *Journal of the Acoustical Society of America, 122*, 2842–2854.
- Jenkins, J. J., Strange, W., & Polka, L. (1995). Not everyone can tell a “rock” from a “lock”: Assessing individual differences in speech perception. In D. J. Lubinski & R. V. Dawis (Eds.), *Assessing individual differences in human behavior: New concepts, methods, and findings* (pp. 297–325). Palo Alto, CA: Davies-Black Publishing.
- Krebs-Lazendic, L., & Best, C. T. (2008). Early and late bilinguals' vowel perception and production: English vowel contrasts that give Serbian–English bilinguals a H(E)AD-ache. In A. S. Rauber, M. A. Watkins, & B. O. Baptista (Eds.), *New sounds 2007: Proceedings of the Fifth International Symposium on the Acquisition of Second Language Speech*. Florianópolis, Brazil: Federal University of Santa Catarina.
- Laufer, B., & Nation, P. (1999). A vocabulary size test of controlled productive ability. *Language Testing, 16*, 33–51.
- McAllister, R., Flege, J. E., & Piske, T. (2002). The influence of the L1 on the acquisition of Swedish vowel quantity by native speakers of Spanish, English and Estonian. *Journal of Phonetics, 30*, 229–258.
- Metsala, J. L. (1999). Young children's phonological awareness and nonword repetition as a function of vocabulary development. *Journal of Educational Psychology, 91*, 3–19.
- Metsala, K. L., & Walley, A. C. (1998). Spoken vocabulary growth and the segmental restructuring of lexical representations: Precursors to phonemic awareness and early reading ability. In J. L. Metsala & L. C. Ehri (Eds.), *Word recognition in beginning literacy* (pp. 89–120). Mahwah, NJ: Erlbaum.
- Morrison, G. S. (2002). Japanese listeners' use of duration cues in the identification of English high front vowels. In J. Larson & M. Paster (Eds.), *Proceedings of the 28th Annual Meeting of the Berkeley Linguistics Society* (pp. 189–200). Berkeley, CA: Berkeley Linguistics Society.
- Munro, M., Flege, J. E., & MacKay, I. (1996). The effect of age of second language learning on the production of English vowels. *Applied Psycholinguistics, 17*, 313–334.
- Nation, P. (1993). Measuring readiness for simplified reading: A test of the first 1000 words of English. *RELC Journal, 31*, 193–203.
- Nation, P., & Beglar, D. A. (2007). *A vocabulary size test, JALT2007*. Retrieved April 1, 2008, from <http://jalt-publications.org/tlt/resources/2007/0707a.pdf>
- Smith, B. L., McGregor, K. K., & Demille, D. (2006). Phonological development in lexically precocious 2-year-olds. *Applied Psycholinguistics, 27*, 355–375.
- Stager, C. L., & Werker, J. F. (1997). Infants listen for more phonetic detail in speech perception than in word learning tasks. *Nature, 388*, 381–382.
- Strange, W. (1998a). Perception of vowels: Dynamic constancy. In J. M. Pickett (Ed.), *The acoustics of speech communication: Fundamentals, speech perception theory, and technology* (pp. 152–165). Boston: Allyn & Bacon.
- Strange, W. (1998b). Perception of consonants: From variance to invariance. In J. M. Pickett (Ed.), *The acoustics of speech communication: Fundamentals, speech perception theory, and technology* (pp. 166–182). Boston: Allyn & Bacon.
- Strange, W., Akahane-Yamada, R., Kubo, R., Trent, S. A., Nishi, K., & Jenkins, J. J. (1998). Perceptual assimilation of American English vowels by Japanese listeners. *Journal of Phonetics, 26*, 311–344.



- Strange, W., Weber, A., Levy, E. S., Shafiro, V., Hisagi, M., & Nishi, K. (2007). Acoustic variability within and across German, French and American English vowels: Phonetic context effects. *Journal of the Acoustical Society of America*, 122, 1111–1129.
- Swingle, D. (2003). Phonetic detail in the developing lexicon. *Language and Speech*, 46, 265–294.
- Tsao, F.-M., Liu, H.-M., & Kuhl, P. K. (2004). Speech perception in infancy predicts language development in the second year of life: A longitudinal study. *Child Development*, 75, 1067–1084.
- Tsujimura, N. (1996). *An introduction to Japanese linguistics*. Malden, MA: Blackwell.
- Tsukada, K., Birdsong, D., Bialystok, E., Mack, M., Sung, H., & Flege, J. E. (2005). A developmental study of English vowel production and perception by native Korean adults and children. *Journal of Phonetics*, 33, 263–290.
- Werker, J., & Tees, R. C. (1984). Phonemic and phonetic factors in adult cross-language speech perception. *Journal of the Acoustical Society of America*, 75, 1866–1878.