

Cross-linguistic studies of children's and adults' vowel spaces^{a)}

Hyunju Chung,^{b)} Eun Jong Kong, Jan Edwards, Gary Weismer, and Marios Fourakis
Department of Communicative Disorders, University of Wisconsin–Madison, Madison, Wisconsin 53706

Youngdeok Hwang

Department of Statistics, University of Wisconsin–Madison, Madison, Wisconsin 53706

(Received 22 September 2010; revised 8 September 2011; accepted 15 September 2011)

This study examines cross-linguistic variation in the location of shared vowels in the vowel space across five languages (Cantonese, American English, Greek, Japanese, and Korean) and three age groups (2-year-olds, 5-year-olds, and adults). The vowels /a/, /i/, and /u/ were elicited in familiar words using a word repetition task. The productions of target words were recorded and transcribed by native speakers of each language. For correctly produced vowels, first and second formant frequencies were measured. In order to remove the effect of vocal tract size on these measurements, a normalization approach that calculates distance and angular displacement from the speaker centroid was adopted. Language-specific differences in the location of shared vowels in the formant values as well as the shape of the vowel spaces were observed for both adults and children.

© 2012 Acoustical Society of America. [DOI: 10.1121/1.3651823]

PACS number(s): 43.70.Kv, 43.70.Ep [AL]

Pages: 442–454

I. INTRODUCTION

It has been widely claimed that vowels are acquired earlier than consonants. This may be why relatively few studies have focused on investigating acquisition patterns of vowels. However, the acquisition patterns of consonants are not separable from those of vowels, and a few studies (e.g., [Davis and MacNeilage, 1990](#)) have suggested that vowel acquisition is considerably more complicated than has been thought. Furthermore, there is relatively little cross-linguistic research on the related question of how children master the language-specific characteristics of vowels in their native language.

Previous cross-linguistic research on vowel production by native adult speakers of different languages has demonstrated systematic differences in how shared vowels (vowels represented by common phonemic symbols) are produced. For example, [Bradlow \(1995\)](#) showed that productions of the shared vowels /i/, /e/, /o/, and /u/, by adult English and Spanish speakers differed systematically with respect to language; all of the English speakers produced vowels with significantly higher $F2$ values than their Spanish counterparts. [Yang \(1996\)](#) also found cross-linguistic differences between shared vowels produced by adult English and Korean speakers. English /u/ had higher $F2$ values than Korean /u/, and English /a/ had lower $F2$ values than Korean /a/. Both observations suggest that the concept of “shared” vowels does not account for subtle differences in vowel production across different languages.

Cross-linguistic differences in the vowel productions of young children have also been documented. [de Boysson-](#)

[Bardies et al. \(1989\)](#) examined cross-linguistic vowel formant patterns in the babbling of 10-month-old monolingual infants of Parisian French, British English, Algerian Arabic, and Hong Kong Cantonese. They found that infants started to show systematic cross-linguistic differences in their vowel formant frequency patterns as early as 10 months of age. [Rvachew et al. \(2006\)](#) and [Rvachew et al. \(2008\)](#) also demonstrated cross-linguistic differences in the vowel spaces of 8- to 18-month-old infants raised in Canadian English and Canadian French linguistic communities. These authors showed that mean $F2$ values in vocalic portions of the babbling of Canadian English infants decreased with age, while these values remained stable in the Canadian French infants. These studies suggest that language-specific characteristics of adults' vowel production are produced by infants even before the age of 1 year. The findings of these studies also demonstrated that research on vowel acquisition patterns needs to examine vowel production at a finer grain than the phonemic level in order to describe developmental patterns. In babbling studies, however, it is impossible to determine if infants are intending to produce a particular vowel sound because babbling, by definition, does not contain words from the ambient language. Therefore, a full account of cross-linguistic vowel acquisition must include a study of young children who are producing words in which the intended vowels can be identified and assumed to be the child's targets.

The present study analyzed the three shared vowels, /a/, /i/, and /u/, as extracted from word-initial consonant-vowel sequences from a word repetition task. The speech samples were from five different languages, Cantonese, American English, Greek, Japanese, and Korean, and were produced by subjects in three different age groups, 2-year-olds, 5-year-olds, and adults. These languages were chosen because of the variation among their vowel inventories and organizations. For example, Cantonese has a large vowel inventory with 11 vowel phonemes (/i:, y:, e, e:, æ:, ø, ø, a:, u:, o, ɔ:/).

^{a)}Portions of this work were presented at the International Child Phonology Conference, West Lafayette, IN 2007; the Symposium on Research in Child Language Disorder, Madison, WI 2007; and the 156th Acoustical Society of America Meeting, Miami, FL 2008.

^{b)}Author to whom correspondence should be addressed. Electronic mail: hyunju1@ualberta.ca

American English also has a relatively large vowel inventory, with four short vowels (/ɪ, ε, ʌ, ʊ/), seven long vowels (/i, e, æ, ɑ, ɔ, o, u/), and one rhotic vowel (/ɝ/), resulting in a total of 12 monophthong vowels (Hillenbrand *et al.*, 1995; Peterson and Barney, 1952). The vowel inventories of Greek and Japanese are similar to each other in terms of number and quality of vowels, both languages having only five vowels (/i, e, a, o, u/). However, to be phonetically accurate, the Japanese high back vowel /u/ should be symbolized as [ɯ], because it is produced with less lip rounding than typically associated with /u/. There is a phonemic length distinction for each of the five vowels in Japanese, though there is no significant difference in the spectral quality of long/short pairs (Keating and Huffman, 1984). Finally, Korean has eight monophthong vowels (/a, e, ε, i, o, u, ʌ, i/). However, the non-high vowels /e/ and /ε/ are almost merged in some dialects, including the Seoul dialect. This is especially notable in female speakers, whose *F1* and *F2* values for /e/ and /ε/ are more similar than those of males (Yang, 1996). For this reason, the Korean vowel inventory is usually analyzed as having seven monophthongs. The vowel inventory of each language is summarized in Table I.

Based on the differences in vowel inventory of each language and results from previous cross-linguistic studies, it was predicted that the acoustic realization of the shared vowels /a/, /i/, and /u/ would show language-specific differences. While these language-specific characteristics may be consistent and systematic, it is possible that they are not easily perceptible to listeners (even to trained phoneticians). Therefore, following previous work, the current study used acoustic analyses to capture the phonetic details of cross-linguistic and developmental differences in vowels across the five target languages and three age groups.

In vowel acoustic studies, it is common to characterize vowels with the first three formant frequencies (*F1*, *F2*, and *F3*), which correspond to the resonant frequencies of the vocal tract. These values, however, do not simply represent linguistic aspects of vowel sounds. Instead, there are confounding effects associated with non-linguistic factors, such as anatomical differences between speakers (differences in vocal tract size). In vowel acoustic studies, therefore, various attempts have been made to eliminate the influence of anatomical factors related to inter-speaker variation on acoustic values, while preserving the linguistic properties of vowel sounds (e.g., Adank *et al.*, 2004; Disner, 1980). While these previous studies normalized the assumed acoustic correlates of vocal tract size differences across speakers, another approach would be to understand the relative distribution of vowel categories in a speaker-specific vowel space. The current study used an approach that transforms the first two formant frequency values into values that characterize vowels in relation to speaker-specific vowel production patterns. This transformation method was inspired by Titze (2000) and Whiteside (2001). Titze (2000) used the neutral vowel /ə/ as a reference to characterize other vowels. Similarly, Whiteside (2001) used speaker-specific centroids of vowel formants by calculating the mean *F1* and *F2* values of each speaker in order to compare vowels across age and sex groups. Adapting their methods, this study calculates the

TABLE I. Vowel inventories of Cantonese, Greek, English, Japanese, and Korean.

Cantonese				
	front	central	back	
high	i: y		u:	
mid	e	ə	o:	
	ε	œ:	ɔ:	
low		ɐ	a:	

English				
	front	central	back	
high	i		u	
	ɪ		ʊ	
mid	e		o	
	ε	ʌ ɜ̃	ɔ	
low	æ		ɑ	

Greek				
	front	central	back	
high	i		u	
mid	e		o	
low	a			

Japanese				
	front	central	back	
high	i		ɯ	
mid	e		o	
low	a			

Korean				
	front	central	back	
high	i	i	u	
mid	e		o	
	ε	ʌ		
low	a			

relative distance and angular displacement of each vowel in relation to a speaker-specific centroid. This way, the vowel spaces of individual speakers can be compared with one another, across languages, age, and sex, presumably free of the effects of varying vocal tract size on the acoustic output of the vocal tract.

TABLE II. Mean age (in months) for 2- and 5-year-olds, and the mean age (in years) for age group by language (standard deviations in parentheses).

		Cantonese	English	Greek	Japanese	Korean
2 yr		30.5 (3.43)	30.7 (3.13)	30.6 (3.72)	30.1 (3.61)	29.8 (3.18)
5 yr		66.3 (4.17)	64.6 (3.19)	65.1 (4.18)	62.5 (1.97)	65.7 (4.79)
Adults	Male	21.4 (1.81)	22.8 (1.64)	21.2 (0.44)	21.2 (0.84)	19.2 (0.83)
	Female	22.4 (1.51)	24 (0.7)	24 (0.84)	21.4 (0.55)	22.2 (2.17)

The aims of this study were to understand how the language-specific phonetic characteristics of shared vowels develop with age, especially whether this pattern differs across languages with different vowel inventories as described earlier. To address these questions, this study employed methods different from those of previous studies. First, it compared vowel production patterns of three different age groups (2-year-olds, 5-year-olds, and adults) across a relatively large number of languages (American English, Cantonese, Greek, Japanese, and Korean). Also, instead of comparing *F1* and *F2* values, data were transformed into speaker-specific values to compare vowel production patterns across different age and language groups free of the influence of vocal tract size.

II. METHODS

The data used in this study are part of a larger study, the *paidologos* project (Edwards and Beckman, 2008a, b), which examined the effects of phoneme and phoneme sequence frequency on phonological acquisition of word-initial lingual consonants across five languages.

A. Participants

Participants in this study were 10 speakers (5 males and 5 females) from each of three age groups (2-year-olds, 5-year-olds, and adults) and five languages. Demographic data for the 150 speakers (10 speakers \times 3 age groups \times 5 languages) are provided in Table II. All participants were monolingual native speakers of Cantonese, American English, Greek, Japanese, or Korean. Dialectal influences on vowel production were controlled by recruiting subjects from only one region of each country; recruitment was done by a native speaker of the same dialect region. Cantonese-speaking subjects were recruited from Hong Kong, China; English-speaking subjects from Columbus, OH; Greek-speaking subjects from Thessaloniki, Greece; Japanese-speaking subjects from Tokyo, Japan; and Korean-speaking subjects from Seoul, Korea. All child participants passed a hearing screening using otoacoustic emissions

and had age-appropriate oromotor skills (Kaufman, 1995). All child participants had normal speech and language development according to parent and teacher reports. All adult participants also passed a hearing screening and reported no history of speech, language, or hearing problems.

B. Stimuli

The stimuli were /a/, /i/, and /u/ vowels in real words that began with lingual obstruent-vowel sequences. The /a/, /i/, and /u/ vowels were chosen because they are common to all five languages, and because they are point vowels that denote the boundaries of the vowel space. For Japanese, both long and short /i/ and /u/ vowels were included in the analysis because there is no systematic difference in formant frequency values between the length variants of the same vowel (Keating and Huffman, 1984). For all languages, the words used to elicit target vowels were real words that were picturable and were most likely to be familiar even to very young children. For each of the word-initial CV sequences, three different words were selected to elicit the target vowel. As an example, the stimulus list for English is provided in Appendix A. All stimulus items that included the target vowels were included in the analysis. Because of the different phonotactic restrictions of each language, it was impossible to have the exact same consonant environment for all languages (e.g., in Cantonese, /u/ is only preceded by the velar /k/).

The number of tokens used for the analysis also varied across languages. A summary of word-initial consonants preceding the target vowels for each language is provided in Table III.

C. Procedures

Speech samples were collected using an auditory-visual word repetition task. The auditory prompts were recorded by a native female speaker of each language, from the same dialect region as participants, in a child-directed speech register. The pictures were culturally appropriate for each language and

TABLE III. Consonant environment for vowels in each language.

	/a/	/i/	/u/
Cantonese	[t, t ^h , s, ts, ts ^h , k, k ^h , k ^w , k ^{wh}]	[t, t ^h , s, ts, ts ^h , k, k ^h]	[k, k ^h]
English	[d, t ^h , s, f, tʃ, g, k ^h]	[d, θ, t ^h , t ^w , s, f, tʃ, g, k ^h]	[d, t ^h , s, f, tʃ, k ^h]
Greek	[t, s, ts, k, k ^l , x]	[θ, t, s, ts, ç, k ^l]	[t, s, ts, k]
Japanese	[d, t, s, f, tʃ, g, k, k ^j]	[t, f, tʃ, g ^j , k ^j]	[s, f, ts, tʃ, dz, dʒ, g, k, k ^j]
Korean ^a	[t, t ^h , t', s, s', tʃ, tʃ ^h , k, k ^h , k']	[t, t ^h , t', f, f', tʃ, tʃ ^h , tʃ', k, k']	[t, t ^h , t', s, s', tʃ, tʃ ^h , tʃ', k, k ^h , k']

^aKorean stops have three-way laryngeal contrast /t/ represents a lax stop, /t^h/ represents an aspirated stop, and /t'/ represents a tense stop.

country. Productions of these single words were presented to children and adults by playing an auditory recording of each item along with a color picture on the 14 in. screen of a laptop computer. The participants were asked to repeat each target word after the prompt was played to them. The responses were digitally recorded using a unidirectional tabletop microphone coupled to a Marantz PMD 660 flashcard recorder (Marantz, Kew Gardens, NY) at a sampling rate of 44 100 Hz.

A trained native phonetician of each language transcribed each token by listening to and examining the acoustic waveform of each repetition. Each target vowel was transcribed as correct or incorrect. For English and Cantonese, vowels that are difficult to discriminate in children's speech (even by trained phoneticians) and that have similar coarticulatory effects were collapsed together and treated as vowels of the same vowel category for coding accuracy. In English, both /i/ and /ɪ/ vowels were collapsed into the /i/ category, /a/, /ʌ/, and /ɔ/ into the /a/ category, and /u/ and /ʊ/ into the /u/ category. Similarly, in Cantonese, /a/ or /ɛ/ were collapsed into the /a/ category. That is, if a target /i/ was heard as /ɪ/ in English, it would still be coded as correct. All adult productions were assumed to be correct and were included in the analysis without a native phonetician's accuracy judgment. Table IV gives the accuracy rate of each vowel for children from all five languages. Only those vowels judged as correctly produced were used for the subsequent acoustic analysis, which resulted in an unbalanced number of vowels across language and age groups. The number of tokens for each vowel in each language and age group is provided in Table V. It should be noted that the accuracy rates for English and Cantonese may be somewhat inflated, given the above-described accuracy-coding procedure. Even so, the accuracy rate for vowels produced by 2-year-old Cantonese speakers is the lowest of all of the languages.

D. Data analysis

1. Vowel formant measurement

The first two formant frequency values ($F1$ and $F2$) were extracted from the vowel midpoint using the LPC solution in PRAAT (version 5.0.29) (Boersma and Weenink, 2006). The size of the analysis window was 25 ms and the dynamic range was set at 30 dB. In order to determine the

TABLE IV. Vowel accuracy rates (%) by age and language.

Age group	Language	Vowel categories		
		/a/	/i/	/u/
2-year-olds	Cantonese	77	85	56
	English	92	88	83
	Greek	87	96	84
	Japanese	94	83	85
	Korean	91	82	73
5-year-olds	Cantonese	98	100	97
	English	96	99.5	95
	Greek	95	100	99
	Japanese	99.6	94	96
	Korean	99.7	97	94

TABLE V. Number of observations of each language and age groups.

		Cantonese	English	Greek	Japanese	Korean
/a/	2-year-olds	56	130	92	197	269
	5-year-olds	91	137	92	212	301
	Adults	91	137	90	121	300
/i/	2-year-olds	129	64	100	102	146
	5-year-olds	191	81	93	116	173
	Adults	190	92	90	55	181
/u/	2-year-olds	16	114	71	181	174
	5-year-olds	30	127	80	189	236
	Adults	30	144	88	104	245

midpoint location at which to measure the formant frequencies, vowel onset and offset were marked manually. Vowel onset was defined as the time at which the first clear glottal pulse was observed in both the waveform and the spectrographic display. Vowel offset was defined as the final glottal pulse extending at least through $F1$ and $F2$. Once vowel onset and offset were defined, the vowel duration was calculated automatically by PRAAT and the midpoint was determined. $F1$ and $F2$ values were measured at vowel midpoint, at which we assumed there would be minimal formant shifts. This measurement approach is well known and has been used in a nearly identical manner in other studies of vowel formant frequency values (e.g., Hillenbrand *et al.*, 1995; Peterson and Barney, 1952). Tokens were excluded from the analysis if the vowel was devoiced, too low in intensity for accurate formant measurements, or overlaid with noise. Formant mistrackings were identified by eye and corrected manually.

2. Vowel formant transformation: Radius and degree

To eliminate the effect of different vocal tract sizes of speakers on the formant frequency values of vowels, each of the three point vowels was described in terms of relative distance and angular displacement from a speaker-specific centroid. The procedures were as follows. First, the raw $F1$ and $F2$ values were log-transformed as done in other normalization schemes discussed in the literature (Adank *et al.*, 2004). Second, for each participant, the mean log $F1$ and log $F2$ values of each of the three vowels /a/, /i/, and /u/ were calculated, and then the mean of the mean log $F1$ and log $F2$ of these three vowels were calculated to obtain a speaker-specific centroid. For Cantonese /u/, there was only a single consonant–vowel sequence (/ku/), which resulted in only 2–3 productions per speaker. Except for Cantonese /u/, there were 7–30 productions per speaker for each vowel. A weighted mean by number of tokens for each vowel category was used to calculate the speaker-specific centroid to account for the different number of tokens available for each speaker. If a subject did not produce all three vowels, he/she was excluded from the analysis since the calculation of a speaker centroid depended on the number of vowel categories contributing to the mean values. For this reason, two Cantonese 2-year-olds were excluded from the analysis because they lacked any correct /u/ productions. Third, the

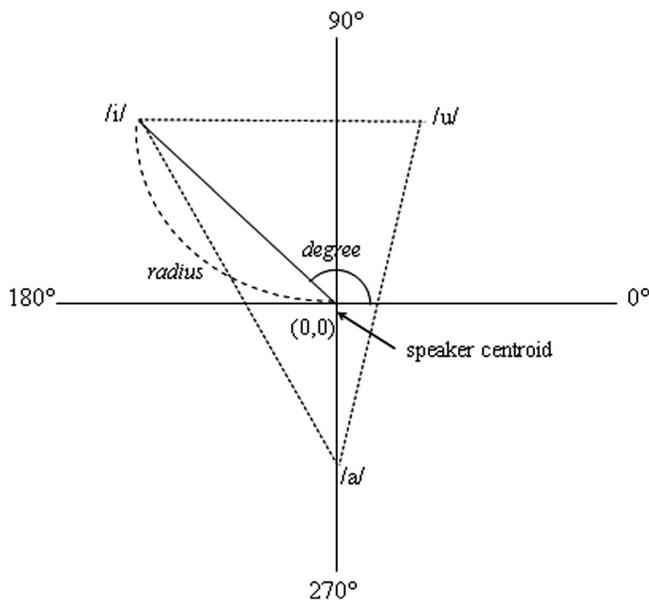


FIG. 1. A Cartesian coordinate system showing how the radius and degree values for /i/ were calculated.

polar coordinate system ($\log F1$ by $\log F2$) was transformed to a Cartesian coordinate system so that the speaker-specific centroid was the origin, (0,0). The transformations were made by subtracting each individual's mean $F1$ and $F2$ values from \log -transformed $F1$ and $F2$ values of each token. Each data point was then expressed with the Cartesian system relative to this speaker-specific centroid. Finally, the location of each vowel, relative to the speaker-specific

centroid, was calculated with respect to (1) *radius* (the distance from the speaker-specific centroid) and (2) *degree* (the angular displacement from the x axis of this system) (see Fig. 1). A speaker-specific centroid can be interpreted as formant frequency values that reflect the neutral vowel configuration or a tube with uniform cross sectional area from glottis to lips. The radius and degree values represent the size and the rotation of the vowel space within this Cartesian coordinate system. Longer radii imply more peripheral locations of vowels relative to a speaker-specific centroid and different degree values imply rotational differences of each vowel, either front or back, or high or low in the vowel acoustic space.

3. Statistical analysis: Radius and degree

For radius values, a linear mixed-effects regression model was used to accommodate the repeated measures in the data structure. The analyses were done using R (R Development Core Team, 2009) and its implemented packages LATTICE (Sarkar, 2009) and LME4 (Bates and Maechler, 2009). The effects of three fixed factors (language, age, and vowel) and one random factor (speakers) on the radius were analyzed. Interactions between factors were also examined. In the present analysis, only two-way interactions were of interest (three-way interactions were not of interest as we expected there to be an effect of vowel).

For the analysis of degree, a standard linear regression analysis was used to account for the periodic property of degree values. A problem with the periodic property of degree

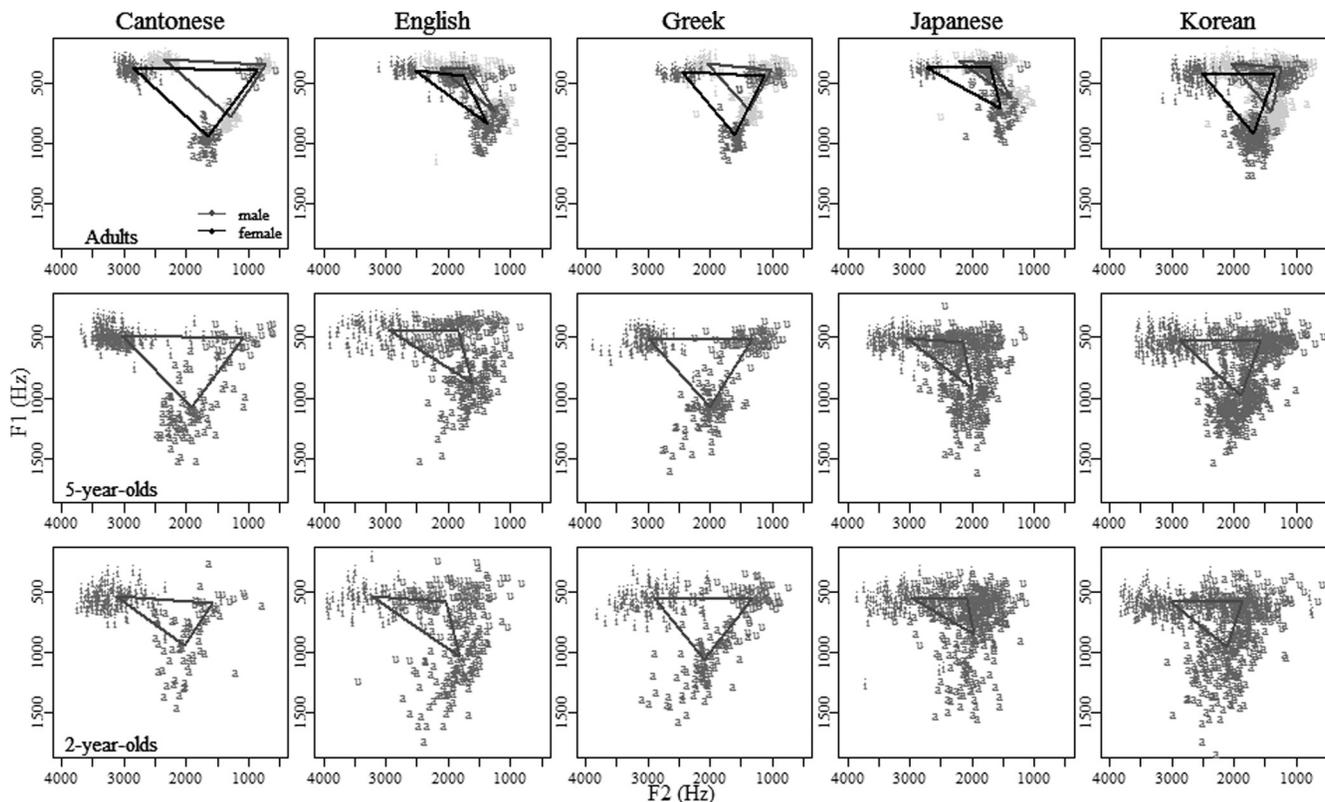


FIG. 2. Scatter plots of $F1$ and $F2$ by language and age for both males (grey) and females (black). Each point represents the production of a single vowel by a single speaker. The vowel space was drawn by connecting the mean $F1$ - $F2$ value of all productions of each vowel for each language and age group.

values is that the absolute value of the difference between 1° and 359° is just 2° , but numerically, the difference between 1° and 359° is 358° . To circumvent this problem, an average degree for each vowel for each speaker was calculated so that all degree values lie on a semi-circle. Because there was only a mean observation for each vowel for each subject, the repeated measure was not necessary. The weighted least-square method was used instead to adjust different number of degree values for each speaker, by giving different weights to each speaker based on the number of observations averaged. In the regression model, the independent variables were language and age. The analyses were done separately for each vowel category.

III. RESULTS

A. Traditional vowel space with F1–F2 values

Scatter plots of $F1$ vs $F2$ by language and age are shown in Fig. 2. The points of the triangles are group means for the $F1$ – $F2$ coordinates of the three corner vowels. For adults, notable cross-linguistic differences observed in Fig. 2 included a large vowel space with the three vowels clearly and distinctly separated in Cantonese and relatively smaller vowel spaces with closely distributed /i/ and /u/ vowels in English and Japanese. Cantonese had an almost-equilateral triangular shape for its vowel space. On the other hand, for English and Japanese, /u/ was closer to /i/ in the $F2$ dimension and /a/ was closer to /u/ in the $F1$ dimension. The English and Japanese vowel spaces, therefore, differed from the Cantonese vowel space in terms of both size and shape.

The identifiable language-specific characteristics were also observed even in vowel productions of young children. The 5-year-olds showed language specificity in their vowel acoustic spaces, which looked remarkably similar to those of adults. More variability was observed in the production patterns of 2-year-olds than in those of 5-year-olds and adults. Nevertheless, the relatively clear separation of the three vowels in Cantonese and the higher $F2$ values of /u/ in English and Japanese were still observable in the vowel production patterns of the 2-year-olds. Overall, the shapes of the vowel space were also similar across the three age groups for all languages. However, the vowel spaces of 2-year-olds were consistently smaller in size in comparison with those of older age groups. The average first three formant frequencies of vowels produced by 2-year-olds, 5-year-olds, and adults of Cantonese, English, Greek, Japanese, or Korean are summarized in Appendix B.

B. Transformed vowel space

Figure 3 shows the transformed vowel spaces of each language and age group following the computational process outlined in Sec. II. This transformation allows the comparison of size and shape of vowel spaces across speakers without having the confounding effect coming from different vocal tract sizes of speakers. When compared to Fig. 2, two cross-linguistic differences that were not found with raw $F1$ and $F2$ values were observed in the transformed data presented in Fig. 3. First, the relative location of /i/ and /u/ in adult English and Japanese productions were different in the

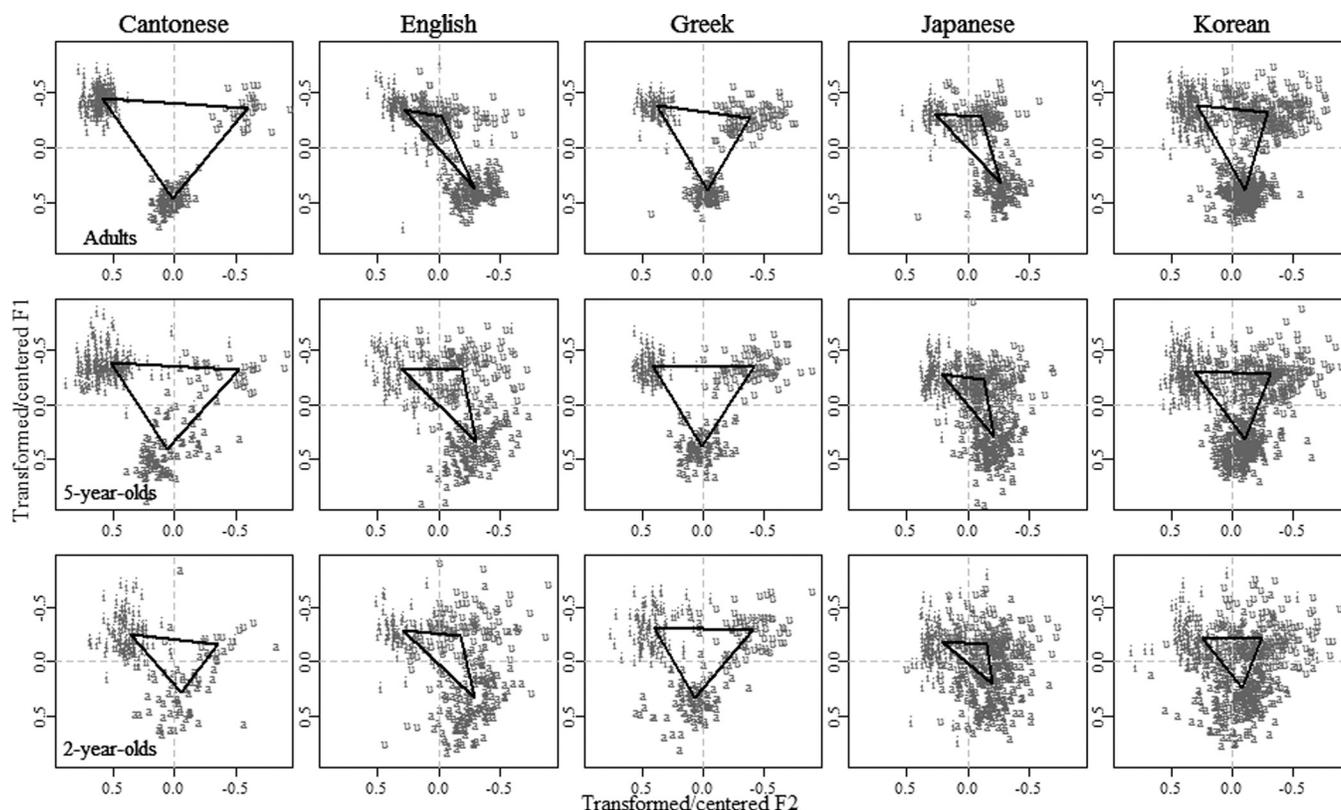


FIG. 3. Scatter plots of transformed $F1$ – $F2$ by language and age. The transformations were made by subtracting each individual's mean $F1$ and $F2$ values from log-transformed $F1$ and $F2$ values of each token. The measurements were centered at the origin (0, 0), the speaker centroid. The vowel space was drawn by connecting the mean log-transformed $F1$ – $F2$ value of all productions of each vowel for each language and age group.

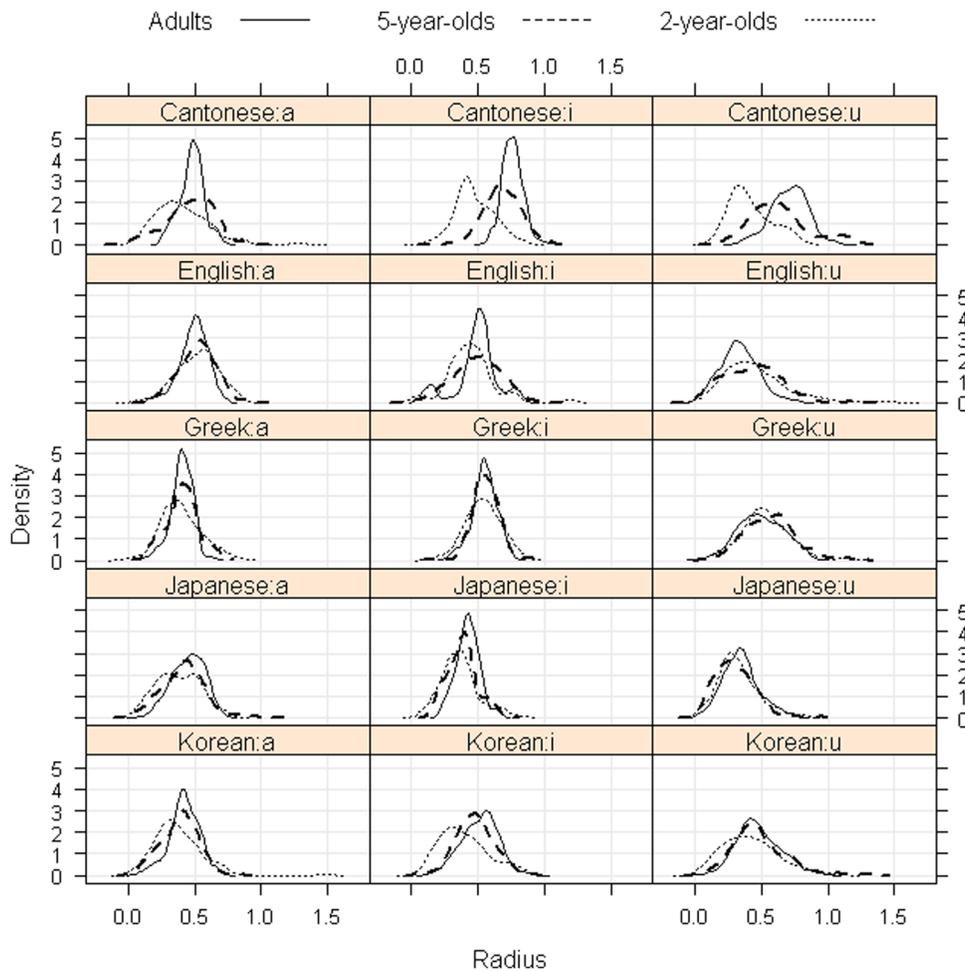


FIG. 4. (Color online) Density plots of radius values sorted by language, age, and vowel. The x axis shows radius values, and the y axis represents the density, which is the relative frequency of occurrence for different radius values. Solid lines show data for adults; dotted lines show data for 5-year-olds; and small dotted lines show data for 2-year-olds.

raw vs transformed data in that the relative distance between /i/ and /u/ appeared to be closer in the transformed data. This resulted in differently shaped vowel spaces between Fig. 3 and Fig. 2. Second, the transformed vowel spaces of 2-year-olds became clearly smaller in size relative to those of 5-year-olds and adults, which indicates that vowel spaces become larger with age.

1. Radius (distance from a speaker-specific centroid)

Figure 4 presents the smoothed distribution of radii for the three point vowels of Cantonese, American English, Greek, Japanese, and Korean for each age group, where the peak and spread of the curves suggest the mode and variability of radii for the vowels. As Fig. 4 shows, for adults' productions (solid lines), the radii for /i/ and /u/ produced by Cantonese adults were longer than for the same vowels in the four other languages (see the modes of the second and third columns of Fig. 4). This difference was greatest between Cantonese vs English and Japanese adults' productions. The result from the linear mixed-effects model confirmed this observation, where a significant effect of language ($F = 47.71$, $df = 4$, $SumSq = 3.88$) was found. The estimated mean radius values of /i/ and /u/ showed that those of Cantonese /i/ and /u/ were the greatest, followed by

Greek, Korean, English, or Japanese (in descending order), suggesting that /i/ and /u/ vowels were the most distant from the speaker-specific centroid in Cantonese. The mean radii for each vowel estimated from the mixed-effects regression model was summarized in Table VI. For /a/ produced by adult speakers, however, no substantial cross-linguistic difference in radii values was observed.

Across age groups within each language, the radius values of younger children were relatively shorter than those of older age groups, especially for Cantonese /i/ and /u/. However, this was not the case in English, in which radius values of vowels produced by adults were shorter than those produced by 2- and 5-year-old children. The result from the

TABLE VI. Mean radii for each vowel for each language and age group estimated from mixed-effects regression models.

	/a/			/i/			/u/		
	2 yr	5 yr	Adults	2 yr	5 yr	Adults	2 yr	5 yr	Adults
Cantonese	0.400	0.466	0.477	0.478	0.682	0.759	0.425	0.640	0.710
English	0.515	0.520	0.494	0.437	0.495	0.490	0.454	0.424	0.337
Greek	0.398	0.404	0.403	0.536	0.559	0.550	0.524	0.565	0.501
Japanese	0.374	0.412	0.452	0.362	0.387	0.422	0.324	0.301	0.347
Korean	0.366	0.386	0.421	0.393	0.477	0.516	0.417	0.478	0.483

linear mixed-effects model also confirmed these observations, where a significant effect of age was found ($F = 11.42$, $df = 2$, $SumSq = 0.46$). In addition, the radius values of the 2-year-olds had a more diffuse distribution of radii as compared to those of the 5-year-olds and the adults, suggesting that the vowels produced by 2-year olds had the most variable distance from the speaker-specific centroid.

A significant interaction between age and language was also found ($F = 6.42$, $df = 8$, $SumSq = 1.04$). The program that was used for the linear mixed-effects analysis (LME4, Bates and Maechler, 2009) does not provide a detailed *post hoc* analysis. Therefore, differences between group means were used to interpret the interaction (see Table VI). There was a larger difference in radius values across the three age groups for Cantonese /i/ and /u/, relative to radius values of the other languages. The radius values of /i/ and /u/ produced by Cantonese 2-year-olds were substantially shorter than those of Cantonese adults (for /i/ —radius for 2-year-olds = 0.48 vs adults = 0.76; for /u/ —radius for 2-year-olds = 0.43 vs adults = 0.71), while these large differences in radius values between 2-year-olds and adults were not observed in the other four languages. This observation suggests that the /i/ and /u/ productions of Cantonese-speaking 2-year-olds are less adult-like than those of 2-year-olds in the other four languages. This may be related to the cross-linguistic patterns observed in adult productions: The longest radius values for /i/ and /u/ across languages were for Cantonese productions, so this production pattern may be particularly difficult for 2-year-olds to master.

2. Degree (angular displacement around the speaker-specific centroid)

Figures 5–7 show the angular displacements of each of the three vowels calculated in relation to the corresponding speaker centroid by vowel types (/a/, /i/, and /u/) for five languages and three age groups. Each point, placed on an arc, represents a vowel produced by each speaker. The center of each circle is the origin, (0,0), of the Cartesian x - y coordinate system. The degree values increase beginning from the top right part of the circle (the first quadrant) to the bottom right part of the circle (the fourth quadrant). The greater the spread of points around the arc, the greater the dispersion of vowels within a category. The mean degree values of each age and language group are summarized in Table VII.

As Fig. 5 shows, the distribution of degree values of /a/ is different across languages. The degree values of /a/ produced by Cantonese, Greek, and Korean adults were distributed between the third and the fourth quadrant, whereas those of English and Japanese adults were mostly in the fourth quadrant. The result from the linear regression analysis is consistent with this observation, where a main effect of language on degree values [$F(4, 133) = 36.18$, $p < 0.001$] was found. The Tukey's HSD *post hoc* test showed that the degree values of English and Japanese /a/ were significantly greater than those of Cantonese, Greek, and Korean. This implies that /a/ produced by English and Japanese adults occupied a more back location in the vowel

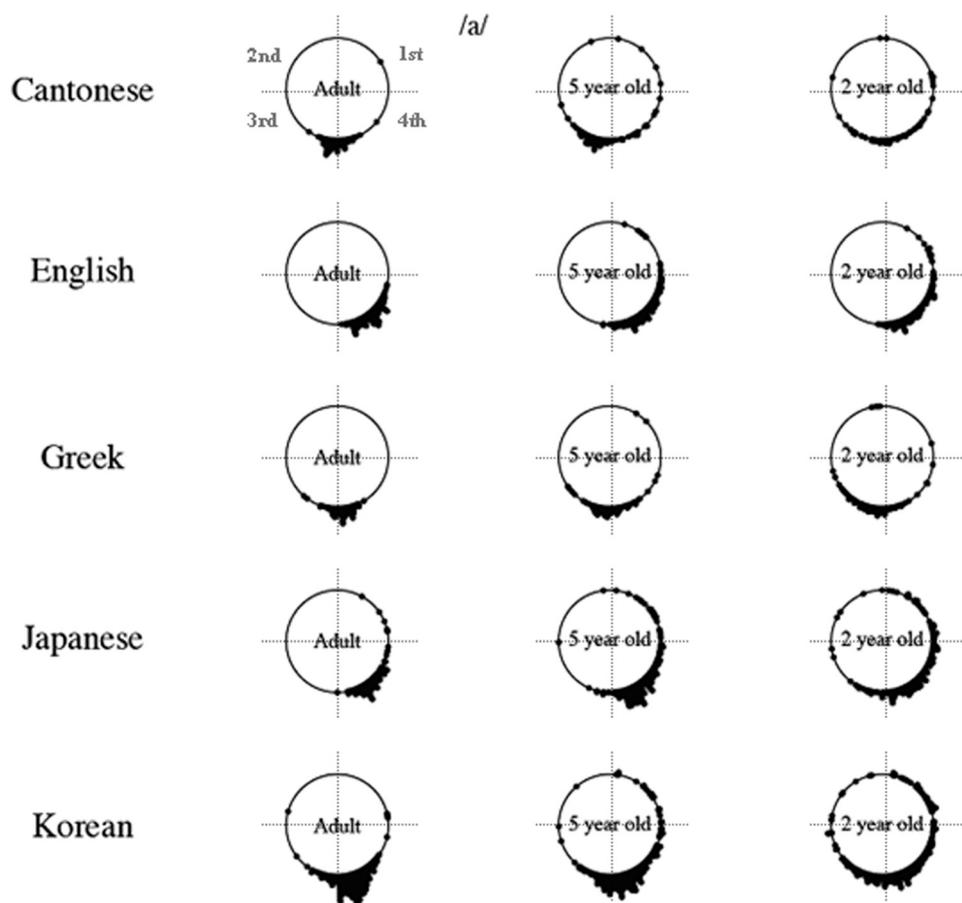


FIG. 5. Circular plots for /a/. Each dot represents degree values of each vowel. The degree values were calculated and placed on the arc of a circle from the origin (0,0).

TABLE VII. Mean degree values for each language and age group.

		Cantonese	English	Greek	Japanese	Korean
/a/	2-year-olds	245.58	278.71	251.42	253.16	242.15
	5-year-olds	247.73	288.49	264.16	275.08	267.60
	Adults	265.70	307.32	274.87	296.63	280.46
/i/	2-year-olds	148.28	139.78	143.63	142.70	139.90
	5-year-olds	141.51	136.03	140.40	131.35	136.71
	Adults	142.69	136.45	134.62	130.70	127.58
/u/	2-year-olds	47.99	88.05	49.82	95.06	76.23
	5-year-olds	34.94	75.04	44.90	76.55	55.92
	Adults	31.22	93.24	39.24	75.89	58.04

acoustic space compared to /a/ produced by Cantonese, Greek, and Korean adults. This is consistent with previous findings, in which American English /a/ was transcribed with a different IPA (International Phonetic Alphabet) symbol ([ɑ]) (Hillenbrand *et al.*, 1995; Ladefoged and Maddieson, 1996), which represents a more back position of the tongue as compared to /a/. The low back position of Japanese /a/ has also been reported in previous studies (e.g., Keating and Huffman, 1984). No clear cross-linguistic difference was observed in the degree distribution pattern of /i/. As shown in Fig. 6, /i/ productions of adults of all languages were clustered mainly in the second quadrant. Although a main effect of language on degree values was found [$F(4, 133) = 6.54, p < 0.001$], the Tukey's HSD *post hoc* analysis confirmed the pattern

shown in Fig. 6; no pair of languages showed a significant difference in their degree values for /i/. As can be seen in Fig. 7, the degree values of /u/ differed considerably across languages. The degree values of /u/ produced by English and Japanese adults were widely spread along the top right (first quadrant) and left parts (second quadrant) of the arc, while those produced by Cantonese, Greek, and Korean adults were clustered in the top right part of the arc (mostly in the first quadrant). The results from the linear regression and Tukey's HSD *post hoc* analyses showed that there was a main effect of language on degree values [$F(4, 133) = 13.59, p < 0.001$] and that the degree values of /u/ produced by English-speaking adults were significantly greater than those produced by speakers of the other four languages. The degree values of Japanese /u/ were also found to be significantly greater than those produced by speakers of Cantonese and Greek, but significantly smaller than those produced by speakers of English, resulting in the order of English, Japanese, Korean, Greek, and Cantonese (from the largest degree values to the smallest). This result suggests that adult productions of English /u/ occupy the most front location in the vowel acoustic space, while Cantonese /u/ occupies the most back location.

The rotational differences of vowels produced by children were found to be significantly different from those of adults' productions for all three point vowels [/a/: $F(2, 133) = 12.75, p < 0.001$; /i/: $F(2, 133) = 4.73, p = 0.010$; /u/: $F(2, 133) = 20.35, p < 0.001$]. The result from the Tukey's HSD *post hoc* test showed that degree values of

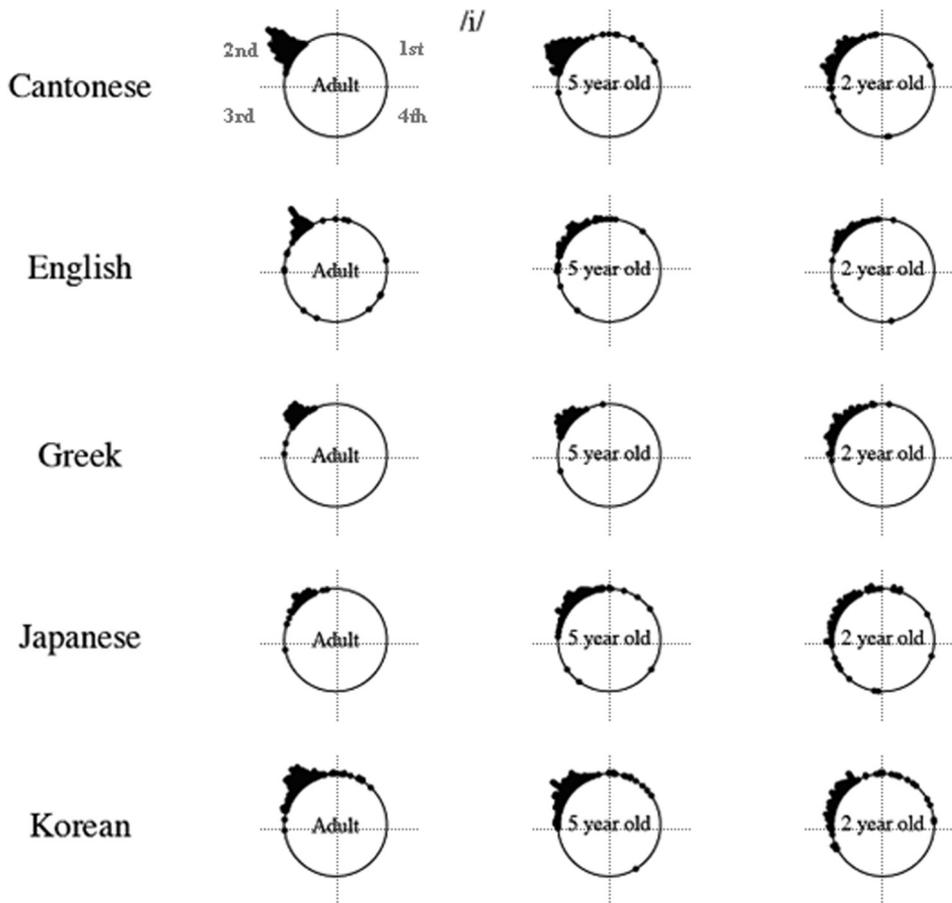


FIG. 6. Circular plots for /i/.

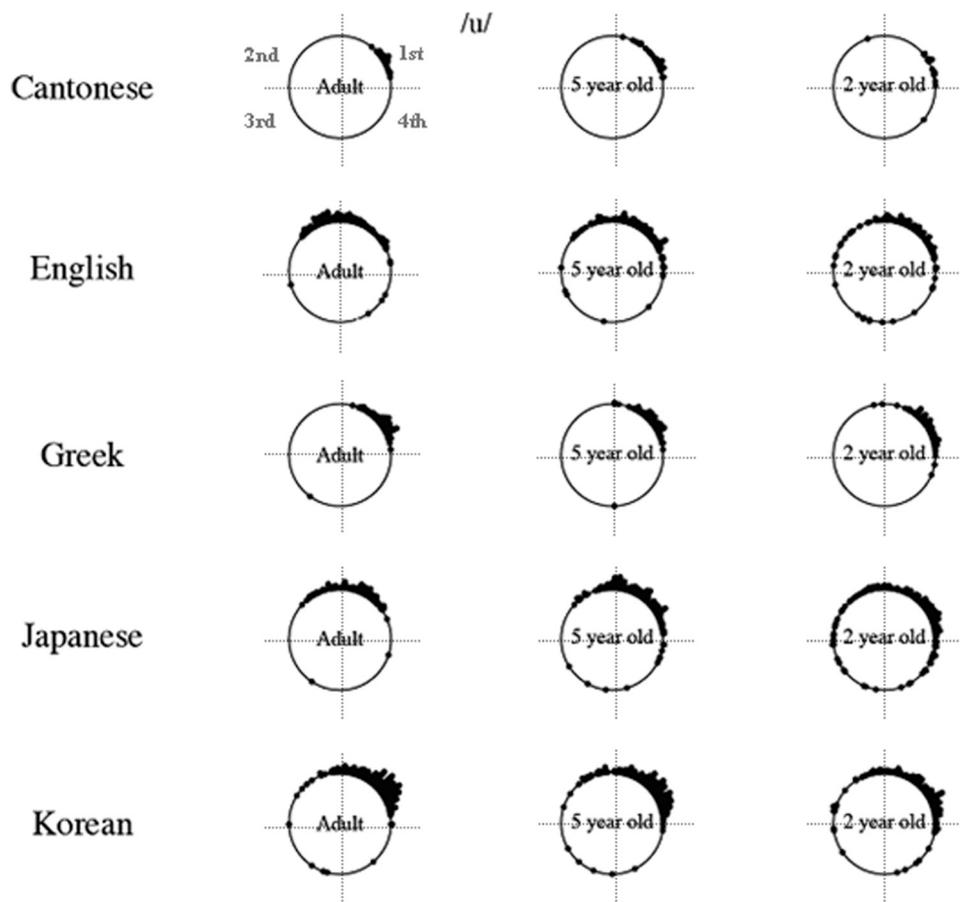


FIG. 7. Circular plots for /u/.

vowels produced by 2-year-olds were consistently smaller than those produced by older age groups for /a/ and /u/. Although smaller degree values may imply more central articulatory locations for /a/ and /u/ productions among 2-year-olds, this interpretation should be made cautiously because variability decreased with age. That is, the smaller degree values for /a/ and /u/ productions by 2-year-olds may be due to more centralized articulation or it may simply be related to the greater variability of their productions.

For /i/, Tukey's HSD *post hoc* analysis showed no significant age group difference. That is, 2-year-olds' productions of /i/ showed a very similar pattern as those of adults, both in terms of the extent of the dispersion and the location of each vowel along the circle. More dispersed distribution of degree values were found in vowels produced by 2-year-olds than in those produced by 5-year-olds and adults.

No language by age interaction was found for either /a/ or /i/, indicating similar vowel developmental patterns across languages. For /u/, however, a significant language by age interaction was found [$F(2, 133) = 20.35, p < 0.001$]. The Tukey's HSD *post hoc* analysis showed that in English, the degree values produced by adults were significantly larger than those of 5- and 2-year-olds, while for Japanese, degree values of /u/ produced by adults were significantly larger than those of 2-year-olds, but were not significantly different from those of 5-year-olds. An effect of age group was not observed in the degree values of /u/ produced by Cantonese,

Greek, and Korean speakers. This result suggests that Cantonese, Greek, and Korean children produced /u/ with a similar rotation as adults by the age of 2. By contrast, English and Japanese 2-year-olds are still in the process of refining the phonetic characteristics of /u/ in order to achieve an adult-like production. This is especially true for English, as even the productions of 5-year-olds have not achieved an adult-like rotational location for /u/.

IV. DISCUSSION

The purpose of this study was to understand cross-linguistic differences in the acoustic realization of three "shared" vowels (/a/, /i/, and /u/) across five languages, and to determine whether children as young as 2 years of age produce the language-specific fine phonetic detail observed in adult productions. To eliminate the confounding effects of different vocal tract sizes on formant frequency values, we compared the relative location of each of the three point vowels in a speaker-specific vowel space by calculating distance and rotation of these vowels.

Vowel spaces formed by the three point vowels of each language were systematically different from each other both in terms of the size and shape of their vowel spaces. Consistent with previous cross-linguistic studies that showed language-specific characteristics in shared vowels (e.g., Bradlow, 1995; Yang, 1996), the vowel space formed

by the three point vowels produced by Cantonese adults was larger in size than those of other languages. Also, Cantonese /i/ and /u/ vowels were more peripherally located in the vowel acoustic space, while those of English and Japanese were located closer to a speaker-specific centroid (more centralized). In terms of rotational differences, cross-linguistic differences were mostly due to a more back location of /a/ and a more fronted location of /u/ produced by English and Japanese speakers relative to those of the three other languages. The fronted /u/ of English and Japanese has been reported in previous studies as a current sound change occurring both in American and British English (Clopper and Pisoni, 2006; Hagiwara, 1995; Harrington *et al.*, 2008; Hawkins and Midgley, 2005; Hillenbrand *et al.*, 1995) and as a phonemic characteristic with higher $F2$ values (e.g., Keating and Huffman, 1984), respectively. The Cantonese /u/ had the most back location in the vowel acoustic space among the five languages studied. The relatively extreme back location of Cantonese /u/, however, may be due, at least in part, to the fact that only a /k- /u/ sequence is allowed in Cantonese, prohibiting alveolar consonants before /u/. Therefore, the coarticulation with the velar stop might have resulted in a more back location for /u/ in Cantonese. Unlike the /u/ and /a/ vowels, /i/ showed the most stable production pattern across languages.

These language-specific characteristics were also observed in children's productions. Vowel productions of 5-year-olds were similar to those of adults in that they showed language-specific patterns in their vowel spaces. These language-specific characteristics were also found in 2-year-olds' productions, though adult-like patterns were less prominent and somewhat less consistent. Moreover, vowel productions of 2-year-olds were more variable than those of older age groups. The variability observed for the 2-year-olds' productions suggests that they are still in the process of learning to produce adult-like vowels, even for productions that were transcribed as correct.

It is also possible, however, that the lower accuracy scores and greater variability observed for the 2-year-olds' productions are related, at least in part, to their smaller vocabularies. While the target words used in this study were chosen to be familiar to young children, not all of these words were likely to be in the expressive vocabulary of all 2-year-olds, although they were likely to be in the expressive vocabulary of all 5-year-olds. It is also possible that some of the differences across vowels could be related to differences in phonotactic probability of each vowel type. For example, as discussed earlier, /u/ is low in frequency in Cantonese because of a phonotactic restriction that allows it to occur only before velar consonants. Two additional analyses were done to explore whether word familiarity or phonotactic probability influenced transcribed accuracy or acoustic variability for the 2-year-olds' productions. First, $F1$ and $F2$ values for stimulus words that were likely to be in English-speaking 2-year-olds' expressive vocabulary (because they were included in the MacArthur-Bates Communicative Development Inventories (CDIs) (Fenson *et al.*, 2007) were compared to words that were not on the CDI. Qualitatively, there was no difference in

variability of $F1$ and $F2$ patterns between these two sets of words. Thus, it is unlikely that the greater variability of $F1$ and $F2$ values for the 2-year-olds was entirely due to the fact that some of these words were not in their expressive vocabulary. We also calculated the average phonotactic probability for each vowel, based on the consonant-vowel sequences that were used for the stimulus items for each vowel in each language (see Appendix C for results and description of procedures). Across languages, phonotactic probability was generally lower for /u/ than for the other two vowels, but again, phonotactic probability was not consistently related to accuracy across languages. Thus, it seems unlikely that differences in phonotactic probability of these consonant-vowel sequences across languages can explain the results for the 2-year-olds.

Across languages, /u/ was the vowel that had the most differences in radius and degree values across age groups. While the radius and degree values for /i/ were stable across age groups in all five languages, these values for /u/ differed between adults and 2-year-olds for English and Japanese. This result suggests that learning how to produce /u/ in English and Japanese is more demanding than learning to produce the same vowel in Cantonese, Greek, and Korean. A possible explanation for this may be that the input for /u/ may be variable, at least in English. Fronting of /u/ has been described as a sound change in progress in both American and British English (Clopper and Pisoni, 2006; Hagiwara, 1995; Harrington *et al.*, 2008; Hawkins and Midgley, 2005; Hillenbrand *et al.*, 1995). Thus, children may be getting more variable input for /u/, with older speakers producing a more backed /u/ and younger speakers producing a more fronted /u/. This variable input for /u/ could make it harder for children to form a vowel category and refine its phonetic characteristics.

To conclude, this study showed that the adult vowel spaces formed by the three point vowels systematically differed from each other in terms of both size and shape. In comparing vowel spaces of children to those of adults, we found that the vowel spaces of the 5-year-olds were mostly similar to those of adults of the same language, while those of 2-year-olds were similar in shape, but were smaller in size than those of adults. This result suggests that children gradually refine their vowel categories even after the adult-like categories have been "roughed out" at an early age. This gradual phonetic refinement process differs by language based on the acoustic characteristic of each vowel.

ACKNOWLEDGMENTS

This work was supported by a Fulbright fellowship to H.C. and by NIDCD Grant No. 02932 and NSF Grant No. 0729140 to J.E. We thank the children and adults who participated in this study. We also thank Mary Beckman for her advice on the analyses, Junko Davis, Sarah Schellinger, Asimina Syrika, and Peggy Wong for their work on data collection and analysis, and Amy Andrzejewski, Stelios Fourakis, and Ann Todd for their comments on previous versions of this paper.

APPENDIX A: STIMULI LIST FOR ENGLISH^a

Target	/a/	/i/	/u/
/t/	taco	teacher	tooth
	tall (/ɔ/)	tepee	tuna
	tongue (/ʌ/)	<i>tickle</i>	tube
/d/	donkey	deer	duke
	dove	<i>digging</i>	duel
	dolphin	<i>ditch</i>	duel
/s/	soccer	seat	soup
	sauce	seal	super
	sock	seashore	suitcase
/ʃ/	shop	shield	shoe
	sauce (/ɔ/)	sheep	chute
	shark	<i>ship</i>	<i>sugar</i>
/tʃ/	chart	<i>chimney</i>	choo-choo
	chalkboard (/ɔ/)	<i>chipmunk</i>	choose
	chop	<i>chips</i>	chewing
/g/	garden	<i>giggles</i>	
	golf (/ɔ/)	<i>giving</i>	—
	gumdrop (/ʌ/)	<i>gift</i>	
/θ/	—	<i>thin</i>	—
		<i>thinking</i>	
		<i>thimble</i>	
/k/	car	key	cougar
	color (/ʌ/)	<i>kitchen</i>	<i>cooking</i>
	cut (/ʌ/)	<i>kicking</i>	<i>cookie</i>

^aThe lax vowels /ɪ/ and /ʊ/ were not included in the acoustic analysis. Only the tense vowels /i/ and /u/ were included in the analysis. Italicized are words with lax vowels.

APPENDIX B: AVERAGE FORMANT FREQUENCIES OF CANTONESE, ENGLISH, GREEK, JAPANESE, AND KOREAN VOWELS OF 2- AND 5-YEAR-OLDS AND ADULTS

		/a/			/i/			/u/		
		F1	F2	F3	F1	F2	F3	F1	F2	F3
Adult males	Cantonese	779	1296	2713	310	2368	3124	353	721	2728
	English	730	1250	2512	370	2146	2838	374	1668	2486
	Greek	720	1372	2457	347	2058	2642	401	1011	2393
	Japanese	634	1363	2857	321	2222	3206	345	1704	2848
	Korean	734	1390	2467	343	2046	2781	379	1241	2446
Adult females	Cantonese	942	1641	3125	375	2851	3467	397	872	2886
	English	843	1363	2634	391	2570	3128	442	1755	2699
	Greek	929	1603	2825	406	2456	3029	443	1126	2801
	Japanese	708	1533	2903	375	2729	3447	380	1714	2841
	Korean	922	1686	2976	419	2528	3277	434	1358	2804
5-year-olds	Cantonese	1076	1917	3459	481	3046	3917	512	1094	3478
	English	876	1613	3319	439	2964	3739	454	1834	3425
	Greek	1076	2007	3786	513	3008	3902	521	1331	3698
	Japanese	918	1982	3800	509	3044	4049	540	2157	3806
	Korean	977	1890	3481	522	2866	3805	533	1571	3522
2-year-olds	Cantonese	895	1944	3727	541	3127	4368	577	1520	3880
	English	1038	1822	3995	544	3220	4210	566	2016	3841
	Greek	1053	2095	3947	546	2907	3923	561	1320	3678
	Japanese	838	1971	3957	553	2966	4059	566	2074	3879
	Korean	949	2103	3989	588	2975	4166	583	1846	3849

APPENDIX C: AVERAGE CV BIPHONE FREQUENCIES FOR EACH VOWEL CONTEXT FOR EACH LANGUAGE

These CV phonotactic probabilities were calculated based on an adult online lexicon for each language the *Segmentation Corpus* (Chan and Tang, 1999; Wong et al., 2002) for Cantonese, the *Hoosier Mental Lexicon* (HML, Nusbaum et al., 1984) for English, the Institute for Language and Speech Processing (ILSP) database (Gavrilidou et al., 1999) for Greek, the Nippon Telegraph and Telephone (NTT) database (Amano and Kondo, 1999) for Japanese, and NIKL database (National Institute of the Korean Language, 2000) for Korean]. For each vowel in each language, the number of times each target CV sequence of the stimulus items occurred in the online lexicon was added up and this number was divided by the total number of stimulus items. Then this number was divided by the total number of words in the online lexicon of the language and took the natural log of this quotient to get the phonotactic probability.

	Cantonese	English	Greek	Japanese	Korean
/a/	-1.171	-1.988	-1.658	-1.548	-1.601
/i/	-1.414	-3.059	-1.628	-1.724	-1.960
/u/	-2.426	-2.750	-2.644	-1.771	-1.936

Adank, P., Smits, R., and van Hout, R. (2004). "A comparison of vowel normalization procedures for language variation research," *J. Acoust. Soc. Am.* **116**, 3099–3107.

Amano, S., and Kondo, T. (1999). *Nihongo-no goitokusei (Lexical Properties of Japanese)* (Sanseido, Tokyo).

Bates, D., and Maechler, M. (2009). "Lme4: Linear mixed-effects models using S4 classes," R package version 0.999375-32.

Boersma, P., and Weenink, D. (2006). "Praat: Doing phonetics by computer [Computer program]," <http://www.praat.org> (Last viewed September 20, 2010).

Bradlow, A. R. (1995). "A comparative acoustic study of English and Spanish vowels," *J. Acoust. Soc. Am.* **97**, 1916–1924.

Chan, S. D., and Tang, Z. X. (1999). "Quantitative analysis of lexical distribution in different Chinese communities in the 1990's," *Yuyan Wenzhi Yingyong (Applied Linguistics)* **3**, 10–18.

Clopper, C., and Pisoni, D. B. (2006). "The Nationwide Speech Project: A new corpus of American English dialects," *Speech Commun.* **48**, 633–644.

Davis, B. L., and MacNeilage, P. F. (1990). "Acquisition of correct vowel production: A quantitative case study," *J. Speech Hear. Res.* **33**, 16–27.

de Boysson-Bardies, B., Halle, P., Sagart, L., and Durand, C. (1989). "A cross-linguistic investigation of vowel formants in babbling," *J. Child Lang.* **16**, 1–17.

Disner, S. F. (1980). "Evaluation of vowel normalization procedures," *J. Acoust. Soc. Am.* **67**, 253–261.

Edwards, J., and Beckman, M. E. (2008a). "Methodology questions in studying consonant acquisition," *Clin. Linguist. Phonetics* **22**, 939–958.

Edwards, J., and Beckman, M. E. (2008b). "Some cross-linguistic evidence for modulation of implicational universals by language-specific frequency effects in the acquisition of consonant phonemes," *Lang. Learn. Dev.* **4**, 122–156.

Fenson, L., Marchman, V. A., Thal, D. J., Dale, P. S., Reznick, J. S., and Bates, E. (2007). *MacArthur-Bates Communicative Development Inventories (CDIs)* (Singular, San Diego), pp. 1–188.

Gavrilidou, M., Labropoulou, P., Mantzari, E., and Roussou, S. (1999). "Prodiagrafes gia ena ipologistiko morfologiko lexiko tis Neas Ellinikis" ("Specifications for a computational morphological lexicon of Modern Greek"), in *Greek Linguistics 97, Proceedings of the 3rd International Conference on the Greek Language*, edited by A. Mozer (Ellinika Grammata, Athens), pp. 929–936.

Hagiwara, R. (1995). "Acoustic realizations of American /r/ as produced by women and men," *UCLA Work. Pap. Phonetics* **90**, 1–187.

- Harrington, J., Kleber, F., and Reubold, U. (2008). "Compensation for coarticulation, /u/ -fronting, and sound change in standard southern British: An acoustic and perceptual study," *J. Acoust. Soc. Am.* **123**, 2825–2835.
- Hawkins, S., and Midgley, J. (2005). "Formant frequencies of RP monophthongs in four age groups of speakers," *J. Int. Phonetics Assoc.* **35**, 183–199.
- Hillenbrand, J., Getty, L. A., Clark, M. J., and Wheeler, K. (1995). "Acoustic characteristics of American English vowels," *J. Acoust. Soc. Am.* **97**, 3099–3111.
- Kaufman, N. R. (1995). *The Kaufman Speech Praxis Test for Children* (Wayne State University Press, Detroit, MI).
- Keating, P. A., and Huffman, M. K. (1984). "Vowel variation in Japanese," *Phonetica* **41**, 191–207.
- Ladefoged, P., and Maddieson, I. (1996). *The Sounds of the World's Language* (Blackwell, Malden, MA), Chap. 9, pp. 281–322.
- National Institute of the Korean Language (2002). "Hyeondae gugeo sayong bindo josa bogoseo" ("The frequency analysis of modern Korean"), <http://www.korean.go.kr> (Last viewed September 20, 2010).
- Nusbaum, H. C., Pisoni, D. B., and David, C. K. (1984). "Sizing up the Hoosier mental lexicon: Measuring the familiarity of 20,000 words," in *Research on Speech Perception Progress Report* (Speech Research Laboratory, Indiana University, Bloomington, IN), pp. 357–377.
- Peterson, G., and Barney, H. (1952). "Control methods used in a study of the vowels," *J. Acoust. Soc. Am.* **24**, 175–184.
- R Development Core Team (2009). "R: A language and environment for statistical computing," R Foundation for Statistical Computing, Vienna, Austria.
- Rvachew, S., Alhaidary, A., Mattock, K., and Polka, L. (2008). "Emergence of the corner vowels in the babble produced by infants exposed to Canadian English or Canadian French," *J. Phonetics* **36**, 564–577.
- Rvachew, S., Mattock, K., Polka, L., and Ménard, L. (2006). "Developmental and cross-linguistic variation in the infant vowel space: The case of Canadian English and Canadian French," *J. Acoust. Soc. Am.* **120**, 2250–2259.
- Sarkar, D. (2009). "Lattice: Lattice Graphics," R package version 0.17-26.
- Titze, I. R. (2000). *Principles of Voice Production* (National Center for Voice and Speech, Iowa City, IA), Chap. 6, pp. 149–184.
- Whiteside, S. P. (2001). "Sex-specific fundamental and formant frequency patterns in a cross-sectional study," *J. Acoust. Soc. Am.* **110**, 464–478.
- Wong, W.-Y. P., Brew, C., Beckman, M. E., and Chan, S. D. (2002). "Using the segmentation corpus to define an inventory of concatenative units for Cantonese speech synthesis," *Proceedings of the First SIGHAN Workshop on Chinese Language Processing* (Academia Sinica, Taipei, Taiwan), pp. 119–123.
- Yang, B. (1996). "A comparative study of American English and Korean vowels produced by male and female speakers," *J. Phonetics* **24**, 245–261.