EVIDENCE OF FINE PHONETIC DETAIL IN CHILDREN AND ADULTS' VOWELS: CROSS-LINGUISTIC PRODUCTION & PERCEPTION STUDIES

BY

HYUNJU CHUNG

A dissertation submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

(Communicative Disorders)

At the

UNIVERSITY OF WISCONSIN-MADISON

2011

EVIDENCE OF FINE PHONETIC DETAIL IN

CHILDREN AND ADULTS' VOWELS:

CROSS-LINGUISTIC PRODUCTION & PERCEPTION STUDIES

Hyunju Chung

Under the supervision of Professor Jan Edwards

At the University of Wisconsin-Madison

Previous studies of vowels produced by adults have shown that there are cross-linguistic differences in the acoustic characteristics of vowels that share the same phonemic symbol (e.g., Bradlow, 1995). It has also been reported that young children start to produce these language-specific phonetic details very early (e.g., de Boysson-Bardies, Halle, Sagart, & Durand, 1989). Although a considerable amount of work has been done to understand the acoustic characteristic of vowels produced by adult speakers, there is less work on vowel acquisition. This study is unique in the literature in that it examines production of the same set of "shared" vowels (vowels with the same phonemic symbol) across five languages and three age groups, using both static and dynamic acoustic measures, and it also examines the relationship between these acoustic measures and naïve-listeners' perception of these vowels.

The overall aim of the three studies included in this dissertation was to examine language-specific fine phonetic detail in shared vowels produced by adults, 5-year-olds, and 2-year-olds, who are monolingual speakers of American English, Cantonese, Greek, Japanese, or Korean. Values of the first two formant frequencies were extracted both at the vowel midpoint in Study 1, and from onset to midpoint (formant trajectories) in Study 3, and examined across vowels, languages, and age groups. The relationship between acoustic measures at vowel midpoint and vowel perception by naïve listeners was investigated in Study 2.

There were three major results of this study. First, significant cross-linguistic differences for some shared vowels were observed, both in F1 and F2 values measured at vowel midpoint, and in formant trajectories. Second, Study 1 found some significant within-language differences across age groups, especially between adults and 2-year-olds. However, Study 3 found that vowel trajectory patterns were highly similar across age groups. Finally, it was observed that native vowels generally received higher goodness ratings than non-native vowels by naïve listeners, and acoustic measures were more related to goodness ratings of non-native than native point vowels. The implication of these results will be discussed in terms of both vowel acoustics and vowel acquisition.

ACKNOWLEDGMENTS

I started the journey called “Ph.D.” five years ago in Sep 2006. I was very excited to start this new adventure in life, but back then, I was not aware of where this journey would take me, how long it would last, and what kinds of people I would meet along the way. After five years, I found myself getting ready to enter a world of “vowel acoustics” as a “baby” speech scientist, with unforgettable (both exhilarating and difficult) memories I had made with incredible people who shared this journey with me.

My advisor, Dr. Jan Edwards, was one of the very first people I met when I started this journey. Without her support, this adventure would not have been possible. Dr. Edwards allowed me to pursue the research I was interested in and guided me by endless encouragement, both academically and emotionally, which made me grow into a stronger and more expressive researcher. I would also like to express my genuine gratitude to Dr. Gary Weismer, who had introduced me to the world of vowel acoustics and eventually made me “vowel acousticky”! Whenever I faced obstacles, he always showed me the wisest way to overcome them, and stopped me from giving up the journey. He taught me not only important concepts of speech science, but also incalculable life lessons that are essential for starting another chapter of my life as a speech scientist. I would also like to deliver my sincere gratitude to Dr. Marios Fourakis for his advice, feedback, and most of all, his kind support that always made me smile.

Additionally, I would like to thank my committee members Dr. Margarita Kaushanskaya and Dr. Keith Kluender for their insightful comments, suggestions, and time. My thanks also goes to former and current members of the *Paidologos* project, Dr. Mary Beckman and Dr. Benjamin Munson for their input and technical support, and to Katerina Derdemezis, Carly

Faulhaber, Eun Jong, Kong, Asimina Syrika, Sarah Schellinger, and Ann Todd for not only being my colleagues, but also my best friends.

I also thank my many close friends, the COMDIS Korean girls, “Drs” Yunjung Kim and Jimin Lee for their endless and realistic advice, as well as their friendship, which was a huge motivation that helped me successfully end this journey. Likewise, I would like to thank my Korean friends both in Korea and in the U.S. who shared happiness, hardships, and prayers with me throughout the past five years.

Most of all, I thank God for giving me this precious opportunity, for opening a door for me when I knocked, and answering me every time I needed Him. I cannot thank enough my parents, grandmother, brother, and other family members who were always there to support me and show me how much they love and are proud of me. The endless love they displayed was **the reason** I could become the person I am today.

This work was supported by a Fulbright fellowship to Hyunju Chung and by NIDCD grant 02932 and NSF grant 0729140 to Jan Edwards.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

For the past several decades, extensive research on vowels has focused on understanding their developmental patterns (using either transcription or acoustic analysis), their acoustic characteristics, and their perception by native and non-native listeners. Developmental studies based on transcription showed that children can produce most or all of the vowels in their native vowel inventory correctly by 24 months (e.g., Otomo & Stoel-Gammon, 1992; Templin, 1957). Acoustic studies, however, showed that though children's vowels begin to have adult-like acoustic characteristics as early as 24 months (Buhr, 1980; Ishizuka, Mugitani, Kato, & Amano, 2007), increased variability in formant frequency values persists well past 24 months, finally reaching adult-like levels around the age of 14 (Lee, Potamianos & Narayan, 1999). This result suggests that it is important to examine fine-grained phonetic details in studies of vowel development, even though vowels may be perceived as correct by adult listeners.

Cross-linguistic acoustic studies of vowels have also provided support for examining fine-grained phonetic detail in vowel production. A number of studies have found that vowels sharing the same phonemic symbols across languages differ from each other acoustically (e.g., Bradlow, 1995; Yang, 1996). Moreover, developmental cross-linguistic studies have found that infants are sensitive to these cross-linguistic differences in vowel perception by 6 months of age (Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992). and begin to reflect these differences in their productions, as early as 6 to 10 months of age (e.g., de Boysson-Bardies, Halle, Sagart, & Durand, 1989; Rvachew, Mattock, Polka, & Menard, 2006). These findings suggest that using a cross-linguistic paradigm is useful for examining the emergence of language-specificity in children's vowel production.

In this study, cross-linguistic developmental patterns of “shared” vowels were investigated using two different acoustic measurements, traditional steady-state measurement of F1 and F2, and also F1 and F2 trajectories from vowel onset to midpoint. In addition, the relationship between acoustic characteristic of “shared” vowels and listeners’ perception of these vowels was examined. The languages and vowel categories examined in this dissertation are summarized in Table 1.1. The research questions and methodology of each of the three studies are summarized briefly in the next section.

Table 1.1 Target languages and vowels of this study

	Focus of the study	**Target languages**		**Target vowels**
Study 1	Production (midpoint measured F1 & F2)	Cantonese English Greek Japanese Korean		/a/, /i/, /u/
Study 2	Perception	Speaker languages	Listener languages	/a/, /i/, /u/, /e/, /o/
		Cantonese English Greek Japanese Korean	English Greek Korean	
Study 3	Production (F1 & F2 trajectories)	English Greek Korean		/a/, /i/, /u/, /e/, /o/

Research Questions

Study 1: Research Questions: Are there acoustic differences among the three point vowels of five languages with different vowel inventories? If so, are these cross-linguistic differences observed by 2- and 5-year-old speakers of these languages?

Method: The acoustic characteristics of the three point vowels, /a/, /i/, and /u/, of each of five languages were examined systematically to understand the language-specific phonetic characteristics of the shared vowels. In addition, this study examined whether 2- and 5-year-old children produce these language-specific characteristics. In order to compare the acoustic characteristics of vowels produced by different age, sex, and language groups, the raw formant frequency values were transformed into speaker-specific values which presumably controlled for the effect of different vocal tract sizes.

Study 2: Research Questions: Are the cross-linguistic differences in fine phonetic detail observed in Study 1 reflected in the vowel categorization and goodness ratings of naïve adult listeners of three of the languages included in the production study? Are the goodness ratings systematically related to the acoustic characteristics of native and non-native vowels? Do vowel categorization and goodness ratings differ systematically across vowels produced by native vs. non-native speakers and by speakers of different ages within a language?

Method: Native adult speakers of American English, Greek, and Korean were asked to categorize and rate the goodness of vowels produced by 2-year-olds, 5-year-olds, and adults of American English, Cantonese, Greek, Japanese, and Korean. Vowels were blocked by age-group, but not by language. The focus of the analysis in this study was on analyzing listeners' goodness ratings of native and non-native vowels that were "correctly" classified (target category matched the perceptual category). The goodness ratings of each listener group were analyzed as a function of nativeness (native vs. non-native), age of speakers, and acoustic characteristics.

Study 3: Research Questions: Are language-specific patterns observed in the F1 or F2 trajectories (from vowel onset to midpoint) of shared vowels? If so, do 2- and 5-year-old children produced this language-specific detail in their F1 and F2 trajectories?

Specific Aims: This study examined F1 and F2 trajectory patterns of five shared vowels /a, i, u, e, o/, produced by native speakers (2- and 5-year-olds and adults) of American English, Greek, and Korean. The differences between language-specific characteristics found in traditional “slice-in-time” F1 and F2 values measured at vowel midpoint and in vowel spectral patterns over time are discussed in terms of their implications to how vowels are described acoustically, as well as to theories of vowel perception.

Significance

The current study differs from previous work on vowel production/perception/acquisition in the following respects: 1) it examines acoustic characteristics of vowels produced by a relatively large number of monolingual native speakers of different languages and age groups; 2) it examines the acoustic characteristics of vowels, how they are perceived by adults, and the relationship between vowel acoustics and vowel perception by naïve listeners; 3) it employs a novel analysis method that allows a comparison of vowels produced by speakers of different age, sex, and language groups in a speaker-specific vowel space; and 4) it investigates not only the “slice-in-time” acoustic characteristics of vowels, but also their trajectory patterns.

Outline of the Chapters

This dissertation consists of three studies. Chapter 1 provides a general background on vowel acoustic studies and explains why the following three studies are an important contribution to the existing studies. Chapter 2 (Study 1) focuses on the acoustic characteristic of vowels (F1 and F2 values measured at vowel midpoint) produced by monolingual adults, 5-year-olds, and 2-year-olds of American English, Cantonese, Greek, Japanese, and Korean. Chapter 3 (Study 2) examines the relationship between language-specific phonetic details of vowels observed in Study 1 and listeners' perception of these characteristics. Chapter 4 (Study 3) examines F1 and F2 trajectories of a subset of vowels and languages to understand cross-linguistic and developmental patterns of vowel spectral movement patterns. Finally, Chapter 5 concludes the dissertation by discussing the implication of the findings of the three studies, both to theories of vowel production and vowel perception.

CHAPTER 1: REFERENCES

Bradlow, A.R. (1995). A comparative acoustic study of English and Spanish vowels. *Journal of the Acoustical Society of America, 97*, 1916-1924.

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

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., and Lindblom, B. (1992). Linguistic Experience Alters Phonetic Perception in Infants by 6 Months of Age. *Science*, 255, 606-608.

Lee, S. Potamianos, A., & Naraynan, S. (1999). Acoustics of children's speech: Developmental changes of temporal and spectral parameters. *The Journal of the Acoustical Society of America, 105 (3).* 1455-1468.

Otomo, K. & Stoel-Gammon, C. (1992). The Acquisition of Unrounded Vowels in English. *Journal of Speech and Hearing Research. 35*. 604-616.

Rvachew, Mattock, Polka, & Menard (2006). Developmental and cross-linguistic variation in the infant vowel space: The case of Canadian English and Canadian French. *The Journal of the Acoustical Society of America, 120*, 2250-2259.

Selby, J. C., Robb, M. P., & Gilbert, H. R. (2000). Normal vowel articulations between 15 and 36 months of age. *Clinical Linguistics & Phonetics. 14(4).* 255-265.

Templin, M. C. (1957). *Certain language skills in children: Their development and interrelationships*. Minneapolis: University of Minnesota Press.

Yang (1996). A Comparative study of American English and Korean vowels produced by male and female speakers, *Journal of Phonetics*. *24*, 245-261.

CHAPTER 2: STUDY 1

Cross-linguistic Studies of Children's and Adults' Vowel Spaces

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, B. L. & MacNeilage, P. F., 1990) suggest 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, differ 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 of these 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, Halle, Sagart, and Durand (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, Mattock, Polka, and Menard (2006) and Rvachew, Alhaidary, Mattock, and Polka (2008) also demonstrated cross-linguistic differences in the vowel spaces of 8- to 18- months-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 one year. The findings of these studies, also demonstrated that research on vowel acquisition patterns need 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, including 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 eleven vowel phonemes (/i: y:, e, ɛ:, œ:, ø, ɐ, a:, u: o, ɔ:/). American English also has a relatively large vowel inventory, with four short vowels (/ɪ, ɛ, ə, ʊ/) and seven long vowels (/i, e, æ, ɑ, ɔ, o, u/), resulting in eleven monophthong vowels altogether (Hillenbrand, Getty, Clark, & Wheeler, 1995; Peterson & 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, Japanese high back vowel /u/ should be symbolized as [ɯ], because this high back vowel 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 & Huffman, 1984). Finally, Korean has eight monophthong vowels (/a, ɛ, e, i, o, u, ʌ, ɨ/). However, the non-high vowels /e/ and /ɛ/ 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 /ɛ/ 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 2.1.

Table 2.1 Vowel inventories of Cantonese, English, Greek, Japanese and Korean

Cantonese

	front	central	back
high	i: y		u:
mid	e	ɵ	o:
	ɛ œ:		ɔ:
low		ɐ a:	

English

	front	central	back
high	i		u
	ɪ		ʊ
mid	e		o
	ɛ	ə ɝ	
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		ɑ	

Korean

	front	central	back
high	i	ɨ	u
mid	e		o
	ɛ	ʌ	
low		a	

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 are 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 (Adank, Smits, & Hout, 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 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.

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 above. 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.

.

METHODS

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

Participants

Participants used 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). Demographic data for the 150 participants are provided in Table 2.2. 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, Ohio; 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.

Table 2.2. Mean age (in month) for 2- and 5-year-olds, and the mean age (in year) 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)

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 & Huffman, 1984). For all languages, the words used to elicit target vowels were real words that were pictureable 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. 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 2.3.

Table 2.3 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, s, ʃ, tʃ, g, k]	[d, t , t^w, s, ʃ, tʃ, g, k, k^w]	[d, t , s, □, tʃ, k, k^j]
Greek	[t, s, ts, k, k^j, x]	[θ, t, s, ts, ç, k^j]	[t, s, ts, k]
Japanese	[d, t , s, ʃ, tʃ, g, k, k^j]	[t, ʃ, tʃ, g^j, k^j]	[s, ts, tʃ, dz, dʃ, g, k, k^j]
Korean[1]	[t, t^h, t', s, s', tʃ, tʃ^h, k, k^h, k']	[t, t^h, t', ʃ, ʃ', tʃ, tʃ^h, tʃ', k, k']	[t, t^h, t', s, s', tʃ, tʃ^h, tʃ', k, k^h, k']

[1]Korean stops have three-way laryngeal contrast /t/ represents lax stop, /t^h/ represents aspirated stop, and /t'/ represents tense stop.

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 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-inch 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 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 and 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, /ɑ/, /ʌ/, and /ɔ/ into the /ɑ/ category, and /u/ and /ʊ/ into the /u/ category. Similarly, in Cantonese, /a/ or /ɐ/ 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. Table 2.4. 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 2.5. It should be noted that the accuracy rates for English and Cantonese may be somewhat inflated, given that the accuracy-coding procedure described above. Even so, the accuracy rate for vowels produced by 2-year-old Cantonese speakers is the lowest of all of the languages.

Table 2.4 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 2.5 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

Data Analysis

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 & Weenink, 2006). The size of the analysis window was 25ms and the dynamic range was set at 30dB. In order to determine the 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, Getty, Clark, & Wheeler, 1995; Peterson & 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.

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. 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 will depend 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 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). 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 Figure 2.1). The radius and degree values represent the size and the rotation of the vowel space around the 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.

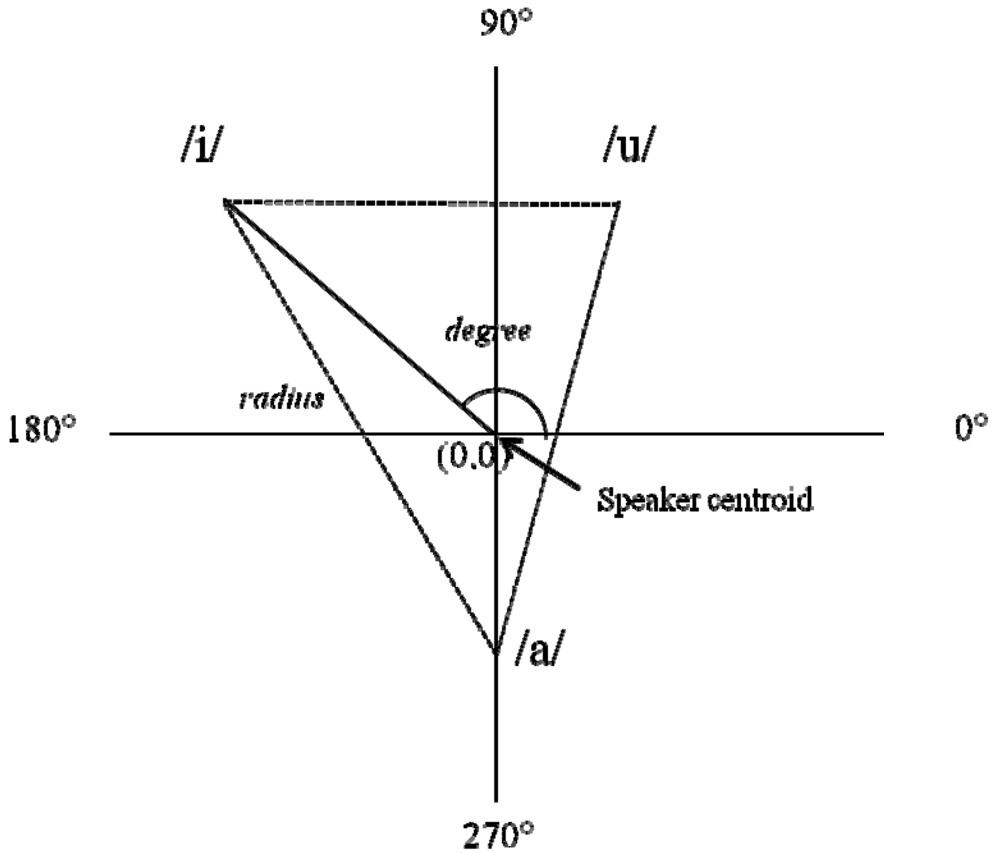

Figure 2.1 A Cartesian coordinate system showing how the radius and degree values for /i/ was calculated.

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 & 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 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°. We used a weighted least-square analysis to circumvent this problem by calculating an average degree value for each vowel for each speaker and thus placing all of degree values on a semi-circle. Because there was only a mean observation for each vowel for each subject, we did not need to account for the effect of individuals on degree values. In the regression model, the independent variables were language and age. The analyses were done separately for each vowel category.

RESULTS

Traditional vowel space with F1 - F2 values

Scatter plots of F1 vs. F2 by language and age are shown in Figure 2.2. 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 Figure 2.2 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. 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 by2-year-olds, 5-year-olds, and adults of Cantonese, English, Greek, Japanese or Korean were summarized in Appendix A.

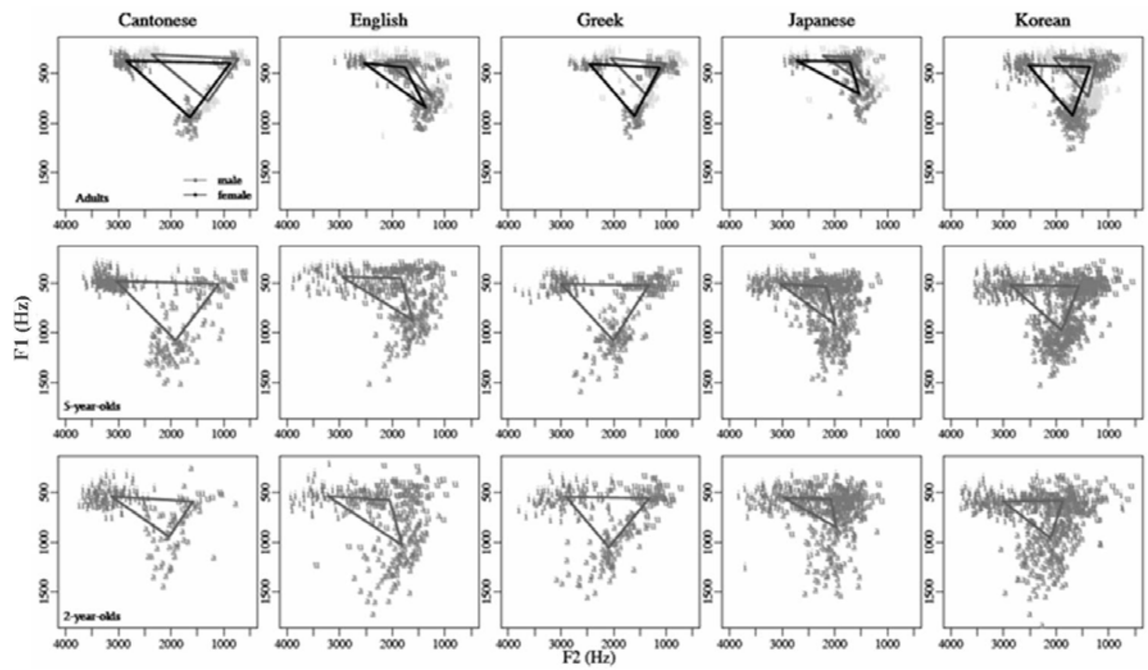

Figure 2.2 Scatter plots of F1 and F2 by language and age for both males (grey) and females (black). Each point represents the production of a single vowel by a single speaker. The vowel space was drawn by connecting the mean F1-F2 value of all productions of each vowel for each language and age group.

Transformed vowel space

Figure 2.3 shows the transformed vowel spaces of each language and age group following the computational process outlined in the Method section. This transformation allows the comparison of size and shape of vowel spaces across speakers without having the confound effect coming from different vocal tract sizes of speakers. When compared to Figure 2.2, two cross-linguistic differences that were not found with raw F1 and F2 values were observed in the transformed data presented in Figure 2.3. First, the relative location of /i/ and /u/ in adult English and Japanese productions were different in the raw versus 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 Figure 2.2 and Figure 2.3. 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.

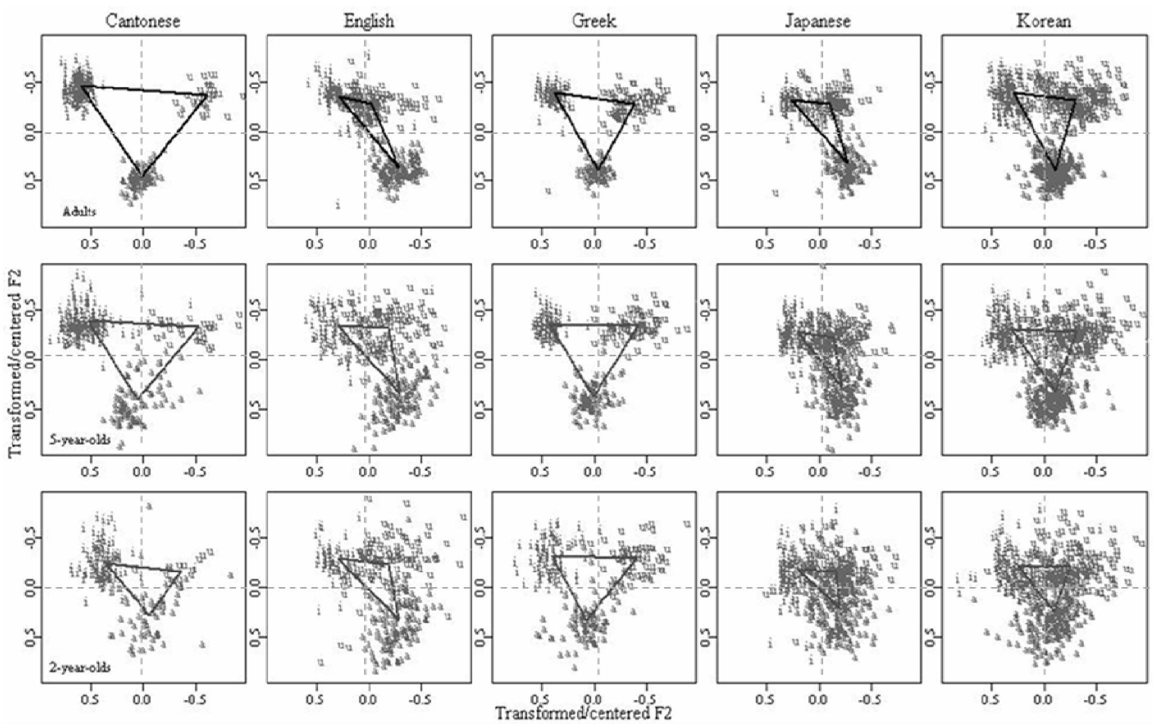

Figure 2.3 Scatter plots of transformed F1-F2 by language and age. The transformations were made by subtracting each individual's mean F1 and F2 values from log-transformed F1 and F2 values of each token. The measurements were centered at the origin, (0,0), the speaker centroid. The vowel space was drawn by connecting the mean log-transformed F1-F2 value of all productions of each vowel for each language and age group.

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

Figure 2.4 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 Figure 2.4 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 2nd and 3rd columns of Figure 2.4). This difference was greatest between Cantonese versus 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/ vowels 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 2.6. For /a/ produced by adult speakers, however, no substantial cross-linguistic difference in radii values was observed.

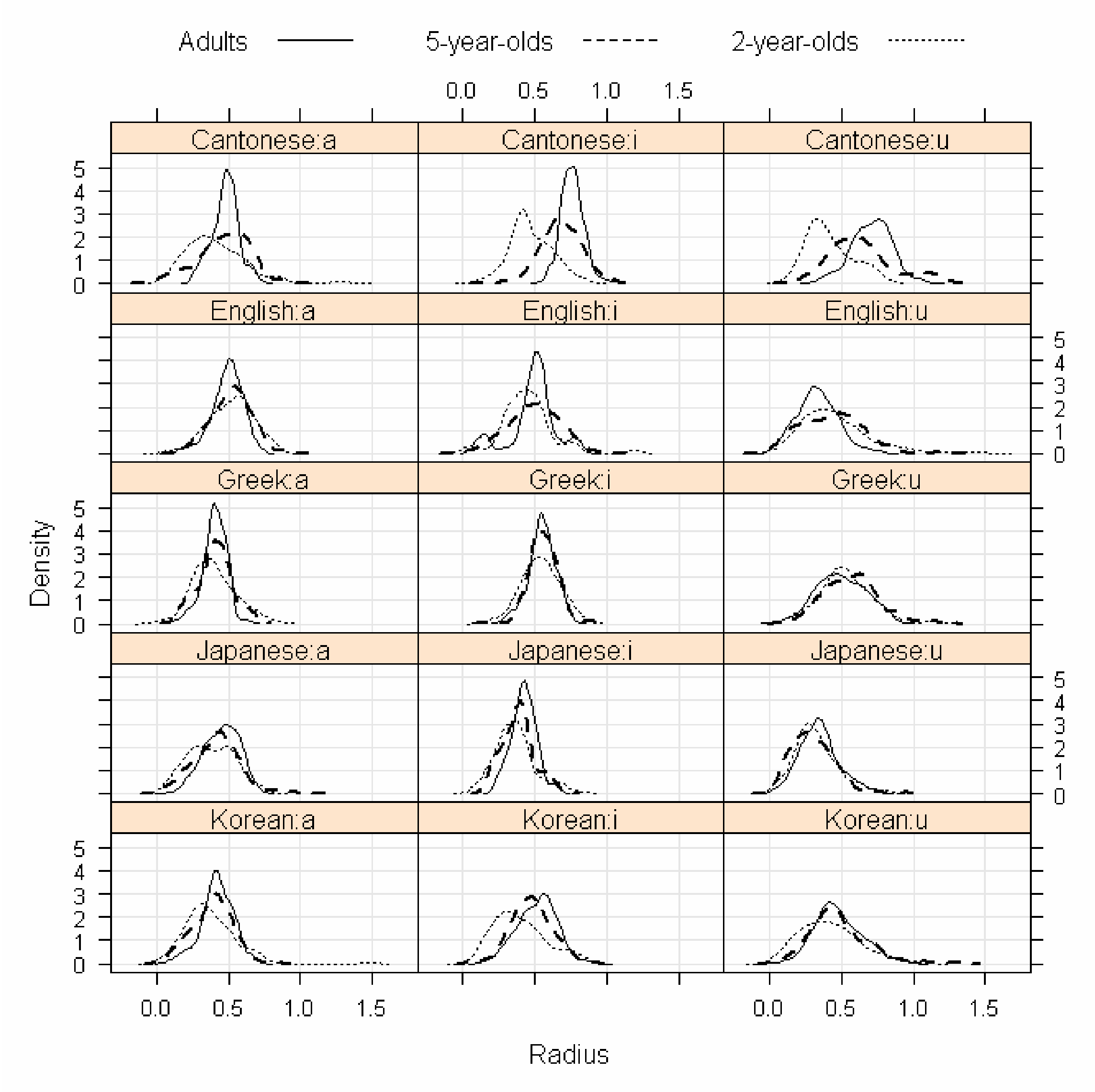

Figure 2.4. Density plots of radius values sorted by language, age, and vowel. The x-axis shows radius values, and the y-axis represents the density, which is the relative frequency of occurrence for different radius values. Solid lines show data for adults; dotted lines show data for 5-year-olds; and small dotted lines show data for 2-yar-olds.

For across age groups of 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 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-old and the adults, suggesting that the vowels produced by 2-year olds have 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 & Maechler, 2009) does not provide a detailed post-hoc analysis. Therefore, differences between group means were used to interpret the interaction (see Table 2.6). 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.

Table 2.6 Mean radii for each vowel for each language and age group estimated from mixed-effects regression models.

	/a/			/i/			/u/		
	2yr.	5yr.	Adults	2yr.	5yr.	Adults	2yr.	5yr.	Adults
Cantonese	0.400	0.466	0.477	0.478	0.682	0.759	0.425	0.640	0.71
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

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

Figure 2.5, 2.6 and 2.7 shows 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 1st quadrant) to the bottom right part of the circle (the 4th 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 2.7.

As Figure 2.5 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 3rd and the 4th quadrant, whereas those of English and Japanese adults were mostly in the 4th 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 occupies a more back location in the vowel acoustic space compared to /a/ produced by Cantonese, Greek, and Korean adults. This is consistent with previous findings, in which English /a/ was transcribed with a different IPA (International Phonetic Alphabet) symbol ([ɑ]), 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 & Huffman, 1984). No clear cross-linguistic difference was observed in the degree distribution pattern of /i/. As shown in Figure 2.6, /i/ productions of adults of all languages were clustered mainly in the 2nd 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 Figure 2.6; no pair of languages showed a significant difference in their degree values with each other. As can be seen in Figure 2.7, 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 (1st quadrant) and left parts (2nd 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 1st quadrant). The result 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 adults' productions of English /u/ occupy the most front location in the vowel acoustic space, while Cantonese /u/ occupies the most back location.

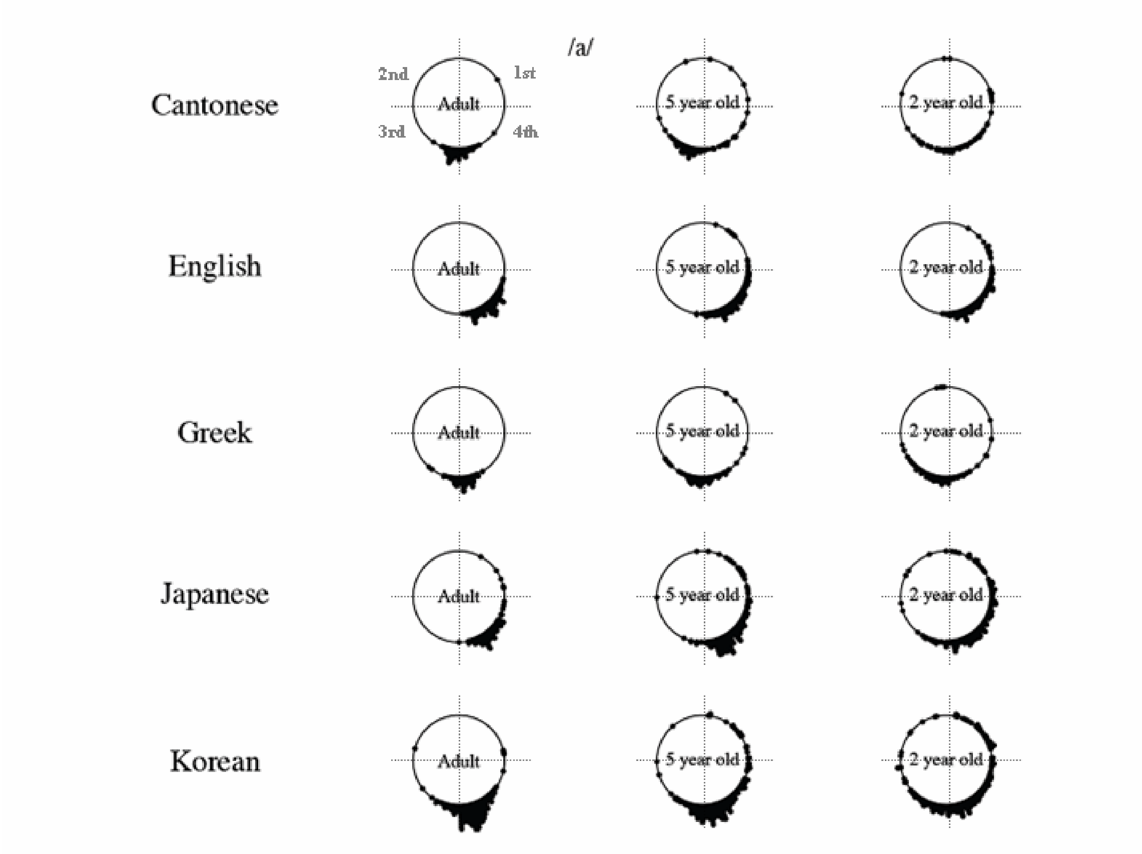

Figure 2.5 Circular plots for /a/. Each dot represents degree values of each vowel. The degree values were calculated and placed on the arc of a circle from the origin (0,0).

Table 2.7 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

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 vowels produced by 2-year-olds' were consistently smaller than those produced by older age groups for /a/ and /u/. This result suggests that vowels produced by 2-year-olds occupy a more central location in the vowel acoustic space than those of older age groups. That is, for /a/, vowels occupy a relatively more low mid position in the vowel acoustic space initially, but move to a more low back location with age (In the Cartesian system, this is a movement from the 3rd quadrant to 4th quadrant). Similarly, for /u/, the degree values of /u/ produced by 2-year-olds clustered around the center of the vowel acoustic space (90 degrees), and these values increased with age (being de-centralized into a more back location in the vowel acoustic space). For /i/, however, Tukey's HSD post hoc analysis showed no significant age group difference. That is, 2-year-olds' productions of /i/ showed 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.

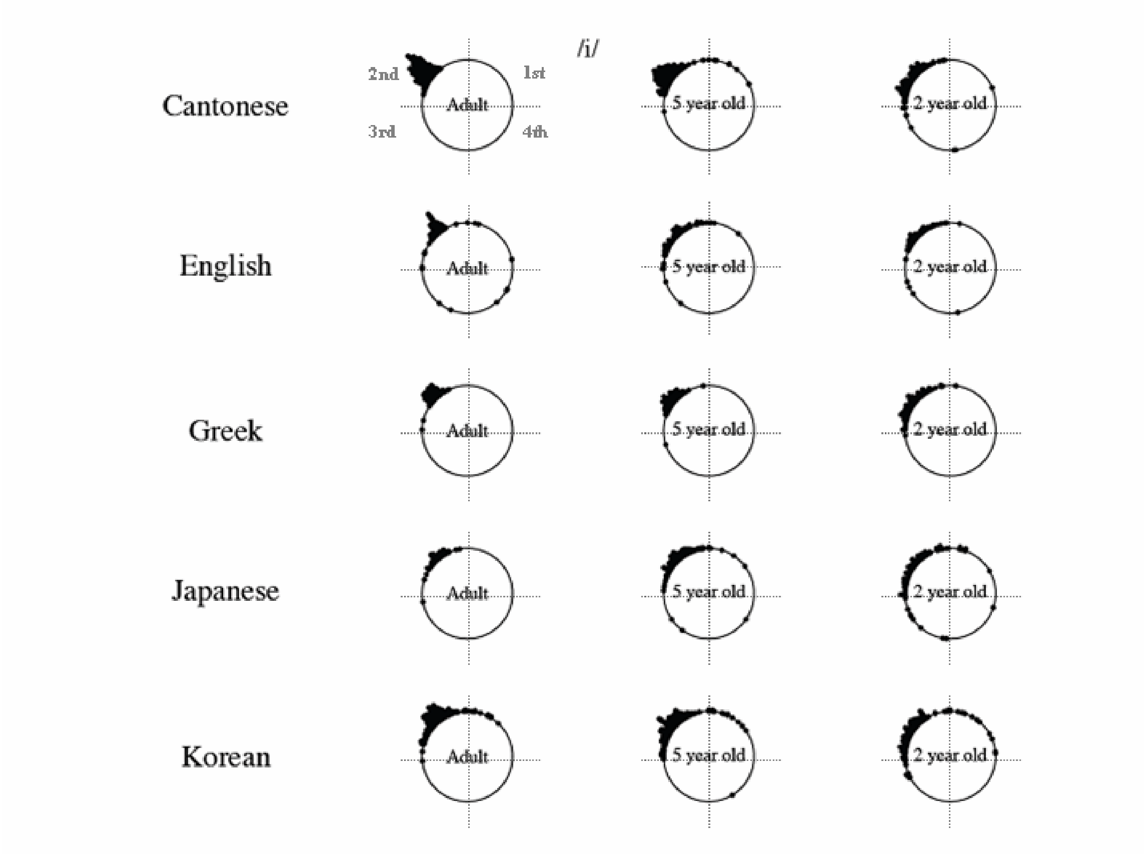

Figure 2.6 Circular plots for /i/.

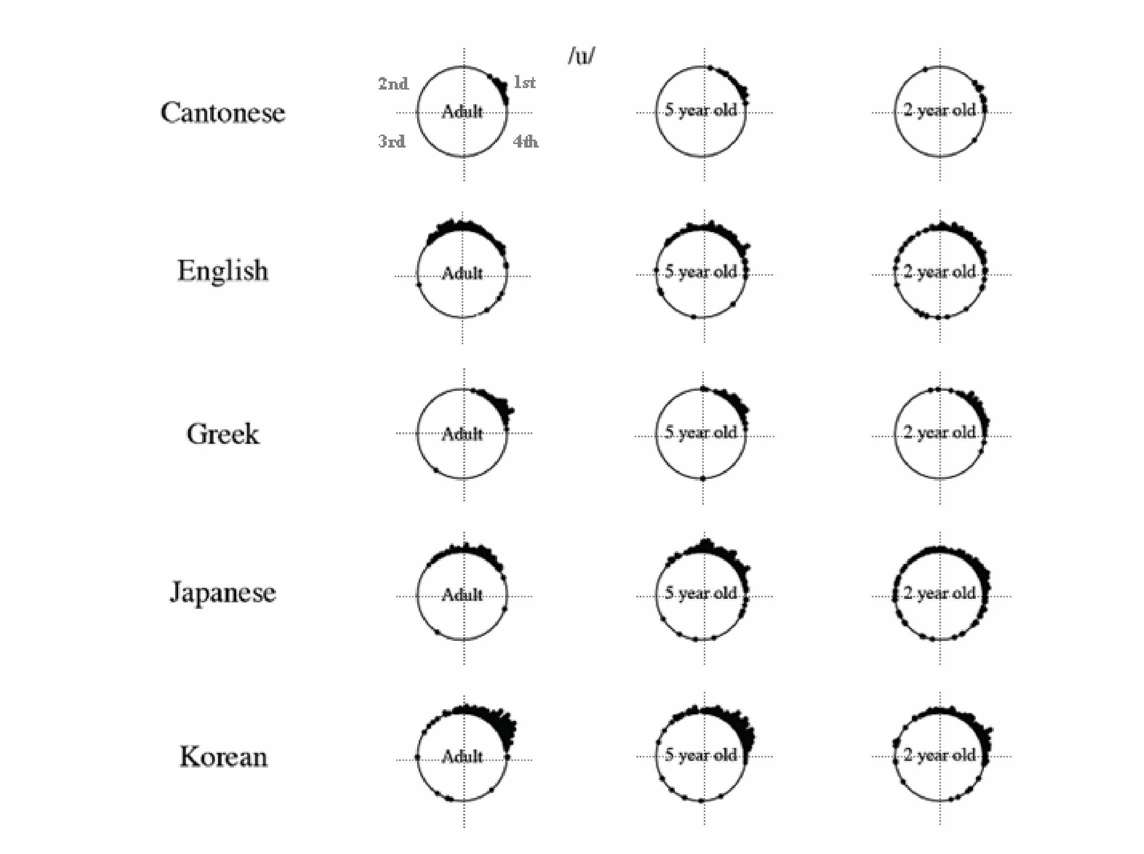

Figure 2.7 Circular plots for /u/.

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 adult were significantly larger than those of 5- and 2-year-old, 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 two. 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/.

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, the cross-linguistic difference was 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 & Pisoni, 2006; Hagiwara, 1995; Harrington, Kleber, & Reubold, 2008; Hawkins & Midgley, 2005; Hillenbrand *et al.*, 1995) and as a phonemic characteristic with higher F2 values (e.g., Keating & 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' vowel 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 that were observed for the 2-year-olds's 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 above, /u/ is low in frequency in Cantonese because of a phonotactic restriction that allows /u/ 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, L., Marchman, V. A., Thal, D. J., Dale, P. S., Reznick, J. S., & Bates, E., 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 B for results and description of procedure). 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 & Pisoni, 2006; Hagiwara, 1995; Harrington, Kleber, & Reubold, 2008; Hawkins & 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.

CHAPTER 2: REFERENCES

Adank, P. Smits. R., & van Hout, R (2004). A comparison of vowel normalization procedure for language variation research. *Journal of the Acoustical Society of America*, *116*, 3099-3107.

Amano, S. & Kondo, T. (1999). *Lexical properties of Japanese.* Tokyo: Sanseido.

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

Boersma, P. & Weenink, D. (2006). Praat: doing phonetics by computer [Computer program]. Retrieved Aug, 2006, from http://www.praat.org/.

Bradlow, A.R. (1995). A comparative acoustic study of English and Spanish vowels. *The Journal of the Acoustical Society of America*, *97*, 1916-1924.

Chan, S. D., & Tang, Z. X. (1999). Quantitative analysis of lexical distribution in different Chinese communities in the 1990's. *Yuyan Wenzi Yingyong [Applied Linguistics]*, 3, 10-18.

Clopper, C., & Pisoni, D. B. (2006). The Nationwide Speech Project: A new corpus of American English dialects. *Speech Communication, 48*, 633-644.

Davis, B. L. & MacNeilage, P. F. (1990). Acquisition of correct vowel production: a quantitative case study. *Journal of Speech and Hearing Research, 33*, 16-27.

de Boysson-Bardies, B., Halle, P., Sagart, L. & Durand, C. (1989). A cross-linguistic investigation of vowel formants in babbling. *Journal of Child Language, 16*, 1-17.
Davis, B. L. & MacNeilage, P. F. (1990). Acquisition of correct vowel production: a quantitative case study. *Journal of Speech and Hearing Research, 33*, 16-27.

Disner, S. F. (1980). Evaluation of vowel normalization procedures. *Journal of the Acoustical Society of America,* 67, 253-261.

Edwards, J. & Beckman, M. E. (2008a). Methodology questions in studying consonant acquisition, *Clinical Linguistics & Phonetics. 22*, 939-958.

Edwards, J. & Beckman, M. E. (2008b). Some cross-linguistic evidence for modulation of implicational universals by language-specific frequency effects in the acquisition of consonant phonemes, *Language Learning & Development*, *4,* 122-156.

Fenson, L., Marchman, V. A., Thal, D. J., Dale, P. S., Reznick, J. S., & Bates, E. (2007). *MacArthur-Bates Communicative Development Inventories (CDIs).* San Diego, California: Singular Publications.

Gavrilidou, M., Labropoulou, P., Mantzari, E., & Roussou, S. (1999). Prodiagrafes gia ena ipologistiko morphologiko lexiko tis Neas Ellininkis [Specifications for a computational morphological lexicon of Modern Greek], In A. Mozer (Ed.), *Proceedings of the 3rd International Conference on the Greek Language*, (pp. 929-936). Athens.

Hagiwara, R. (1995). Acoustic Realizations of American /r/ as produced by women and men, *UCLA Working Papers in Phonetics. 90*, 1-187.

Harrington, J., Kleber, F., & Reubold, U. (2008). Compensation for coarticulation, /u/-fronting, and sound change in standard southern British: An acoustic and perceptual study. *Journal of the Acoustical Society of America*, *123*, 2825-2835.

Hawkins, S., & Midgley, J. (2005). Formant frequencies of RP monophthongs in four age groups of speakers. *Journal of International Phonetic Association, 35*, 183-199.

Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels, *Journal of the Acoustical Society of America,* 97, 3099-3111.

Kaufman, N. R. (1995). *The Kaufman Speech Praxis Test for Children.* Detroit, Michigan: Wayne State University Press.

Keating, P. A. & Huffman, M. K. (1984). Vowel variation in Japanese, *Phonetica. 41,* 191-207.

NIKL. (2002). "Hyeondae gugeo sayong bindo josa bogoseo" [The frequency analysis of modern Korean]. Electronic Document. The National Institute of the Korean Language (Seoul, Korea). http://www.korean.go.kr.

Nusbaum, H. C., Pisoni, D. B., & David, C. K. (1984). "Sizing up the Hoosier mental lexicon: Measuring the familiarity of 20,000 words," in *Research on Speech Perception Progress Report*. (Bloomington, IN: Speech Research Laboratory, Indiana University). pp. 357-377.

Peterson, G., & Barney, H. (1952). Control methods used in a study of the vowels, *Journal of the Acoustical Society of America, 24*, 175-184.

R Development Core Team (2009). *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0.

Rvachew, S., Alhaidary, A., Mattock, K, & Polka, L. (2008). Emergence of the corner vowels in the babble produced by infants exposed to Canadian English or Canadian French, *Journal of Phonetics,* 36, 564-577.

Rvachew, S. M. (2006). Developmental and cross-linguistic variation in the infant vowel space: The case of Canadian English and Canadian French. *Journal of the Acoustical Society of America, 120(4),* 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, Iowa), Chap. 6, pp. 149-184.

Whiteside, S. P. (2001). Sex-specific fundamental and formant frequency patterns in a cross-sectional study, *Journal of the Acoustical Society of America,* 110, 464-478.

Wong, W. –Y. P., Brew, C., Beckman, M. E., & 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 . *Journal of Phonetics, 24*, 245-261.

CHAPTER 2: APPENDIX

Appendix A. 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	7 30	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 B: 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 & Tang, 1999; Wong, Brew, Beckman, & Chang, 2002) for Cantonese, the *Hoosier Mental Lexicon* (HML, Nusbaum, Pisoni, & David, 1984) for English, the ILPS database (Gavrilidou, Labropoulou, Mantzari, & Roussou, 1999) for Greek, and the NTT database (Amano & Kondo, 1999) for Japanese, and NIKL database (NIKL, 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 the natural log of this quotient was taken 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

CHAPTER 3: STUDY 2

Listeners' sensitivity to native versus non-native vowels produced by children and adults

INTRODUCTION

The previous chapter examined language-specific characteristics of the three shared point vowels, /a/, /i/, and /u/, produced by monolingual 2-year-olds, 5-year-olds, and adults of five different languages, Cantonese, American English, Greek, Japanese, and Korean. The results showed that vowel spaces formed by the three point vowels of each language differ systematically across languages. Furthermore, the results showed that children's vowel productions were similar to those of adults (especially those of 5-year-olds'), although they were more variable (especially those of 2-year-olds) than those of adults'. The interpretation of this developmental pattern was that vowel production continues to be refined even after children's vowels are perceived as "correct" by native listeners, and that this phonetic refinement continues at least through the age of 5 years.

The purpose of the present study is to assess whether naïve adult listeners perceive the cross-linguistic and developmental differences observed in the previous chapter. Adult speakers of three different first languages, American English, Greek, and

Korean[1], categorized and rated goodness of both native and non-native vowels. Based on previous research, three predictions were made. First, it was predicted that non-native vowels would be categorized in the same phonemic category as native vowels except in those cases where there were significant phonetic differences between the non-native and native categories (e.g., English and Japanese /u/ relative to Greek, Cantonese, and Korean /u/). Second, it was predicted that native vowels would be rated as better vowels than non-native vowels categorized into the same vowel category. Finally, it was predicted that while native vowels produced by 5-year-olds would be categorized and rated similarly to those produced by adults, native vowels produced by 2-year-olds would not be categorized as accurately or rated as highly as those produced by 5-year-olds and adults.

Previous Cross-linguistic Studies

While the earliest studies of cross-linguistic vowel perception suggested that vowel perception is shaped by language-universal mechanisms (e.g., Stevens, Libermann, Studdert-Kennedy, & Ohman, 1969), more recent studies have shown that it is strongly influenced by linguistic experience (e.g., Best, 1995; Flege, 1992; 1995). Stevens *et al.* (1969) studied American English-speaking and Swedish-speaking listeners' perception of synthetic steady-state rounded versus unrounded vowels. The vowel systems of English

[1] The experiment was conducted with Cantonese listeners as well. However, the results from Cantonese listeners were not included in this study due to the nature of the Cantonese orthographic system. Cantonese has an orthographic system that does not separate vowels from consonants. Therefore, native Cantonese listeners had extreme difficulty in performing the vowel categorization task, where they were asked to tease apart vowels from consonants and click the vowels they heard. On average, Cantonese listeners needed 2 ½ or up to 3 times longer to complete the perception task, as compared to English, Greek or Korean listeners. Their "accuracy" rates (whether they classified vowels in the same categories as in the native language) were also substantially lower than the accuracy rates for the listeners of the other three languages.

and Swedish differ from each other in that unrounded vowels are phonemic in both languages (English: /i/, /ɪ/, /ɛ/; Swedish: /i/, /e/, /a/), while rounded vowels are phonemic only in Swedish (/i/, /y/, /ʉ/), but not in English. Stevens and colleagues found that when both English-speaking and Swedish-speaking listeners were asked to identify and discriminate synthetic vowels in /i/ to /ɛ/ or /i/ to /ʉ/ continua, both groups of listeners showed similar discrimination patterns for both rounded and unrounded vowels, although more variability was found for English-speaking listeners. Stevens *et al*. (1969) interpreted this result to mean that general auditory factors rather than specific linguistic factors influence listeners' ability to discriminate phonetic differences in vowel categories.

However, more recent research has shown that linguistic experience also plays a role. In another cross-language vowel perception study, Flege, Munro, and Fox (1994) asked both English-speaking and Spanish-speaking subjects to rate the perceptual similarity of vowel pairs that varied in their F1-F2 distance. They found that vowel pairs with greater distance in the F1-F2 space were perceived as more dissimilar by both English-speaking and Spanish-speaking speakers. However, when the two listener groups were tested with both adjacent and nonadjacent vowel pairs, English-speaking speakers' ratings for adjacent vowel pairs were more dissimilar than those of Spanish-speaking speakers, although the ratings for nonadjacent vowel pairs were similar in both groups of listeners (adjacent vowel pairs are vowels that could belong to the same vowel category for Spanish speakers, but not for English speakers (e.g., Spanish /a/-English /ɛ/, Spanish /a/-English /ɑ/, or Spanish /a/-English/æ/) and nonadjacent vowel pairs are vowel pairs that belong to two different vowel categories for both Spanish and English speakers (e.g., Spanish /a/-English /i/, Spanish /a/-English /ɪ/, or Spanish /a/-English/e$^{\text{ɪ}}$/)). These two different results suggest that listeners' perception of vowels is governed by either general auditory factors or linguistic experience, depending on the nature of the task.

Recent research has focused on the role of linguistic experience on vowel perception. That is, how does experience with the native language (L1) affect listeners' perception of the second language (L2)? Two general models of cross-language speech perception have been posed: Best (1995)'s PAM (Perceptual Assimilation Model) and Flege (1992, 1995)'s SLM (Speech Learning Model). Although these two models differ in explaining how unfamiliar F2 vowels are assimilated by listeners, they both explain adults' perception of vowels and consonants based mainly on experience with the first language. Best's PAM model claims that when a non-native sound that is phonetically similar to L1 category is heard, second language (L2) learners assimilate it into a L1 category. However, if they hear a non-native sound that does not exist in L1, then listeners classify it as "uncategorizable". Similar to the PAM model, Flege's SLM model also highlights the role of L1 on L2 perception. However, these two models differ in that the SLM model claims that when listeners hear a L2 sound that is acoustically similar to a L1 category, it will be assimilated to the corresponding L1 category, and eventually it will form a sound category that has the characteristics of both L1 and L2 sounds. However, when listeners hear L2 sounds that are not similar to any of the L1 phonetic category, listeners will form a new category that is in between L1 and L2 sound categories and with more experience, the sound will eventually be produced and perceived accurately as the target L2 sound.

Several studies have shown how these two models can be used to explain the interplay between L1 and L2 sound in cross-language vowel perception. Strange, Akahana-Yamada, Kubo, Trent, Nishi, and Jenkins (1998) examined Japanese listeners' perceptual assimilation pattern of American English vowels produced in a /hVd/ context. Strange and colleagues found that American English vowels that have similar spectral patterns to Japanese vowels, mostly the three point vowels /i/, /a/, and /u/ and the diphthongized vowels /ei/ and /ou/, were assimilated into "phonetically-similar" Japanese vowel categories and rated as good examples. However, American English vowels that do not have similar spectral counterparts in the Japanese vowel inventory (such as /æ/) showed variability in their assimilation pattern to Japanese vowel categories and were not rated as good examples of Japanese vowels. The results of this study suggested that listeners were capable of attending to small differences in the spectral characteristics of native and non-native sounds and make judgments based mainly on the phonetic similarities between native and non-native vowels, though variability in the response was also observed. These results are consistent with both the Best (1995) PAM model and the Flege (1992) SLM model. The phonetic similarity between native and non-native vowels, as characterized by F1 and F2 values, influenced non-native vowel perception of Japanese speakers. A similar study (Nishi, Strange, Akahane-Yamada, Kubo, & Trent-Brown, 2008) examined American English listeners' perceptual assimilation pattern of the five Japanese vowels in a nonsense [hVba] disyllable both in citation and carrier sentence forms. This study found that American English listeners assimilated long (2-mora) Japanese vowels, [ii], [ee], [aa], [oo], and [uu], into English tense vowel categories, [i:], [e], [ɑ:], [o], and [u:], respectively, and short (1-mora) Japanese vowels,

[i], [ɯ], and [o], into [i:], [u:], and [o], respectively. Similar to the previous study, listeners assimilated L2 vowels into L1 vowel categories based mostly on the spectral similarities between two vowel categories, although this pattern of perceptual assimilation was not consistent across all vowel categories. In sum, previous cross-language vowel perception studies have shown that L1 plays an important role on listeners' perception of vowels. In particular, these studies found that listeners assimilate non-native vowels into the native vowel category if possible, based on acoustic similarity, and also attend to fine-grained cross-linguistic differences during vowel categorization.

Despite the large literature on cross-linguistic vowel perceptual patterns, there are relatively few studies that have examined how the influence of linguistic experience on vowel perception changes as a function of speaker age. The majority of studies have focused on understanding adults' perception of vowels produced only by adult speakers of two languages or synthesized vowels. Moreover, most of the previous studies have related adults' perceptual judgments to the raw F1 and F2 values of vowels. No attempt has been made to understand the relationship between listeners' perception and the location of each vowel in a speaker-specific vowel space.

Aim of the Current Study

The present study is different from previous studies in several respects. First, this study investigates adult listeners' perception of vowels produced by speakers of different age groups. Second, this study relates listeners' vowel goodness judgment to transformed values of each vowel, represented by radius and degree values, instead of raw F1 and F2 values. The advantage of using radius and degree values is that it allows a comparison of listeners' vowel goodness judgments as a function of speaker age without being complicated by influences coming from different vocal tract sizes of speakers. Finally, this study uses native speakers and listeners of the larger data set with several languages. The vowels produced by monolingual children and adults of Cantonese, English, Greek, Japanese and Korean were judged by native adult speakers of American English, Greek, and Korean.

The following research questions and hypotheses were proposed. First, how do listeners map both native and non-native vowels onto their native vowel categories? What influences this perceptual mapping? The null hypothesis is that listeners will categorize vowels of both native and non-native languages into the same vowel category of similar phonetic features. In other words, during categorization, what matters is the broader conception of vowel categories but not the fine-grained phonetic details. The alternative hypothesis is that listeners' categorization of vowels will be based on the age and native language of the speakers who produce the stimuli. That is, the alternative hypothesis proposes that even during a categorization task, listeners attend to language-specific, fine phonetic details. A second research question is how listeners will rate the native and non-native vowels that are categorized into the same native vowel category. The hypothesis is that listeners will rate the native vowels as "better" examples than non-native vowels. If, however, no such effect is found, this will suggest that listeners ignore the language-specific, fine phonetic detail once they determine the category of vowel sounds. A third research question is how listeners will perceive vowels produced by 2-year-olds, 5-year-olds and adults of the same native language. Given the results from the previous study in which many significant differences were observed between the vowel productions of 2-

year-olds and those of adults, it is hypothesized that vowels produced by this youngest age group will not be rated as highly as those produced by 5-year-olds and adults.

METHODS

Participants

Twenty native speakers of American English, Greek, and Korean, ranging in age from 18 to 30- years old, participated in this study. All English-speaking listeners were recruited and tested in Ohio, Columbus, Greek-speaking listeners in Thessaloniki, Greece, and Korean-speaking listeners in Seoul, Korea. These were the same cities in which the three sets of vowel stimuli had been collected. All participants reported no history of speech, language, or hearing impairment.

Stimuli

There were two sets of stimuli: one set included word-initial consonant-vowel (CV) sequences extracted from real words produced by 2-year-olds, 5-year-olds, and adults of five languages, Cantonese, English, Greek, Japanese and Korean, and second set included vowel-only sounds excised and resynthesized from these CV sequences. The stimuli used in this perception experiment were selected from the *paidologos* database (Edwards & Beckman, 2008; Beckman & Edwards, 2009). The target vowels had all been coded as either correct or incorrect based on the judgment of a native speaker who was a trained phonetician. As has been described in the previous chapter, some similar English and Cantonese vowels were treated as vowels of the same vowel category for coding accuracy. For example, both English /i/ and /ɪ/ vowels were collapsed into the /i/

category, /ɑ/, /ʌ/, and /ɔ/ into the /ɑ/ category, and /u/ and /ʊ/ into the /u/ category. For Cantonese, /a/ or /ɐ/ were collapsed into the /a/ category. In this study, only vowels that were transcribed as correct were used. All adult productions were assumed to be correct and were not transcribed. The vowels and consonants in the CV sequences were the vowels /a/, /e/, /i/, /o/, and /u/ paired with velar and alveolar osbstruents. The preceding consonant environment was velar (/k/ or /g/) and alveolar (/t/. /d/, /s/, or /ʃ/). A summary of the number of stimuli used in this study is provided in Table 3.1. For English, only tense ([i], [u], and [e]) vowels were included; in Japanese, both long and short allophones of /i/ and /u/ vowels were included, as vowel length does not influence vowel quality (Keating & Huffman, 1984). The CV sequences were extracted from the real words. At least 10 ms prior to the consonant burst (or onset of frication) at the onset of each CV sequence and at least 10 ms after the end of the second format frequency in the vowel at the offset of each CV sequence were selected. The length of the CV stimuli varied depending on how each speaker produced the target words (mean = 250 ms, SD = 108 ms). The vowel-only stimuli were resynthesized using an overlap-add synthesis method (Moulines & Charpentier, 1990) in *Praat* and had a uniform 250ms duration. The pitch of the selected part was stretched over the re-synthesized signal so that the pitch of each speaker was maintained. In the current study, however, the focus will be on describing listeners' responses to CV sequences. The reason why the results for vowel-only stimuli were not discussed in the current study is described in the discussion section.

Table 3.1 Number of vowel stimuli by language and vowel type used for the perception experiment

Language	Age	/a/	/e/	/i/	/o/	/u/	TOTAL
Cantonese	2yr	4	4	4	4 [a]	4	20
	5yr	4	4	4	4	4	20
	Adults	4	4	4	4	4	20
English	2yr	4	4	2 [b]	4	4	18
	5yr	4	4	4	4	4	21
	Adults	4	4	3	4	4	18
Greek	2yr	4	4	4	4	4	20
	5yr	4	4	4	4	4	20
	Adults	4	4	4	4	4	20
Japanese	2yr	4	4 [c]	4	6	4	22
	5yr	4	6	4	6	6	16
	Adults	4	6	4	6	6	16
Korean	2yr	4	4	4	4	4	20
	5yr	4	4	4	4	4	20
	Adults	4	4	4	4	4	20
	TOTAL	60	64	57	66	65	**312**

[a] Both Cantonese [o:] and [ɔ:] were used for /o/ category. [b] 2-year-olds had different numbers of [i] vowel stimuli because there were only two ‘correct’ vowels with the same consonant environment as adults and 5-year-olds. [c] Both long and short vowels were included. 2-year-olds had different numbers of vowels than adults and 5-year-olds because of the limited number of ‘correct’ vowels.

Procedure

The perception experiment was programmed in *e-prime* and allowed listeners to progress through the experiment at their own pace and to record their responses with mouse clicks on the computer screen. All experiments were conducted in a quiet room.

Prior to the actual experiment, listeners were given instructions and audio examples. The audio examples used in the instruction session were not used in the real experiment. Instructions were given visually on the computer screen in the listener's native language. Audio stimuli were played with a mouse click and listeners heard them over Sennheiser headphones (HD 280 professional). Listeners were told that the purpose of the study was to examine how people perceive vowels and that their task was to identify the vowels from the stimuli they heard. The two blocks (CV and V-only sequences) were counter-balanced and each was preceded by a brief practice block, specific to that stimuli type. Within each stimuli-block, the stimuli were blocked by age (adults, 5-year-olds and 2-year-olds), and each block was presented randomly for each listener. Each of the three age blocks included all five vowels from five languages (English, Cantonese, Greek, Korean, and Japanese). Vowels of all five languages were presented randomly within each age block. Listeners were not informed that they would be listening to vowels of different languages or of different age groups. For both the CV and the V-only blocks, listeners were asked to perform the same two tasks, 1) vowel categorization and 2) goodness rating. In the vowel categorization task, listeners were asked to categorize each stimulus into one of their native vowel categories displayed on a computer screen after the stimulus item was played. For English listeners, 12 keywords containing the 12 target vowels of English were given as response options using the Latin alphabet. For Greek listeners, the 5 vowels of Greek were given as response options using the Greek alphabet and for Korean listeners, the 7 vowels of Korean were given as response options using Hangul (the Korean writing system). Figure 3.1 illustrates these response options for the three languages. Keywords were used as response options only for English vowels because the English writing system, unlike the Greek and Korean systems, does not have a one-to-one grapheme-to-phoneme mapping. The second experimental task was a rating task. Listeners were asked to judge vowel category goodness by clicking on a line where one endpoint was labeled as ‘very good example’ and the other endpoint was labeled as ‘very bad example’. Once the subject had finished responding to both the categorization and goodness rating tasks, the next stimulus was presented after a mouse click by the subject. Thus, the experiment was self-paced.

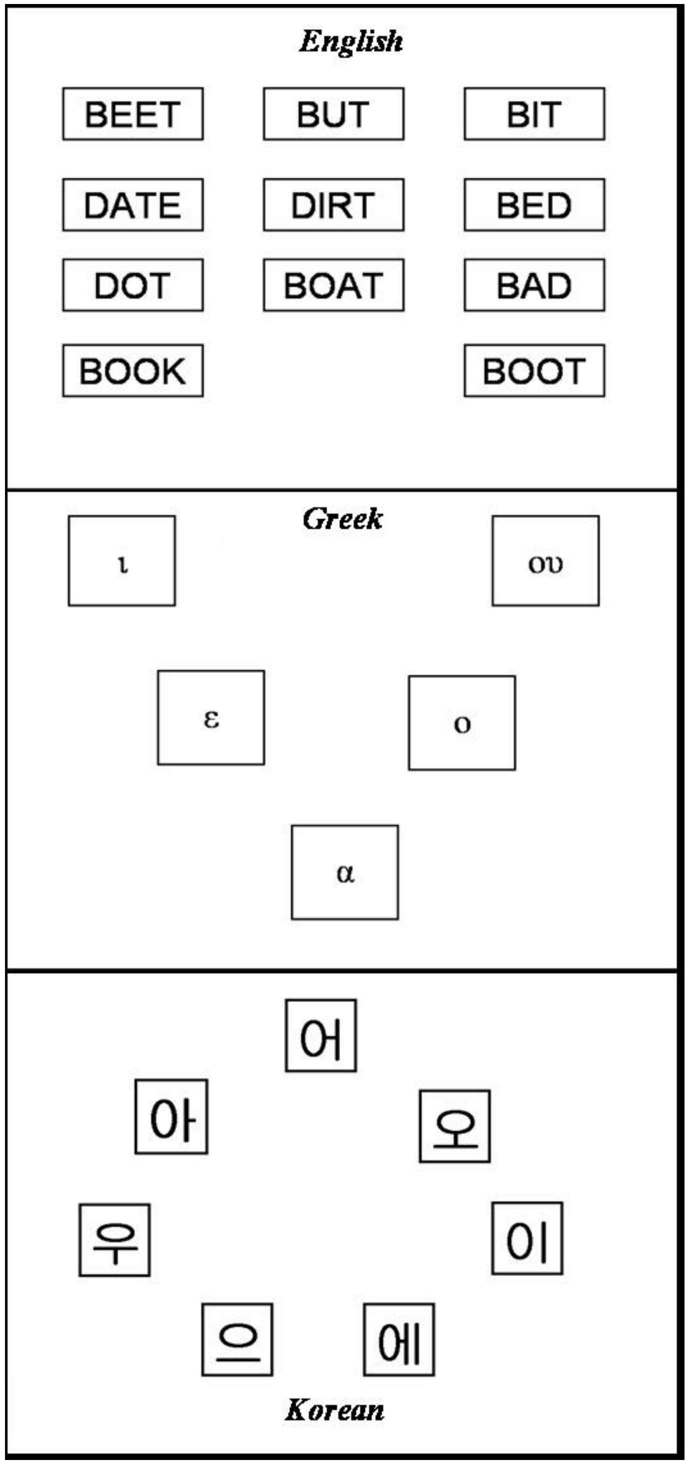

Figure 3.1 Examples of the response screen (with native vowel categories) as presented to listeners of each language, English, Greek, and Korean, for the vowel categorization task.

Acoustic Analysis

F1 and F2 values were measured at the vowel temporal midpoint of each CV sequence using LPC solution in Praat (version 5.0.29) (Boersma & Weenink, 2006). The size of the analysis window was 25 ms and the dynamic range was set at 30 dB. First, vowel duration was measured from the vowel onset, the time at which the first clear glottal pulse was observed, to the vowel offset, the time at which the final glottal pulse began to fade out. Once the duration was calculated, the temporal midpoint was determined and F1 and F2 values were extracted. F1 by F2 scatter plots are shown in Figure 3.2.

It can be observed in Figure 3.2 that the acoustic realization of these five shared vowels differs by language. While Cantonese, Greek, and Korean vowels produced by adults showed relatively good separation across vowel categories, those of English and Japanese showed some overlap. Korean vowels are relatively well separated among vowel categories, but there is some overlap between /u/ and /o/. The vowels that showed the greatest acoustic overlap were the Japanese vowels, where /i/, /e/ and /u/ were all clustered together around the same F1 region. The acoustic characteristics of children's vowel productions were similar to those of adults, except that the children's productions were more variable.

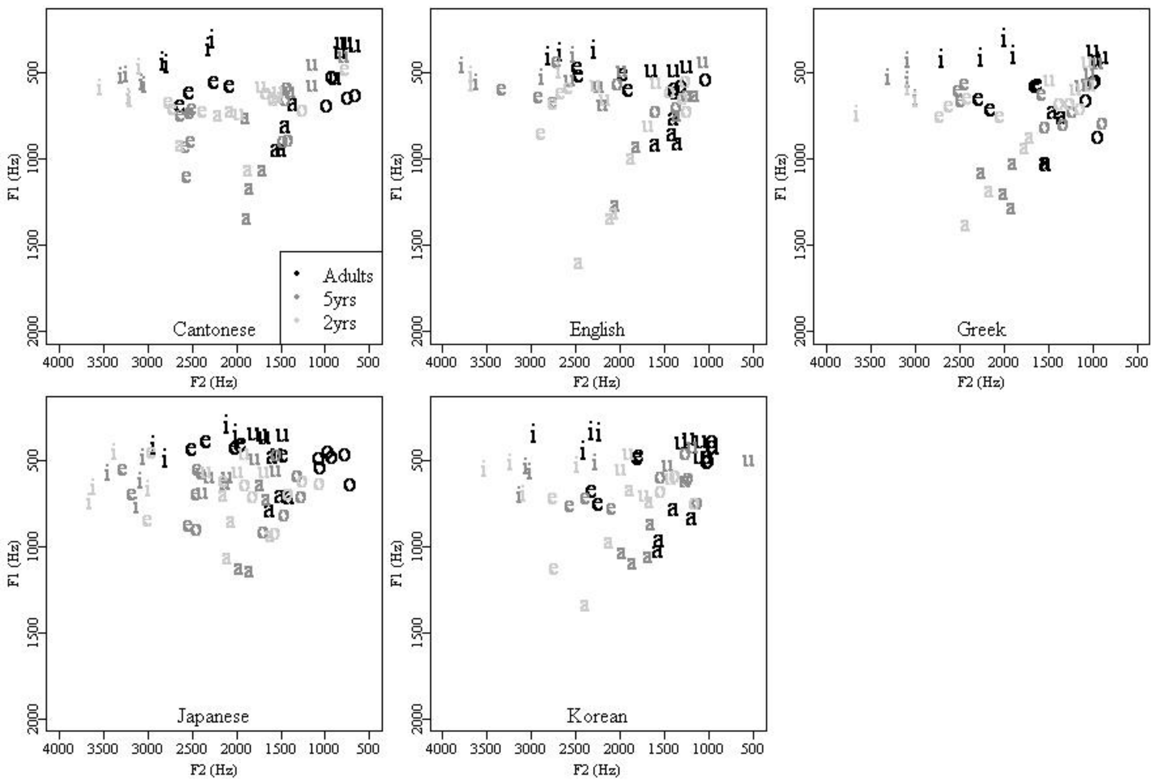

Figure 3.2 Scatter plots of vowel stimuli used in the perception experiment by language and age for both adults (black) and children (5-year-olds: dark grey; 2-year-olds: grey). X-axis represents F2 (Hz) and y-axis represents F1 (Hz). Each point represents the production of a single vowel by a single speaker.

In order to relate the acoustic characteristics of vowels across speakers with different-sized vocal tracts to the results of the perception experiments, radius and degree values for each stimulus item were calculated based on a speaker-specific vowel space. The procedure includes calculating the distance from the speaker-specific centroid (radius) and the angular displacement from the x-axis (degree) after the raw F1 and F2 values of each data point has been log-transformed and expressed relative to the speaker-specific centroid, which is the mean log F1 and log F2 of each of the five vowels. For more detail on this procedure, please see Chapter 2.

RESULTS

The results for the two stimuli blocks, CV sequences and V-only sequences were similar in many respects, except that the results for the V-only sequences were much more variable. Therefore, in this section the focus will be only on results from the CV stimuli. It is possible that listener judgments were influenced by the native or non-native consonants preceding the vowel in these CV tokens. This limitation will be considered in the discussion.

Task 1: Vowel Categorization

To analyze the results of the vowel categorization task, an emphasis was placed on whether or not there was a match between the target vowel and the listener's response. Statistical analysis was not employed for the vowel categorization results because the main purpose of the current study was examining whether or not listeners rate both native and non-native vowels categorized into the same vowel category. In this section, therefore, the categorization pattern of vowels of each language will be discussed descriptively in order to provide a general picture of how listeners categorize native versus non-native vowels of the same category.

Vowel categorization responses for vowels produced by adults are depicted in Figures 3.3 to 3.5 for each listener group (Greek, English, and Korean listeners, respectively). First, the results from Greek listeners will be described since they showed the most consistent and straightforward patterns as illustrated in Figure 3.3. As expected, Greek listeners categorized both native and non-native /a, i, e, o, u/ vowels consistently as their Greek counterparts /a, i, e, o, u/, with the exception of Korean /o/ being assimilated into Greek /u/. This pattern was predicted since Korean /o/ is more acoustically similar to Greek /u/ than to Greek /o/ as can be seen in Figure 3.2.

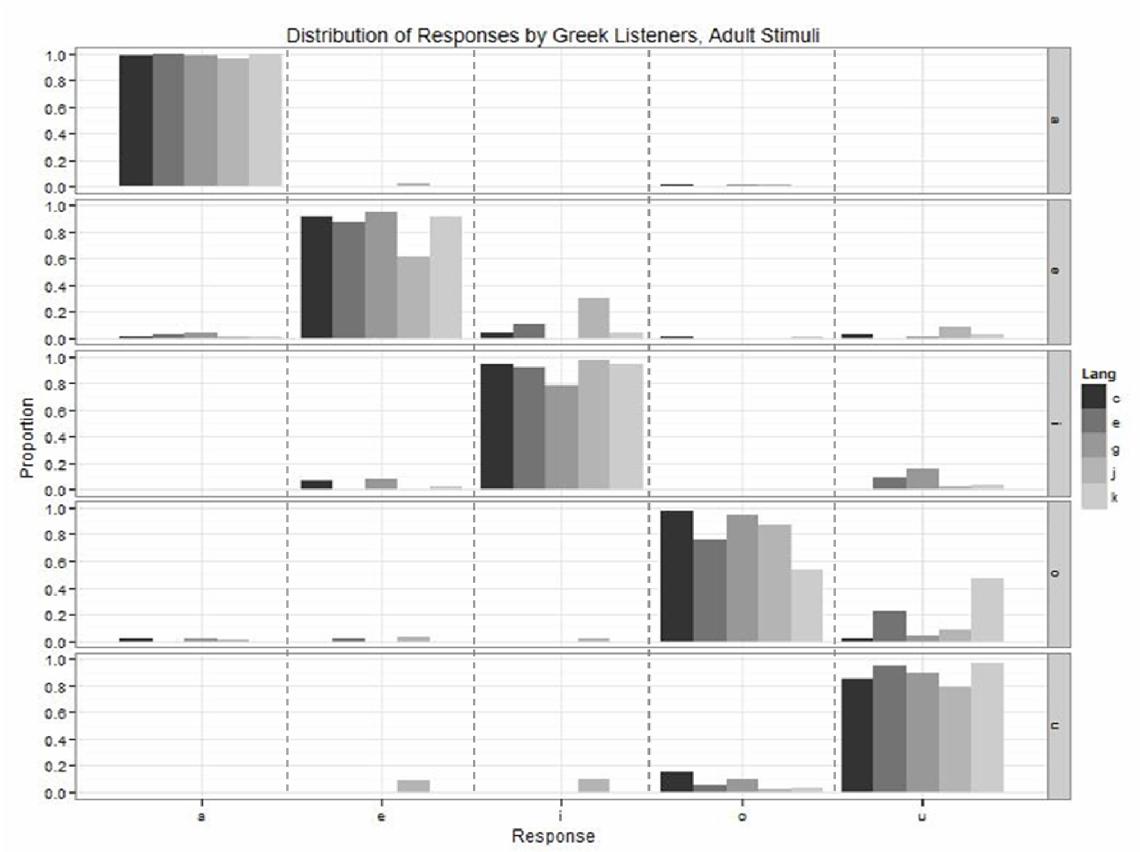

Figure 3.3 Bar graph of Greek listeners' categorization results for both native and non-native vowels produced by adult speakers of five languages. The x-axis shows the vowel response categories of Greek (/a/, /e/, /i/, /o/, and /u/). Right part of the y-axis represents the percentage match between transcribed vowel categories and the listeners' response. The left part of the y-axis represents the vowel categories of the stimuli. The vowels of each language were depicted with different bar colors: Cantonese in black, English in darkest gray, Greek in dark grey, Japanese in light grey, and Korean in the lightest grey. Each row shows the categorization result for each of the target vowel stimuli, /a/, /e/, /i/, /o/, and /u/, from top to bottom.

English listeners, as Figure 3.4 shows, categorized vowels of all five languages mostly into the same vowel category as the stimulus target. That is, /a/ vowels of all languages were categorized as English [ɑ] or [ʌ], /e/ as either [e] or [ɛ], /i/ primarily as [i], but also as [ɪ], and /o/ primarily as [o], but also as [u], and /u/ as either [u] or [ʊ]. The two vowels of English with the most differences in categorization between native and non-native vowels were /o/ and /u/. The vowel /o/ produced by Korean speakers was mostly categorized as [u]. Again, Korean /o/ is more acoustically similar to English /u/ than to English /o/. For /u/, the categorization pattern also differed across languages. Only 22% of /u/'s produced by Cantonese speakers were categorized as [u] by English listeners; most often they were categorized as [ʊ]. In contrast, 54% of /u/'s produced by Greek speakers were categorized by English listeners as [u]; only 29% of them were categorized as [ʊ]. Korean listeners also categorized both native and non-native vowels into the same vowel category as the stimulus target or as a near neighbor. The two noticeable confusion patterns found for Korean listeners were that Cantonese /o/ was assimilated into the Korean /ʌ/ category, and Japanese /u/ was assimilated into the Korean /ɨ/ category. Given that Japanese /u/ is phonetically symbolized as [ɯ], presumably because it is produced with less lip rounding and thus has higher F2 values than typically associated with /u/, and thus shares similar F2 values with Korean /ɨ/, this confusion was expected. Other than this confusion pattern, Korean listeners consistently categorized the five /a, i, e, o, u/ vowels of all languages into comparable Korean vowel categories /a, i, e, o, u/.

The response patterns to vowels produced by 2-year-olds and 5-year-olds of each of the five languages were similar to the response patterns to vowels produced by adults.

The similarity between response patterns to vowels produced by 5-year-olds and adults was greater than between 2-year-olds and adults, probably as a result of the greater acoustic variability of the vowels produced by 2-year-olds (see Chapter 2).

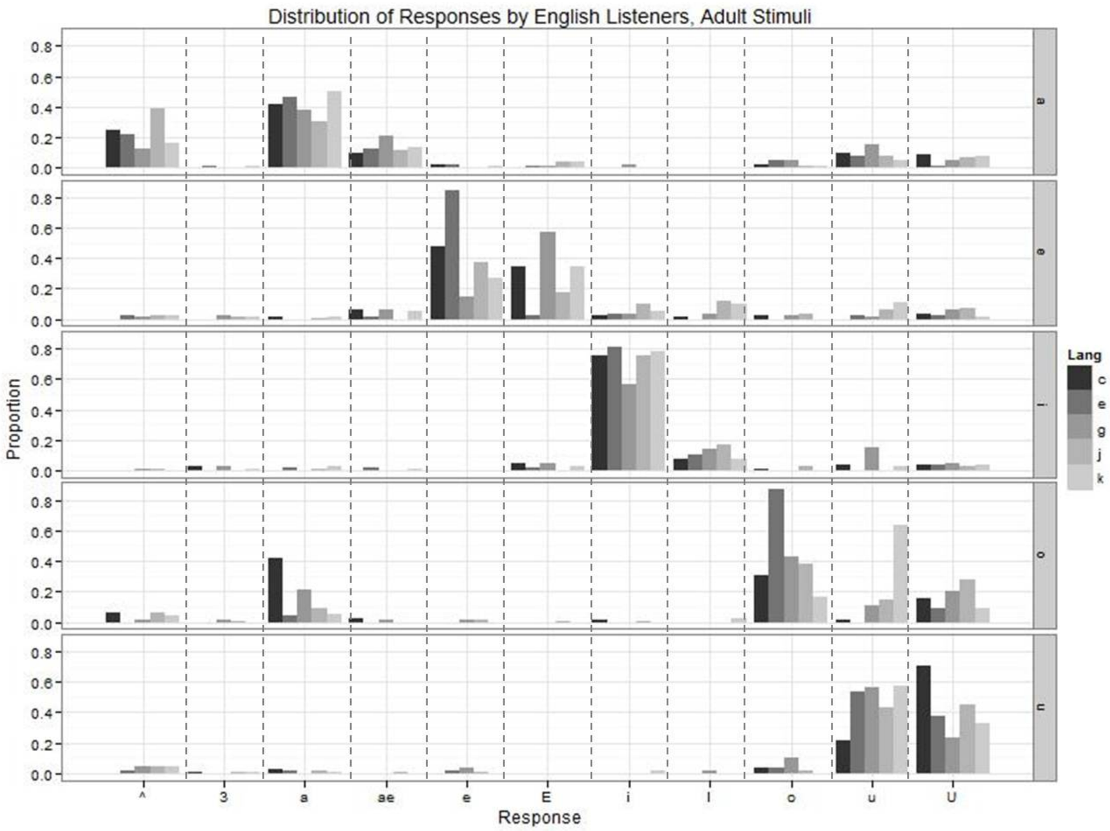

Figure 3.4 Bar graph of English listeners' categorization results for both native and non-native vowels produced by adult speakers of five languages. The x-axis shows the vowel response categories of American English ("^"= /ə/, "3" = /ɚ/, "a" = /ɑ/, "ae" = /æ/, "e" =/e/, "E" =/ɛ/, "i"=/i/, "I" =/ɪ/, "o" = /o/, "u"= /u/, and "U"= /ʊ/).

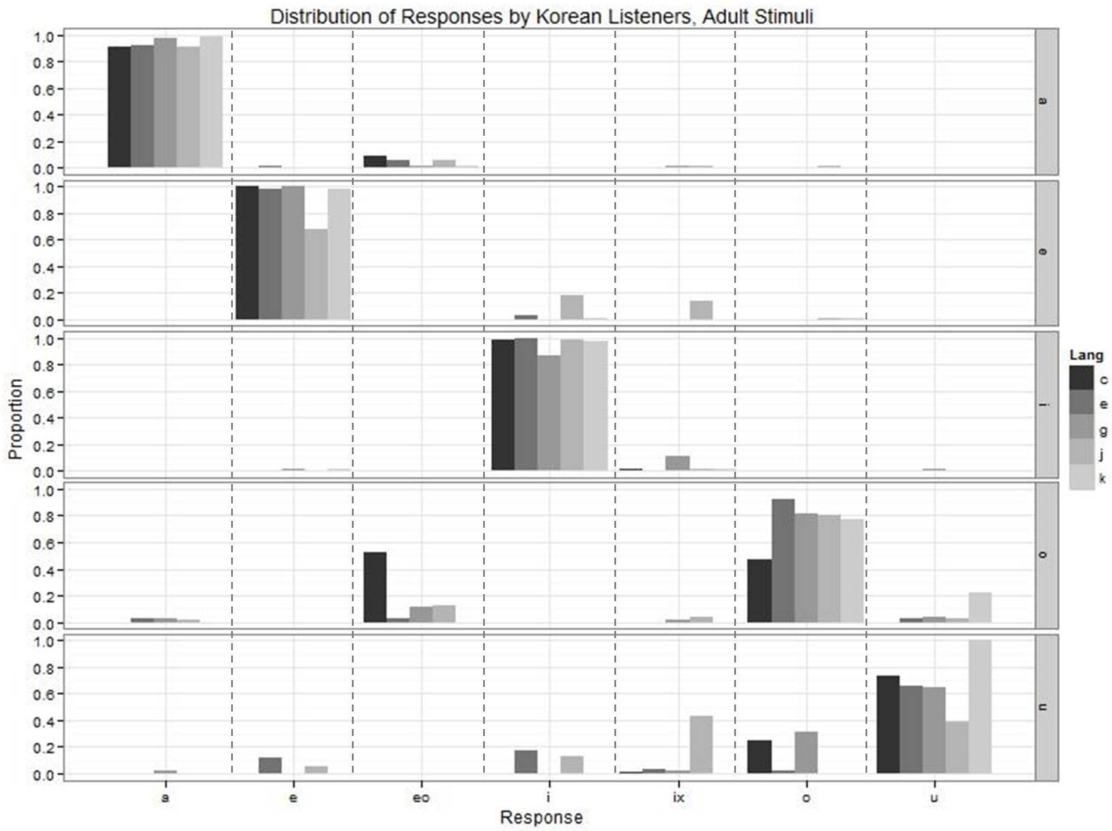

Figure 3.5 Bar graph of Korean listeners' categorization results for both native and non-native vowels produced by adult speakers of five languages. X-axis shows the vowel categories of Korean ("a" = /a/, "e" =/e/, "eo" =/ʌ/, "i"=/i/, "ix" =/ɨ/, "o" = /o/, and "u"= /u/).

In summary, the vowel categorization results showed that listeners generally categorized both native and non-native vowels as vowels of the same phonemic category as their native language, or as a vowel that was nearby in the vowel acoustic space. However, there were some cases that could not entirely be explained by acoustic similarities. For example, although Korean /o/ has very similar spectral characteristics to Korean /u/, Korean listeners categorized Korean /o/ vowels into the Korean /o/ category, but not into the Korean /u/ category. This was unlike English and Greek listeners who categorized Korean /o/ into their native /u/ category. This result indicates that listeners respond differently to their native vowels than to non-native counterparts despite acoustic similarities based on F1 and F2 values measured at the vowel midpoint. The next section examined vowel goodness ratings. The analysis was done only for vowels that were "matched" in terms of the target vowel and listeners' response categories and, as noted above, the hypothesis was that listeners would rate native vowels as better examples of a given vowel category than non-native vowels of the same category.

Task 2: Vowel Goodness Ratings

The analysis of vowel goodness ratings includes only ratings to tokens that had been categorized in the same vowel category as the stimulus item. The ratings were divided into two groups: ratings of native vowels (match between listener and speaker language) and ratings of non-native vowels (listener and speaker language did not match). This section is divided into two parts: First, the goodness rating scores of both native and non-native vowels (/a/, /e/, /i/, /o/, /u/) are presented for English, Greek, and Korean listeners, as a function of speaker language (native versus non-native), and speaker age. The results of statistical analyses, using a linear mixed-effects model, are also reported. Second, the goodness rating patterns of the three listener groups are related to acoustic properties of the vowel stimuli, defined by their radius and degree values.

a. Effect of Speaker Language, and Age on Vowel perception

Figure 3.6 shows the goodness rating judgments from the three listener groups. It can be observed that vowels produced by native speakers were rated as better vowels than those produced by non-native speakers, for all three age groups across listener groups and vowel types. Preference toward native vowels was shown most clearly for English listeners. This pattern was especially pronounced for English /e/ and /o/, and least obvious for /a/. Across vowel types, all listener groups consistently preferred native /u/ vowels over non-native counterparts. Across age groups, vowels produced by older age-groups were judged as better vowels than those produced by younger age-groups.

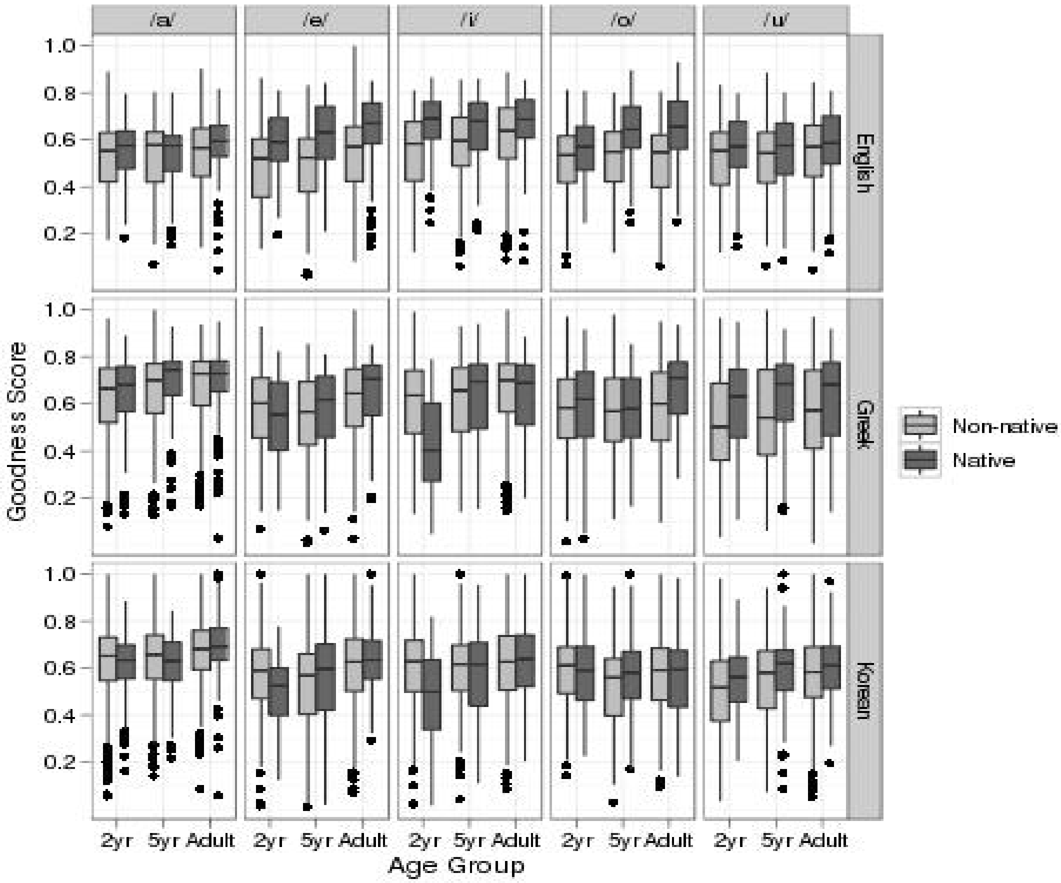

Figure 3.6 Box plots of goodness rating scores from English (top row), Greek (middle), and Korean (bottom) listener groups for both native (black) and non-native (grey) vowels produced by 2-year-olds, 5-year-olds, and adults. Each column represents each of five vowel categories of the stimuli (/a/, /e/, /i/, /o/, and /u/). The x-axis represents three speaker age groups, and the y-axis represents goodness rating score, ‘0’ being a very bad example and ‘1’ being a very good example.

To quantify the relationship between listeners' vowel perception and speaker's native language and age, a linear mixed-effects model was applied. The dependent variable was the goodness rating and the explanatory factors were nativeness (native vs. non-native), age-group (2-year-olds vs. 5-year-olds, vs. adults), and the nativeness-by-age interaction. Each listener was entered into the analysis as a random factor, to account for the repeated measurements (each listener contributed to several different responses) and different individual patterns of goodness rating. The analysis was done separately for each vowel type and listener group, resulting in a total of 15 analyses. If the initial model that included the language-by-age interaction term showed no significant interactions, a subsequent analysis was performed without the interaction term. The analyses were done using R (R Development Core Team, 2010) and its implemented packages lme4 (Bates & Maechler, 2009).

Table 3.2 gives the results for the linear mixed-effect models. The default category in these models was the adult native vowels. Although the patterns were not consistent across all 15 analyses, in general, nativeness and age-group were significant predictors of the vowel goodness rating. That is, native vowels were rated significantly higher than non-native vowels for most of the vowel categories. Also, vowels produced by adults were generally rated significantly higher than those produced by younger age groups (5-year-olds and 2-year-olds). The nativeness-by-age-group interaction was significant in some cases. For example, this interaction was significant for the Greek listeners for all vowels except /u/. In general, the significant nativeness-by-age-group interaction was because there was a stronger effect of nativeness on vowel goodness ratings for vowels produced by adults, as compared to vowels produced by the two younger age groups.

Table 3.2 The output from the linear mixed effect model for three listener groups for each vowel type and speaker age group. Significant *t*-values are in bold.

With Interaction Formula: score ~ nativeness * ageGrp + (1|Subject) No Interaction Formula: score ~ nativeness + ageGrp + (1|Subject)				
		ENGLISH LISTENERS		
No interaction		Coefficient	SE	t-value
/a/	(Intercept)	0.517	0.028	**18.485**
	Nativeness (non-native)	-0.030	0.018	-1.691
	ageGrp (2yr)	0.002	0.019	0.114
	ageGrp (5yr)	0.008	0.018	0.440
/e/	(Intercept)	0.653	0.021	**31.181**
	Nativeness (non-native)	**-0.139**	**0.014**	**-10.182**
	ageGrp (2yr)	**-0.046**	**0.015**	**-3.088**
	ageGrp (5yr)	**-0.029**	**0.014**	**-2.023**
/i/	(Intercept)	0.661	0.023	**28.636**
	Nativeness (non-native)	**-0.105**	**0.013**	**-8.072**
	ageGrp (2yr)	**-0.043**	**0.013**	**-3.370**
	ageGrp (5yr)	-0.009	0.012	-0.741
/u/	(Intercept)	0.580	0.017	**33.192**
	Nativeness (non-native)	**-0.047**	**0.012**	**-3.797**
	ageGrp (2yr)	0.010	0.012	0.821
	ageGrp (5yr)	-0.018	0.012	-1.589
With Interaction		Coefficient	SE	t-value
/o/	(Intercept)	0.683	0.022	**31.286**
	Nativeness (non-native)	**-0.223**	**0.022**	**-10.201**
	ageGrp (2yr)	**-0.140**	**0.028**	**-5.024**
	ageGrp (5yr)	-0.036	0.025	-1.451
	Nativeness(non):ageGrp(2yr)	**0.165**	**0.034**	**4.893**
	Nativeness(non):ageGrp(5yr)	**0.075**	**0.031**	**2.410**
		GREEK LISTENERS		
No interaction		Coefficient	SE	t-value
/u/	(Intercept)	0.656	0.025	**25.868**
	Nativeness (non-native)	**-0.083**	**0.015**	**-5.716**
	ageGrp (2yr)	**-0.037**	**0.014**	**-2.544**
	ageGrp (5yr)	**-0.042**	**0.013**	**-3.112**
With Interaction		Coefficient	SE	t-value
/a/	(Intercept)	0.684	0.029	**23.906**
	Nativeness (non-native)	-0.027	0.016	-1.665
	ageGrp (2yr)	**-0.079**	**0.021**	**-3.747**
	ageGrp (5yr)	0.020	0.021	0.975
	Nativeness(non):ageGrp(2yr)	0.039	0.024	1.652
	Nativeness(non):ageGrp(5yr)	**-0.051**	**0.023**	**-2.208**
/e/	(Intercept)	0.645	0.028	**23.017**
	Nativeness (non-native)	**-0.067**	**0.021**	**-3.131**

	ageGrp (2yr)	**-0.157**	**0.032**	**-4.921**
	ageGrp (5yr)	-0.038	0.029	-1.301
	Nativeness(non):ageGrp(2yr)	**0.096**	**0.036**	**2.699**
	Nativeness(non):ageGrp(5yr)	-0.037	0.033	-1.136
/i/	(Intercept)	0.626	0.029	**21.434**
	Nativeness (non-native)	0.015	0.021	0.710
	ageGrp (2yr)	**-0.185**	**0.032**	**-5.737**
	ageGrp (5yr)	0.001	0.027	0.048
	Nativeness(non):ageGrp(2yr)	**0.130**	**0.035**	**3.747**
	Nativeness(non):ageGrp(5yr)	-0.052	0.030	-1.745
/o/	(Intercept)	0.669	0.028	0.669
	Nativeness (non-native)	-0.119	0.021	-0.119
	ageGrp (2yr)	-0.115	0.030	-0.115
	ageGrp (5yr)	-0.080	0.028	-0.080
	Nativeness(non):ageGrp(2yr)	0.116	0.034	0.116
	Nativeness(non):ageGrp(5yr)	0.085	0.032	0.085
	KOREAN LISTENERS			
No interaction		Coefficient	SE	t-value
/a/	(Intercept)	0.670	0.018	**37.923**
	Nativeness (non-native)	**-0.028**	**0.010**	**-2.831**
	ageGrp (2yr)	-0.057	0.010	**-5.712**
	ageGrp (5yr)	-0.031	0.010	**-3.207**
/e/	(Intercept)	0.602	0.020	**30.693**
	Nativeness (non-native)	-0.016	0.012	-1.332
	ageGrp (2yr)	-0.085	0.012	**-7.256**
	ageGrp (5yr)	-0.074	0.011	**-7.013**
/u/	(Intercept)	0.622	0.021	**29.848**
	Nativeness (non-native)	**-0.065**	**0.013**	**-5.103**
	ageGrp (2yr)	-0.049	0.014	**-3.424**
	ageGrp (5yr)	-0.019	0.015	-1.276
With Interaction		Coefficient	SE	t-value
/i/	(Intercept)	0.597	0.022	**27.054**
	Nativeness (non-native)	-0.005	0.019	-0.247
	ageGrp (2yr)	**-0.108**	**0.025**	**-4.242**
	ageGrp (5yr)	-0.040	0.025	-1.623
	Nativeness(non):ageGrp(2yr)	**0.105**	**0.029**	**3.666**
	Nativeness(non):ageGrp(5yr)	0.036	0.028	1.302
/o/	(Intercept)	0.610	0.023	**26.938**
	Nativeness (non-native)	-0.033	0.019	-1.712
	ageGrp (2yr)	-0.053	0.027	-1.992
	ageGrp (5yr)	-0.051	0.028	-1.784
	Nativeness(non):ageGrp(2yr)	**0.070**	**0.029**	**2.388**
	Nativeness(non):ageGrp(5yr)	0.009	0.031	0.278

b. Acoustic Characteristics: radius and degree values

The second analysis focused on the relationship between the vowel acoustics and the vowel goodness ratings. The dependent variable was again vowel goodness ratings and the independent ratings were nativeness, age-group, and either stimuli radius or stimuli degree. Separate mixed-effects models for each listener group and vowel were applied and the effect of radius and degree were analyzed separately, for a total of 30 analyses (five vowels * three listener groups for radius (15) and for degree (15)).

Subjects were entered as a random factor in the model (random intercept) and radius or degree values were also entered a random factor because it was hypothesized that their effect on goodness rating judgments might vary by each listener (random slope). The other two predictor variables, nativeness and age-group, were included as fixed effects. All possible two- and three-way interactions were also included in this model. Only the results for the point vowels /i/, /a/, /u/ will be discussed as few significant relationships between the acoustic measures and the non-point vowels /e/ and /o/ were observed.

Table 3.3, 3.4, and 3.5 provide the results from the 9 linear mixed-effect models (three vowels * three listener groups) that included radius as a random factor. Figure 3.7 shows the model results for each listener group and vowel category for these analyses. As can be observed from Figure 3.7, in general, native vowels were rated more highly than non-native vowels, and vowels produced by adults were rated more highly than those produced by younger speakers across vowel types and listener groups. This result is consistent with those of the previous mixed-effects models, which did not include the acoustic variables.

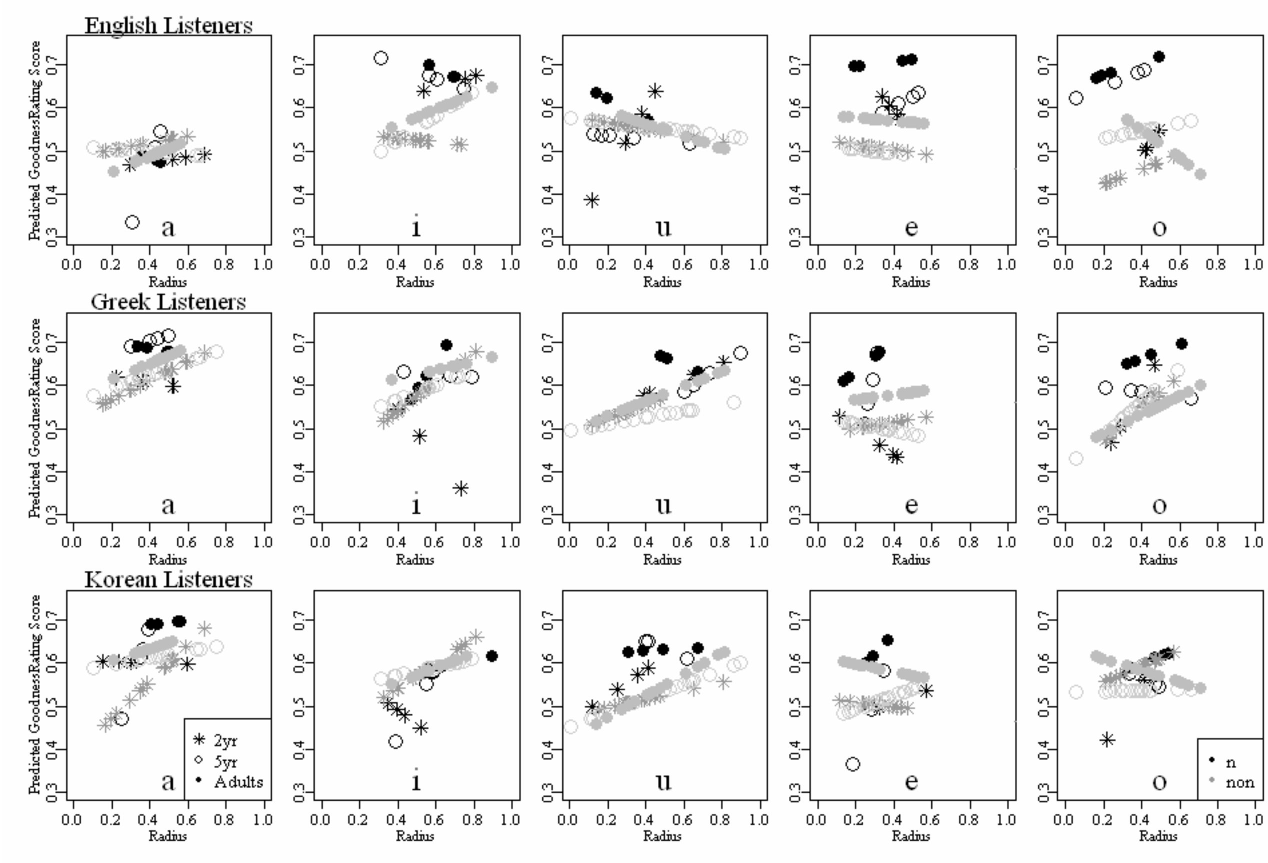

Figure 3.7 Model predicted goodness rating scores from the linear mixed effects model for the effect of radius values, speaker age, and speaker language (nativeness) for English (top row), Greek (middle), and Korean (bottom) listeners for both native (black) and non-native vowels (grey) produced by 2-year-olds (asterisk), 5-year-olds (empty circle), and adults (filled circle). The x-axis represents radius values, and the y-axis represents model predicted goodness rating scores.

Radius was a significant predictor of goodness ratings for non-native adult vowels for 5 of 9 analyses, the radius by nativeness interaction was significant for 2 of 9 analyses, and the three-way interactions among radius, nativeness, and age-groups were significant for 6 of 18 analyses. In general, the longer the radius, the higher the goodness ratings for non-native vowels, indicating that non-native vowels which were more peripheral in the vowel space tended to be perceived as better vowels than those in more central locations. This relationship between radius and goodness ratings was consistent for /u/ across listeners groups. The most clear-cut pattern can be observed in Korean listeners' perception of /u/. As can be observed in Figure 3.7, there is a clear effect of radius on goodness ratings for non-native vowels produced by speakers of all age groups. In contrast, no such effect was found for native vowels. Regardless of how peripherally located native vowels were, they were consistently rated as better vowels than non-native vowels. For native vowels produced by 2-year-olds, however, the effect of radius on goodness ratings was observed, suggesting that listeners perceive native vowels produced by 2-year-olds in a similar way as they perceive non-native vowels.

One unique pattern worth noting is the relationship between goodness ratings and radius values for non-native /u/ for English listeners. As can be seen in Figure 3.6, in contrast to Greek and Korean listeners', there was a negative relationship between the goodness ratings and radius values of non-native /u/ for English listeners. This implies that English listeners tend to rate more centrally located non-native vowels more highly than those in a more peripheral location in the vowel space. This result is in accordance with the acoustic characteristic of English /u/, whose radius values were significantly shorter than those of Greek and Korean vowels. English /u/ thus occupies a more central location in the vowel space as compared to Greek and Korean /u/.

Table 3.3 The output from the mixed-effects model for English listeners for the effect of radius, speaker age and language for the three point vowels, /a/, /i/, and /u/. Significant values are in bold[2].

Formula: score ~ radius * ageGrp * nativeness + (radius|Subject)			
/a/	Coefficient	SE	t-value
(Intercept)	0.477	0.083	5.771
Radius	0.020	0.168	0.121
Nativeness (native)	0.070	0.327	0.213
AgeGrp (2yr)	-0.028	0.087	-0.323
AgeGrp (5yr)	-0.053	0.087	-0.611
Radius:Nativeness (native)	-0.052	0.759	-0.068
Radius:AgeGrp (2yr)	0.090	0.197	0.456
Radius:AgeGrp (5yr)	0.174	0.193	0.900
Nativeness (native):AgeGrp (2yr)	-0.100	0.347	-0.287
Nativeness (native):AgeGrp (5yr)	-0.035	0.344	-0.102
Radius:Nativeness(native):AgeGrp(2yr)	0.127	0.790	0.160
Radius:Nativeness(native):AgeGrp(5yr)	-0.024	0.791	-0.030
/i/	Coefficient	SE	t-value
(Intercept)	0.358	0.051	7.001
Radius	**0.319**	**0.075**	**4.229**
Nativeness (native)	**0.671**	**0.216**	**3.104**
AgeGrp (2yr)	**0.178**	**0.066**	**2.684**
AgeGrp (5yr)	-0.012	0.063	-0.187
Radius:Nativeness (native)	**-0.906**	**0.331**	**-2.735**
Radius:AgeGrp (2yr)	**-0.381**	**0.118**	**-3.214**
Radius:AgeGrp (5yr)	0.040	0.103	0.391
Nativeness (native):AgeGrp (2yr)	**-0.741**	**0.253**	**-2.926**
Nativeness (native):AgeGrp (5yr)	-0.438	0.236	-1.852
Radius:Nativeness(native):AgeGrp(2yr)	**1.221**	**0.385**	**3.170**
Radius:Nativeness(native):AgeGrp(5yr)	0.654	0.367	1.781
/u/	Coefficient	SE	t-value
(Intercept)	0.596	0.029	20.690
Radius	**-0.126**	**0.052**	**-2.397**
Nativeness (native)	0.006	0.056	0.112
AgeGrp (2yr)	-0.027	0.035	-0.795
AgeGrp (5yr)	**-0.069**	**0.033**	**-2.081**
Radius:Nativeness (native)	0.047	0.182	0.260
Radius:AgeGrp (2yr)	0.052	0.074	0.696
Radius:AgeGrp (5yr)	0.103	0.060	1.736
Nativeness (native):AgeGrp (2yr)	-0.055	0.093	-0.597
Nativeness (native):AgeGrp (5yr)	0.019	0.067	0.287
Radius:Nativeness(native):AgeGrp(2yr)	0.233	0.271	0.859
Radius:Nativeness(native):AgeGrp(5yr)	-0.004	0.205	-0.019

---

[2] The default category was non-native adult vowels. That is, the effect of 'Radius' shows the effect of radius on goodness ratings for non-native adults vowels. The 'Nativeness (native)' shows whether or not goodness ratings for native adult vowels were significantly different from those of non-native adult vowels. The two-way interaction, 'Radius:Nativeness (native)' shows whether or not the effect of radius on goodness ratings for native adult vowels were significantly different from those of non-native adult vowels. The other two two-way interaction terms can be interpreted in a similar way. The three-way interaction, 'Radius:Nativeness (native): AgeGrp (2yr)' shows whether or not the effect of radius on goodness rating scores for native vowels produced by 2-year-olds were significantly different from those for non-native adult vowels.

Table 3.4 The output from the mixed-effects model for Greek listeners for the effect of radius, speaker age and language for the three point vowels, /a/, /i/, and /u/. Significant values are in bold.

Formula: score ~ radius * ageGrp * nativeness + (radius|Subject)			
/a/	Coefficient	SE	t-value
(Intercept)	0.579	0.042	13.678
Radius	0.181	0.092	1.971
Nativeness (native)	0.111	0.056	1.970
AgeGrp (2yr)	-0.056	0.045	-1.241
AgeGrp (5yr)	-0.017	0.043	-0.393
Radius:Nativeness (native)	-0.257	0.222	-1.159
Radius:AgeGrp (2yr)	0.043	0.101	0.424
Radius:AgeGrp (5yr)	-0.025	0.098	-0.252
Nativeness (native):AgeGrp (2yr)	-0.027	0.111	-0.241
Nativeness (native):AgeGrp (5yr)	-0.050	0.128	-0.387
Radius:Nativeness(native):AgeGrp(2yr)	-0.034	0.257	-0.134
Radius:Nativeness(native):AgeGrp(5yr)	0.233	0.301	0.774
/i/	Coefficient	SE	t-value
(Intercept)	0.577	0.057	10.082
Radius	0.098	0.087	1.126
Nativeness (native)	**-0.333**	**0.151**	**-2.204**
AgeGrp (2yr)	**-0.169**	**0.061**	**-2.777**
AgeGrp (5yr)	-0.076	0.061	-1.247
Radius:Nativeness (native)	**0.591**	**0.267**	**2.216**
Radius:AgeGrp (2yr)	**0.236**	**0.099**	**2.392**
Radius:AgeGrp (5yr)	0.068	0.099	0.693
Nativeness (native):AgeGrp (2yr)	**0.695**	**0.191**	**3.638**
Nativeness (native):AgeGrp (5yr)	**0.474**	**0.190**	**2.494**
Radius:Nativeness(native):AgeGrp(2yr)	**-1.484**	**0.331**	**-4.487**
Radius:Nativeness(native):AgeGrp(5yr)	**-0.785**	**0.320**	**-2.448**
/u/	Coefficient	SE	t-value
(Intercept)	0.334	0.077	4.340
Radius	0.362	0.117	**3.095**
Nativeness (native)	0.093	0.049	1.880
AgeGrp (2yr)	0.169	0.061	**2.777**
AgeGrp (5yr)	-0.893	0.196	**-4.560**
Radius:Nativeness (native)	-0.168	0.090	-1.869
Radius:AgeGrp (2yr)	-0.236	0.099	**-2.392**
Radius:AgeGrp (5yr)	-0.221	0.164	-1.342
Nativeness (native):AgeGrp (2yr)	-0.695	0.191	**-3.638**
Nativeness (native):AgeGrp (5yr)	0.699	0.265	**2.640**
Radius:Nativeness(native):AgeGrp(2yr)	1.484	0.331	**4.487**
Radius:Nativeness(native):AgeGrp(5yr)	0.334	0.077	**4.340**

Table 3.5 The output from the mixed-effects model for Korean listeners for the effect of radius, speaker age and language for the three point vowels, /a/, /i/, and /u/. Significant values are in bold.

Formula: score ~ radius * ageGrp * nativeness + (radius|Subject)			
/a/	Coefficient	SE	t-value
(Intercept)	0.576	0.042	13.885
Radius	0.141	0.097	1.457
Nativeness (native)	0.092	0.107	0.863
AgeGrp (2yr)	**-0.190**	**0.051**	**-3.695**
AgeGrp (5yr)	0.005	0.044	0.112
Radius:Nativeness (native)	-0.090	0.222	-0.407
Radius:AgeGrp (2yr)	**0.286**	**0.114**	**2.515**
Radius:AgeGrp (5yr)	-0.065	0.101	-0.644
Nativeness (native):AgeGrp (2yr)	0.127	0.116	1.092
Nativeness (native):AgeGrp (5yr)	**-0.557**	**0.160**	**-3.482**
Radius:Nativeness(native):AgeGrp(2yr)	-0.350	0.246	-1.423
Radius:Nativeness(native):AgeGrp(5yr)	**1.440**	**0.400**	**3.602**
/i/	Coefficient	SE	t-value
(Intercept)	0.490	0.054	9.034
Radius	**0.166**	**0.081**	**2.060**
Nativeness (native)	0.045	0.101	0.447
AgeGrp (2yr)	-0.064	0.063	-1.007
AgeGrp (5yr)	0.046	0.061	0.742
Radius:Nativeness (native)	-0.074	0.151	-0.489
Radius:AgeGrp (2yr)	0.123	0.104	1.180
Radius:AgeGrp (5yr)	-0.073	0.100	-0.733
Nativeness (native):AgeGrp (2yr)	0.152	0.189	0.804
Nativeness (native):AgeGrp (5yr)	**-0.481**	**0.207**	**-2.325**
Radius:Nativeness(native):AgeGrp(2yr)	-0.548	0.416	-1.318
Radius:Nativeness(native):AgeGrp(5yr)	**0.804**	**0.356**	**2.257**
/u/	Coefficient	SE	t-value
(Intercept)	0.425	0.037	11.628
Radius	**0.248**	**0.056**	**4.470**
Nativeness (native)	**0.194**	**0.066**	**2.947**
AgeGrp (2yr)	0.052	0.044	1.184
AgeGrp (5yr)	0.026	0.044	0.591
Radius:Nativeness (native)	-0.226	0.133	-1.702
Radius:AgeGrp (2yr)	-0.150	0.091	-1.656
Radius:AgeGrp (5yr)	-0.082	0.080	-1.030
Nativeness (native):AgeGrp (2yr)	**-0.209**	**0.091**	**-2.297**
Nativeness (native):AgeGrp (5yr)	0.086	0.084	1.026
Radius:Nativeness(native):AgeGrp(2yr)	0.435	0.228	1.906
Radius:Nativeness(native):AgeGrp(5yr)	-0.140	0.158	-0.888

Table 3.6, 3.7, and 3.8 provides the results from the 9 linear mixed-effect models that included degree values as a random factor. The default category for these models was also non-native adult vowels. Figure 3.9 shows the model results for each listener group and vowel category for these analyses. The distribution of degree values can be interpreted through the Cartesian coordinate system. The Cartesian coordinate system is used to describe the location of vowels in the vowel acoustic space. For example, vowels distributed in the first quadrant can be interpreted as vowels occupying the high back location of the vowel acoustic space, those in the second quadrant in the high front acoustic space, those in the third quadrant in the low front acoustic space, and those in the fourth quadrant in the low back acoustic space (See Figure 3.8).

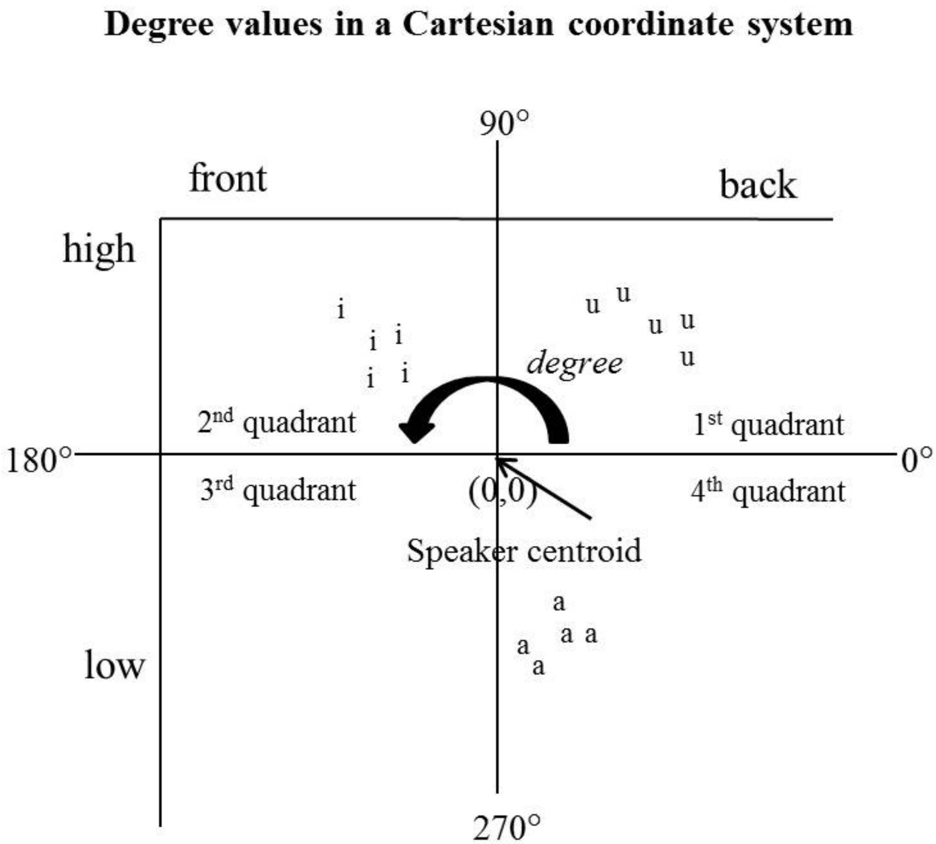

Figure 3.8 The distribution of each vowel expressed as degree values in a Cartesian coordinate system.

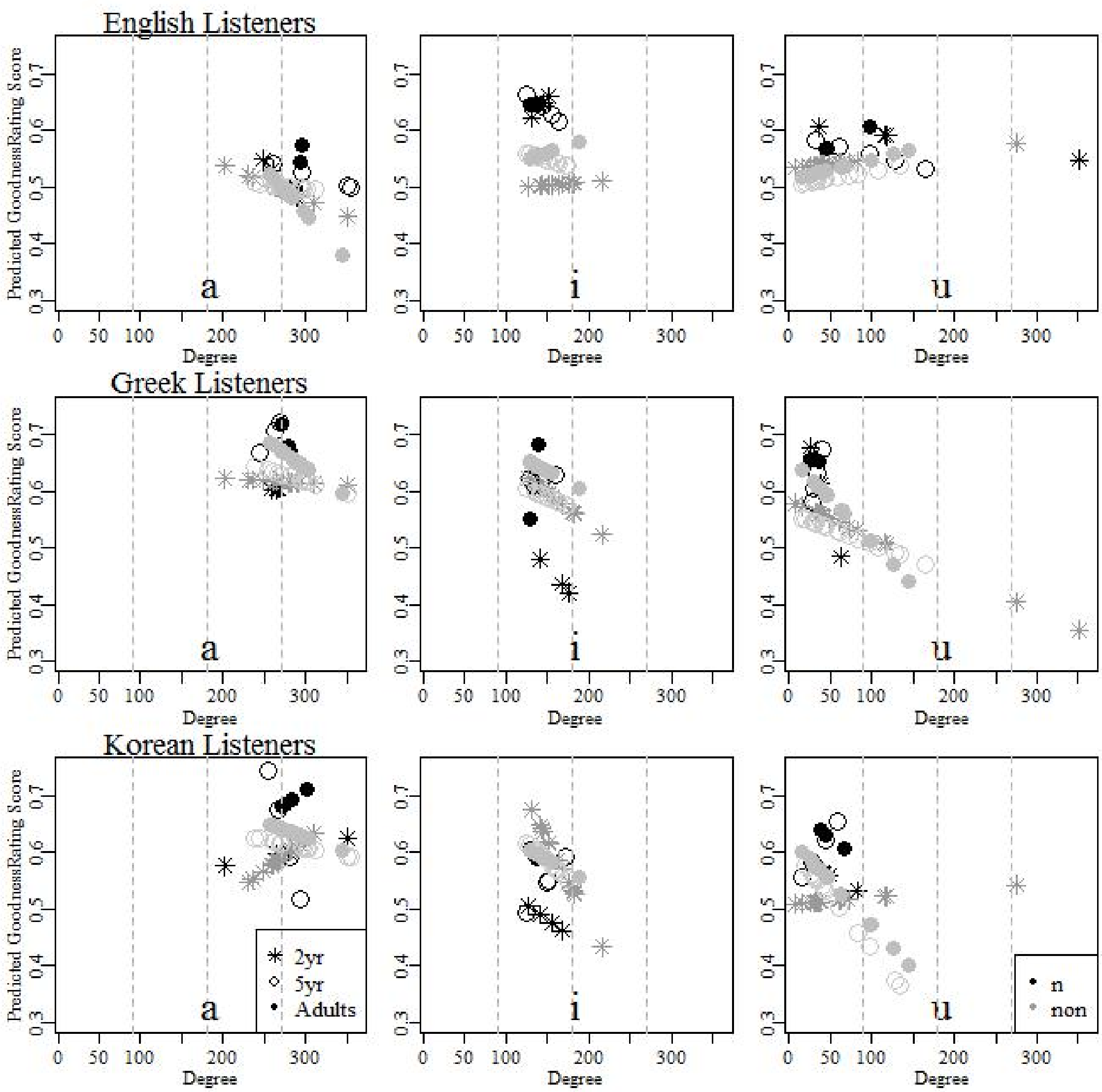

Figure 3.9 Model predicted goodness rating scores from the linear mixed effects model for the effect of degree values, speaker age, and speaker language (nativeness) for English (top row), Greek (middle), and Korean (bottom) listeners for both native (black) and non-native vowels (grey) produced by 2-year-olds (asterisk), 5-year-olds (empty circle), and adults (filled circle). The x-axis represents degree values, and the y-axis represents model predicted goodness rating scores. The grey dotted lines mark 90 degrees, 180 degrees, and 270 degrees.

In general, degree values were not as consistent a predictor of vowel goodness ratings as radius values. Only 4 of 9 analyses showed a significant effect of degree values on goodness ratings for non-native adult vowels. For English listeners, degree values were a significant predictor only for /a/ produced by non-native adult speakers (the only vowel category for which radius values were not a significant predictor). For the other two listener groups, degree values were significant predictors of goodness ratings for non-native adult vowels for 3 of 6 analyses. The degree by nativeness interaction was significant for 5 of 9 analyses. Most of the significant terms were from Greek and Korean listeners. This suggests that the effect of degree on goodness rating was different for native versus non-native vowels for Greek and Korean listeners, while no such effect was found for English listeners. The three-way interactions involving degree, nativeness, and age-group were significant for 3 of 18 analyses. A significant negative relationship between listeners' goodness ratings and degree values was also found for non-native /u/ vowels for Greek and Korean listeners. As shown in Figure 3.9, non-native /u/ vowels occupy both the first and the second quadrants, with those occupying the first quadrant rated as better examples than those occupying the second quadrant. That is, non-native /u/ vowels in the high central vowel acoustic space were rated more poorly than those associated in the high back vowel acoustic space. This negative relationship between goodness ratings and degree values of non-native /u/ was not found for English listeners. Considering that English /u/ vowels occupy both the first and the second quadrants, having the characteristic of fronted /u/ as compared to the /u/ of Greek and Korean, a different perceptual pattern for English listeners was predicted.

The effect of age-group on listeners' vowel goodness ratings was not consistent across vowel types and listener groups for either native or non-native vowels. However, the ratings of vowels produced by adult speakers were generally higher than those of the younger age groups, and this pattern was clearer for non-native vowels than for native vowels. In other words, the effect of age-group on listeners' goodness rating judgments was stronger for non-native vowels than for native vowels. In relation to the radius values, the radius by age-group interaction was significant for four of 18 analyses. Among four significant interaction terms, two vowels showed a stronger radius effect for non-native adult vowels as compared to non-native children's vowels, while the other two vowels showed a stronger effect of radius for non-native 2-year-olds vowels than those of adults'. In relation to the degree values, the degree by age-group interaction was significant for two of 18 analyses, indicating that the effect of degree on goodness rating for non-native vowels were generally similar across age groups, except for Greek /i/ and Greek /u/. For these two vowels, the stronger effect of degree was found for vowels produced by adults than 2-year-olds and 5-year-olds, respectively.

Table 3.6 The output from the mixed-effects model for English listeners for the effect of degree, speaker age and language for the three point vowels, /a/, /i/, and /u/. Significant values are in bold.

Formula: score ~ degree * ageGrp * nativeness + (degree |Subject)			
/a/	Coefficient	SE	t-value
(Intercept)	0.947	0.217	4.368
Degree	**-0.002**	**0.001**	**-2.143**
AgeGrp (2yr)	-0.286	0.246	-1.163
AgeGrp (5yr)	-0.403	0.307	-1.315
Nativeness (native)	-2.984	1.672	-1.785
Degree:Nativeness (native)	0.001	0.001	1.185
Degree:AgeGrp (2yr)	0.001	0.001	1.354
Degree:AgeGrp (5yr)	0.010	0.006	1.823
Nativeness (native):AgeGrp (2yr)	3.279	1.735	1.890
Nativeness (native):AgeGrp (5yr)	3.095	1.704	1.816
Degree:Nativeness(native):AgeGrp(2yr)	-0.011	0.006	-1.919
Degree:Nativeness(native):AgeGrp(5yr)	-0.011	0.006	-1.839
/i/	Coefficient	SE	t-value
(Intercept)	0.490	0.093	5.263
Degree	0.000	0.001	0.748
AgeGrp (2yr)	-0.001	0.130	-0.010
AgeGrp (5yr)	0.130	0.141	0.926
Nativeness (native)	0.155	0.651	0.238
Degree:Nativeness (native)	0.000	0.001	-0.440
Degree:AgeGrp (2yr)	-0.001	0.001	-1.000
Degree:AgeGrp (5yr)	0.000	0.005	-0.095
Nativeness (native):AgeGrp (2yr)	-0.252	0.747	-0.337
Nativeness (native):AgeGrp (5yr)	0.044	0.690	0.064
Degree:Nativeness(native):AgeGrp(2yr)	0.002	0.005	0.390
Degree:Nativeness(native):AgeGrp(5yr)	0.000	0.005	-0.057
/u/	Coefficient	SE	t-value
(Intercept)	0.513	0.021	24.478
Degree	0.000	0.000	1.422
AgeGrp (2yr)	0.020	0.021	0.961
AgeGrp (5yr)	-0.010	0.024	-0.424
Nativeness (native)	0.025	0.059	0.423
Degree:Nativeness (native)	0.000	0.000	-0.683
Degree:AgeGrp (2yr)	0.000	0.000	-0.275
Degree:AgeGrp (5yr)	0.000	0.001	0.394
Nativeness (native):AgeGrp (2yr)	0.056	0.069	0.810
Nativeness (native):AgeGrp (5yr)	0.067	0.072	0.930
Degree:Nativeness(native):AgeGrp(2yr)	-0.001	0.001	-0.757
Degree:Nativeness(native):AgeGrp(5yr)	-0.001	0.001	-0.988

Table 3.7 The output from the mixed-effects model for Greek listeners for the effect of degree, speaker age and language for the three point vowels, /a/, /i/, and /u/. Significant values are in bold.

Formula: score ~ degree * ageGrp * nativeness + (degree |Subject)			
/a/	Coefficient	SE	t-value
(Intercept)	0.955	0.109	8.729
Degree	**-0.001**	**0.000**	**-2.830**
AgeGrp (2yr)	**-0.315**	**0.124**	**-2.539**
AgeGrp (5yr)	-0.204	0.138	-1.478
Nativeness (native)	1.195	0.957	1.248
Degree:Nativeness (native)	**0.001**	**0.000**	**2.188**
Degree:AgeGrp (2yr)	0.001	0.000	1.242
Degree:AgeGrp (5yr)	-0.004	0.003	-1.229
Nativeness (native):AgeGrp (2yr)	-1.285	1.502	-0.855
Nativeness (native):AgeGrp (5yr)	-1.815	1.030	-1.762
Degree:Nativeness(native):AgeGrp(2yr)	0.005	0.006	0.811
Degree:Nativeness(native):AgeGrp(5yr)	0.007	0.004	1.840
/i/	Coefficient	SE	t-value
(Intercept)	0.754	0.095	7.968
Degree	-0.001	0.001	-1.236
AgeGrp (2yr)	0.012	0.112	0.104
AgeGrp (5yr)	-0.072	0.123	-0.584
Nativeness (native)	**-2.085**	**0.771**	**-2.704**
Degree:Nativeness (native)	0.000	0.001	-0.444
Degree:AgeGrp (2yr)	0.000	0.001	0.205
Degree:AgeGrp (5yr)	**0.015**	**0.006**	**2.678**
Nativeness (native):AgeGrp (2yr)	**2.039**	**0.842**	**2.423**
Nativeness (native):AgeGrp (5yr)	**2.002**	**0.826**	**2.423**
Degree:Nativeness(native):AgeGrp(2yr)	**-0.016**	**0.006**	**-2.617**
Degree:Nativeness(native):AgeGrp(5yr)	**-0.015**	**0.006**	**-2.387**
/u/	Coefficient	SE	t-value
(Intercept)	0.664	0.031	21.121
Degree	**-0.002**	**0.000**	**-4.269**
AgeGrp (2yr)	**-0.079**	**0.028**	**-2.826**
AgeGrp (5yr)	**-0.102**	**0.027**	**-3.709**
Nativeness (native)	0.002	0.157	0.010
Degree:Nativeness (native)	**0.001**	**0.000**	**2.344**
Degree:AgeGrp (2yr)	**0.001**	**0.000**	**2.490**
Degree:AgeGrp (5yr)	0.001	0.005	0.244
Nativeness (native):AgeGrp (2yr)	0.229	0.169	1.358
Nativeness (native):AgeGrp (5yr)	-0.199	0.224	-0.887
Degree:Nativeness(native):AgeGrp(2yr)	-0.006	0.005	-1.153
Degree:Nativeness(native):AgeGrp(5yr)	0.007	0.007	1.069

Table 3.8 The output from the mixed-effects model for Korean listeners for the effect of degree, speaker age and language for the three point vowels, /a/, /i/, and /u/. Significant values are in bold.

Formula: score ~ degree * ageGrp * nativeness + (degree |Subject)			
/a/	Coefficient	SE	t-value
(Intercept)	0.783	0.111	7.029
Degree	-0.001	0.000	-1.312
AgeGrp (2yr)	**-0.487**	**0.181**	**-2.684**
AgeGrp (5yr)	-0.087	0.132	-0.655
Nativeness (native)	-0.367	0.349	-1.054
Degree:Nativeness (native)	**0.002**	**0.001**	**2.439**
Degree:AgeGrp (2yr)	0.000	0.000	0.480
Degree:AgeGrp (5yr)	0.001	0.001	1.219
Nativeness (native):AgeGrp (2yr)	0.581	0.385	1.507
Nativeness (native):AgeGrp (5yr)	**1.910**	**0.573**	**3.335**
Degree:Nativeness(native):AgeGrp(2yr)	-0.002	0.001	-1.648
Degree:Nativeness(native):AgeGrp(5yr)	**-0.007**	**0.002**	**-3.444**
/i/	Coefficient	SE	t-value
(Intercept)	0.709	0.098	7.227
Degree	-0.001	0.001	-1.246
AgeGrp (2yr)	**0.341**	**0.115**	**2.964**
AgeGrp (5yr)	0.076	0.123	0.621
Nativeness (native)	0.147	0.634	0.232
Degree:Nativeness (native)	**-0.002**	**0.001**	**-2.705**
Degree:AgeGrp (2yr)	-0.001	0.001	-0.627
Degree:AgeGrp (5yr)	-0.001	0.005	-0.238
Nativeness (native):AgeGrp (2yr)	-0.555	0.670	-0.829
Nativeness (native):AgeGrp (5yr)	-0.690	0.671	-1.028
Degree:Nativeness(native):AgeGrp(2yr)	0.003	0.005	0.587
Degree:Nativeness(native):AgeGrp(5yr)	0.004	0.005	0.909
/u/	Coefficient	SE	t-value
(Intercept)	0.628	0.025	24.985
Degree	**-0.002**	**0.000**	**-4.344**
AgeGrp (2yr)	**-0.119**	**0.027**	**-4.399**
AgeGrp (5yr)	-0.013	0.032	-0.398
Nativeness (native)	0.053	0.070	0.761
Degree:Nativeness (native)	**0.002**	**0.000**	**4.072**
Degree:AgeGrp (2yr)	0.000	0.001	-0.484
Degree:AgeGrp (5yr)	0.000	0.001	0.344
Nativeness (native):AgeGrp (2yr)	0.038	0.093	0.413
Nativeness (native):AgeGrp (5yr)	-0.146	0.091	-1.618
Degree:Nativeness(native):AgeGrp(2yr)	-0.001	0.002	-0.814
Degree:Nativeness(native):AgeGrp(5yr)	0.004	0.002	1.902

DISCUSSION

In the present study, adult listeners' perception of both native and non-native vowels produced by children and adults of both native and non-native languages were examined in relation to the location of a given vowel in the vowel acoustic space. The results showed that adult listeners categorized native and non-native vowels based primarily on the spectral similarities between the vowel stimuli and the native vowel category, but less perceptual confusion occurred for native vowels than non-native vowels, and for vowels produced by adult speakers than young children. Further analysis of adult listeners' goodness judgments of native and non-native vowels within a vowel perceptual category showed that adult listeners perceived native vowels as better examples regardless of how peripherally located the given vowel was, while for non-native point vowels, those that were more peripherally located were judged as better examples. In addition, a systematic relationship between listeners' goodness rating judgment and degree values was found for non-native point vowels, while no such relationship was found for native counterparts.

These findings suggest that listeners use different perceptual cues when perceiving vowels depending on the nature of the task and vowel stimuli. The results of the vowel categorization task showed that listeners were primarily constrained by language-specific phonetic categories during phonemic-level perceptual processing. However, there was evidence of sensitivity to subphonemic fine phonetic detail in the goodness rating results. This evidence was most compelling when listeners were rating non-native vowels. Given that listeners were not told whether vowels they were listening to were produced by native or non-native speakers, this finding clearly demonstrates listeners' ability to perceive fine phonetic details of the vowel stimuli. For the point vowels that were produced by non-native speakers, listeners perceived those closer to the periphery of the vowel acoustic space as better examples of the vowel category. The reason for this perceptual preference for the three point vowels approaching the periphery of the acoustic vowel space could be explained by the patterns of 'directional asymmetries' suggested by Polka and Bohn (2003; 2010). In a series of studies, Polka and Bohn (2003; 2010) showed that infants have a perceptual preference for peripheral vowels, /i/, /a/, and /u/, which are located in the corners of the vowel spaces. The reason for this perceptual asymmetry was explained by the intrinsic characteristics of these three point vowels that make them perceptually more salient as compared to other vowels (the Dispersion-Focalization theory, Schwartz, Boe, Vallee, & Abry, 1997). Although this explanation was made to explain vowel perceptual patterns of infants, a similar pattern of perceptual asymmetries has also been documented in the adult vowel perception (e.g., Nishi *et al.*, 2008). However, the reason why such relationship did not occur for native point vowels remains unexplained. Similarly, degree values also systematically predicted listeners' goodness rating judgments for non-native point vowels. Listeners tended to rate non-native point vowels that shared similar locations in the vowel acoustic space with native counterparts as better examples than those that were not. In sum, the current study showed that acoustic characteristics of vowels, expressed here as radius and degree values, calculated from the temporal midpoint values of F1 and F2, predicted, at least to some extent, adult listeners' vowel goodness ratings for the three non-native point vowels.

An effect of speaker age on adult listeners' vowel perception was also found. Overall, vowels produced by adult speakers were categorized more systematically to the target vowel categories. In addition, vowels produced by adult speakers were generally rated as better examples than those produced by young children. This pattern was clearer for non-native vowels than for native vowels. This suggests that speaker age plays a more important role for vowel goodness judgment of non-native vowels than for native vowels. Moreover, the preference toward native vowels over non-native vowels was most clearly observed for vowels produced by adult speakers and least clearly for vowels produced by 2-year-olds. This implies that language-specific characteristics present in vowels produced by adult speakers were less distinct in vowels produced by 2-year-olds.

The potential limitation of this study, however, is that the stimuli were CV syllables. Therefore, the coarticulatory effects coming from the preceding consonant could have influenced listeners' perception of both native and non-native vowels. Also, goodness ratings could have been influenced by whether the initial consonant sounded native or non-native. While we had also included a vowel-only task, the results for this task were much more variable, perhaps because the stimuli were synthetic rather than natural speech, or because the consonant-vowel transitional information had been removed. A future study with natural vowel-only stimuli or CVC stimuli with a more carefully controlled consonant environment would be useful in determining the generalizability of these results.

This study showed that radius and degree values well predicted adult listeners' goodness judgments of non-native point vowels, but no such relationship was found for native counterparts. Moreover, neither radius nor degree values predicted the goodness rating judgment for either native or non-native /e/ and /o/ vowels. Nonetheless, native /e/ and /o/ vowels were judged as better examples than non-native counterparts by all listener groups, especially by English listeners. These results suggest that adult listeners use different perceptual cues for listening to native versus non-native vowels, as well as when listening to point versus non-point vowels. One possible explanation for this pattern is that listeners focus on the phonetic information when perceiving vowels that are not familiar to them, but they rely on some other non-phonetic factors, such as linguistic experience, when perceiving familiar vowels within the same vowel category. Another possible explanation for the different perceptual responses of adult listeners to native versus non-native vowels is that the acoustic characteristics of the vowels used in this study, defined by temporal midpoint measures of F1 and F2 values, did not fully characterize vowel acoustic characteristics. It has been reported that vowel spectral change plays an important role in vowel identification (e.g., Hillenbrand and Nearey, 1999; Nearey & Assmann, 1986; Strange, 1989). These studies showed that vowel identification error rate increased when vowel dynamic information (e.g., changing formant frequencies over time) was not present in the stimuli. Given that English /e/ and /o/ vowels are, in many dialects, phonetically diphthongized (e.g., Hillenbrand, Getty, Clark, & Wheeler, 1995), and thus are known to have greater spectral movement than the three point vowels, it is possible that English listeners heard this characteristic of English /e/ and /o/ vowels and rated them as better examples than those with reduced or different patterns of spectral movements. It can also be hypothesized that Greek and Korean /e/ and /o/ vowels have different magnitude or direction of spectral change than English vowels. Moreover, the results of the present study showed that among the three point vowels, the preference towards the native over non-native vowel was the most consistent for /u/ across listener groups. Since /u/ is the vowel that has a relatively greater magnitude of spectral change than /a/ and /i/ for both American English and Korean (Chung, Kong, & Weismer, 2010), it may be that Greek /u/ will also show relatively greater magnitude of spectral change than /a/ and /i/.

The potential effect of language-specific vowel spectral movement patterns on listeners' vowel perception will be investigated in the next chapter, by comparing the vowel trajectory patterns of five shared vowels /a, e, i, o, u/ produced by monolingual 2-year-olds, 5-year-olds and adults of American English, Greek, and Korean.

One implication of the findings of this study to general theories of speech perception is that adult listeners are capable of using different strategies, and that they flexibly change these strategies depending on the nature of the task (vowel categorization vs. goodness rating tasks), nativeness of vowels (native vs. non-native language), and vowel types (point vs. non-point vowels). This was clearly evidenced by listeners' response to vowels produced by speakers of different age groups. As compared to listeners' responses to native vowels produced by 5-year-olds or adults, which consistently contain relatively clear language-specific characteristics, listeners perceived vowels produced by 2-year-olds by using different strategies depending on how language-specific their vowels were. For vowels that have adult-like language-specific characteristics, listeners used the same strategy as when they heard native vowels produced by adults or 5-year-olds. However, when vowels produced by 2-year-olds contained minimal to no language-specific characteristics, listeners shifted to a different strategy and perceived these vowels in a similar way they perceived non-native vowels. When vowels are produced that are not clearly in the native category, such as vowels produced by non-native speakers or very young children, listeners may focus greater attention on fine phonetic detail.

In sum, the results of the current study showed that adult listeners are flexible when asked to perceive and rate vowels. Furthermore, the use of one perceptual strategy does not mean that a listener is incapable of using another strategy. Instead, as suggested by Werker and Tees (1984), it suggests that adult listeners “shift attentional focus and or processing strategies” from one task to another.

CHAPTER 3: REFERENCES

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

Best, C. (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, M. A.: York Press.

Boersma, P. & Weenink, D. (2006). Praat: doing phonetics by computer [Computer program]. Retrieved Aug, 2006, from http://www.praat.org/.

Chung, H., Kong, E-J., & Weismer, G. (2010). Vowel formant trajectory patterns for shared vowels of American English and Korean. *Journal of the Korean Society of Speech Science*, *2(4)*, 67-74.

Flege, J. E., Munro, M. J., & Fox, R. A. (1994). Auditory and categorical effects on cross-language vowel perception. *Journal of the Acoustical Society of America*, 95, 3623-3641.

Flege, J. E. (1992). Speech learning in a second language. In C. Ferguson, L. Menn, & C. Stoel-Gammon (Ed.), *Phonological Development: Models, Research, and Applications.* Timonium, MD: York Press.

Flege, J. E. (1995). Second language speech learning: Theory, finding, and problems. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross-language Research* (233-277). Timonium, MD: York Press.

Hillenbrand, J. & Gayvert, R. T. (1993). Identification of steady-state vowels synthesized from the Peterson and Barney measurements. *Journal of the Acoustical Society of America, 94 (2),* 668-674.

Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America, 97*, 3099-3111.

Hillenbrand, J. & Nearey, T. M. (1999). Identification of resynthesized /hVd/ utterances: Effects of formant contour. *Journal of the Acoustical Society of America, 105 (6),* 3509-3523.

Keating, P. A. & Huffman, M. K. (1984). Vowel variation in Japanese. *Phonetica. 41*, 191-207.

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., Lindblom, B. (1992). Linguistic Experience Alters Phonetic Perception in Infants by 6 Months of Age. Science, Vol. 255. 606-608.

Moulines, E. & Charpentier, F. (1990). Pitch Synchronous Waveform Processing Techniques for Text-to-Speech Synthesis Using Diphones. *Speech Communication.* 9, 453-467, 1990.

Nearey, T. M. & Assmann, P. F. (1986). Modeling the role of inherent spectral change in vowel identification. *Journal of the Acoustical Society of America, 80 (5),* 1297-1308.

Nishi, K., Strange, W., Akahane-Yamada, R., Kubo, R., & Trent-Brown, S. A. (2008). Acoustic and perceptual similarity of Japanese and American English vowels. *Journal of the Acoustical Society of America. 124 (1),* 576-588.

Polka, L. and Werker, J. F. (1994). Developmental Changes in perception of nonnative vowel contrasts. *Journal of Experimental Psychology.* 20 (2), 421-435.

Polka, L., & Bohn, O.-S. (2003). Asymmetries in vowel perception. *Speech Communication. 41*, 221-231.

Polka, L., & Bohn, O.-S. (2010). Natural Referent Vowel (NR1V) framework: An emerging view of early phonetic development. *Journal of Phonetics*. Doi:10.1016/j.wocn.2010.08.007

R Development Core Team (2010). *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0.

Schwartz, J. –L., Boe, L.-J., Vallee, N., & Abry, C. (1997). The Dispersion-Focalization Theory of vowel systems. *Journal of Phonetics, 25*, 255-286.

Stevens, K. , Libermann, A., Studdert-Kennedy, M., & Ohman, S. (1969). Crosslanguage study of vowel perception. *Language and Speech, 12*, 1-23.

Strange, W. (1989). Dynamic specification of coarticulated vowels spoken in sentence context. *Journal of the Acoustical Society of America, 85(5),* 2135-2153.

Strange, W., Akahana-Yamada, R., Kubo, R., Trent, S. A., Nishi, K., and Jenkins. J. J. (1998). Perceptual Assimilation of American English vowels by Japanese listeners. *Journal of Phonetics. 26*, 311-344.

Werker, J. F., & Tees, R. C. (1984). Phonemic and phonetic factors in adult cross-language speech perception. *Journal of the Acoustical Society of America*, 75(6), 1866-1878.

CHAPTER 3: APPENDIX

Appendix A: Confusion Matrices for the Vowel Assimilation

English Listeners

			Response Category										
Lang	Age	Target V	ə	ɚ	ɑ	æ	e	ɛ	i	ɪ	o	u	ʊ
c	2yr	a	**34**	3	**11**	**11**	1	4	0	3	9	9	15
		e	8	5	1	1	**9**	**35**	0	23	8	5	5
		i	4	6	1	0	12	3	**41**	**22**	1	6	4
		o	9	4	6	0	0	0	0	1	**41**	5	34
		u	4	1	1	0	0	0	0	1	6	**10**	**76**
	5yr	a	**13**	0	**39**	**21**	0	0	1	0	5	11	10
		e	5	8	3	20	**11**	**38**	0	9	4	0	4
		i	0	1	0	3	0	1	**54**	**26**	1	5	9
		o	14	0	14	1	1	0	0	0	**44**	13	14
		u	1	1	4	3	0	0	3	6	4	**13**	**66**
	Adults	a	**25**	0	**41**	**10**	3	0	0	0	3	10	9
		e	0	0	1	6	**48**	**34**	3	1	3	0	4
		i	0	3	0	0	0	5	**76**	**8**	1	4	4
		o	6	0	42	3	0	0	1	0	**31**	1	15
		u	0	1	3	0	0	0	0	0	4	**22**	**71**
e	2yr	a	**23**	0	**35**	**18**	6	0	0	1	8	5	4
		e	0	4	1	4	**53**	**12**	9	12	1	3	3
		i	3	2	2	0	0	5	**77**	**7**	0	3	2
		o	8	3	8	1	0	0	0	0	**53**	16	13
		u	10	2	3	0	0	0	0	1	9	**55**	**20**
	5yr	a	**14**	1	**38**	**20**	0	0	0	3	14	0	10
		e	0	1	1	0	**50**	**5**	20	15	0	3	5
		i	1	3	0	0	4	1	**71**	**5**	0	10	5
		o	1	1	5	0	0	3	1	0	**78**	3	9
		u	4	3	2	0	2	1	0	2	2	**63**	**21**
	Adults	a	**22**	1	**47**	**13**	3	1	0	0	5	8	1
		e	3	0	0	1	**85**	**3**	4	0	0	3	3
		i	0	0	2	2	0	2	**81**	**10**	0	0	3
		o	0	0	4	0	0	0	0	0	**87**	0	9
		u	2	0	2	0	2	0	0	0	3	**53**	**38**
g	2yr	a	**23**	0	**26**	**18**	6	1	1	0	13	1	10
		e	8	4	1	6	**9**	**29**	0	20	5	9	9
		i	4	8	4	0	5	1	**22**	**29**	0	16	11
		o	1	3	11	0	1	0	0	0	**28**	24	33
		u	5	0	1	0	1	1	0	3	9	**46**	**34**

	5yr	a	**5**	1	**28**	**23**	1	1	1	1	10	18	11
		e	5	11	3	8	**6**	**34**	1	8	6	5	14
		i	0	1	0	1	3	3	**65**	**13**	0	13	3
		o	14	0	13	3	1	1	0	0	**29**	14	26
		u	5	3	5	0	0	0	1	0	18	**42**	**26**
	Adults	a	**13**	0	**38**	**21**	0	1	3	0	5	15	5
		e	1	3	0	6	**15**	**57**	4	4	3	1	6
		i	1	3	0	0	0	5	**57**	**14**	0	15	5
		o	1	1	21	1	1	0	0	0	**43**	11	20
		u	5	0	0	0	3	0	0	1	8	**54**	**29**
j	2yr	a	**49**	0	**17**	**5**	1	0	0	1	10	8	9
		e	8	9	4	12	**14**	**29**	4	8	5	0	8
		i	5	0	3	0	5	4	**55**	**13**	1	12	3
		o	23	2	13	4	0	0	0	0	**28**	9	21
		u	5	0	0	0	0	0	0	1	0	**37**	**57**
	5yr	a	**51**	3	**16**	**3**	0	0	1	0	6	11	9
		e	8	3	3	3	**1**	**29**	7	18	8	8	13
		i	1	1	0	1	5	4	**39**	**23**	0	14	11
		o	8	1	13	3	0	2	0	0	**35**	26	14
		u	13	2	0	0	1	0	0	10	1	**45**	**28**
	Adults	a	**38**	0	**31**	**12**	0	4	0	0	1	8	6
		e	3	2	1	0	**38**	**18**	10	12	3	6	8
		i	1	0	1	0	0	0	**75**	**17**	3	0	3
		o	6	1	9	0	2	1	1	0	**38**	15	27
		u	4	1	2	1	1	0	0	0	2	**44**	**46**
k	2yr	a	**24**	1	**46**	**10**	1	1	0	1	6	1	8
		e	6	1	1	4	**30**	**26**	3	16	6	4	3
		i	6	6	0	0	9	9	**18**	**44**	3	0	5
		o	6	4	5	1	1	8	0	0	**42**	14	19
		u	4	1	3	1	0	1	0	0	4	**15**	**71**
	5yr	a	**35**	0	**25**	**6**	0	3	1	0	11	10	8
		e	14	5	3	11	**8**	**10**	0	15	9	23	4
		i	3	1	1	1	1	5	**43**	**16**	1	16	10
		o	4	0	10	1	0	1	0	0	**21**	53	10
		u	8	1	5	0	0	1	0	0	1	**41**	**43**
	Adults	a	**16**	1	**50**	**14**	1	4	0	0	1	5	8
		e	3	1	1	5	**28**	**35**	5	10	0	11	1
		i	0	1	3	1	0	3	**78**	**8**	0	3	4
		o	4	0	5	0	0	0	0	2	**16**	63	9
		u	5	1	1	0	0	0	2	0	0	**58**	**33**

Greek Listeners

			Response Category				
Lang	Age	Target V	a	e	i	o	u

c	2yr	a	**51**	14	4	13	19
		e	1	**65**	5	11	18
		i	0	6	**91**	0	3
		o	5	0	0	**80**	15
		u	0	3	3	50	**45**
	5yr	a	**98**	0	0	1	1
		e	15	**78**	6	0	1
		i	0	8	**91**	0	1
		o	10	0	0	**63**	28
		u	0	1	3	4	**93**
	Adults	a	**99**	0	0	1	0
		e	1	**91**	4	1	3
		i	0	6	**94**	0	0
		o	1	0	0	**98**	1
		u	0	0	0	15	**85**
e	2yr	a	**91**	1	0	8	0
		e	3	**65**	30	0	3
		i	0	2	**98**	0	0
		o	1	0	0	**64**	35
		u	4	0	2	25	**69**
	5yr	a	**59**	3	0	38	1
		e	1	**46**	46	0	6
		i	0	5	**93**	0	3
		o	4	3	0	**69**	25
		u	0	11	2	3	**84**
	Adults	a	**100**	0	0	0	0
		e	3	**88**	10	0	0
		i	0	0	**92**	0	8
		o	0	1	0	**76**	23
		u	0	0	0	5	**95**
g	2yr	a	**90**	1	0	8	1
		e	4	**53**	16	5	22
		i	0	25	**48**	0	28
		o	1	0	0	**83**	16
		u	0	0	0	5	**95**
	5yr	a	**99**	0	0	0	1
		e	0	**70**	10	1	19
		i	0	10	**79**	0	11
		o	20	0	3	**74**	4
		u	0	0	0	16	**84**

	Adults	a	**99**	0	0	1	0
		e	4	**95**	0	0	1
		i	0	8	**78**	0	15
		o	1	0	0	**95**	4
		u	0	0	0	10	**90**
j	2yr	a	**69**	6	0	19	6
		e	9	**56**	13	11	11
		i	0	5	**80**	0	15
		o	18	3	0	**62**	18
		u	0	1	4	4	**91**
	5yr	a	**85**	1	0	11	3
		e	8	**51**	18	5	18
		i	0	13	**75**	0	13
		o	18	0	1	**50**	31
		u	2	14	15	1	**68**
	Adults	a	**96**	3	0	1	0
		e	1	**61**	30	0	8
		i	0	0	**98**	0	3
		o	1	3	2	**87**	8
		u	0	9	9	2	**80**
k	2yr	a	**98**	1	0	1	0
		e	5	**69**	14	5	8
		i	0	30	**60**	1	9
		o	1	1	4	**36**	58
		u	0	0	4	13	**84**
	5yr	a	**84**	0	0	15	1
		e	20	**40**	1	6	33
		i	0	1	**94**	0	5
		o	1	0	1	**36**	61
		u	0	0	0	4	**96**
	Adults	a	**100**	0	0	0	0
		e	1	**91**	4	1	3
		i	0	3	**94**	0	4
		o	0	0	0	**53**	47
		u	0	0	0	3	**97**

Korean Listeners

			Response Category						
Lang	Age	Target V	a	e	ʌ	i	ɨ	o	u

c	2yr	a	**38**	8	13	1	26	11	4
		e	0	**64**	5	15	10	6	0
		i	0	14	0	**80**	6	0	0
		o	4	0	15	0	0	**75**	6
		u	1	0	1	0	9	59	**30**
	5yr	a	**95**	0	4	0	0	0	1
		e	3	**90**	4	1	2	0	0
		i	0	6	0	**93**	1	0	0
		o	5	1	36	1	4	**49**	4
		u	1	0	0	0	0	59	**40**
	Adults	a	**91**	0	9	0	0	0	0
		e	0	**100**	0	0	0	0	0
		i	0	0	0	**99**	1	0	0
		o	0	0	53	0	0	**48**	0
		u	0	0	0	0	1	25	**74**
e	2yr	a	**94**	0	6	0	0	0	0
		e	3	**83**	0	14	1	0	0
		i	0	2	0	**97**	2	0	0
		o	1	0	16	0	0	**83**	0
		u	1	1	11	6	16	16	**49**
	5yr	a	**60**	0	24	0	0	14	3
		e	0	**78**	0	8	14	1	0
		i	0	0	0	**78**	13	0	10
		o	6	0	20	0	1	**71**	1
		u	0	14	0	30	13	5	**38**
	Adults	a	**93**	1	6	0	0	0	0
		e	0	**97**	0	3	0	0	0
		i	0	0	0	**100**	0	0	0
		o	3	0	3	0	0	**93**	3
		u	0	12	0	17	3	2	**66**
g	2yr	a	**75**	3	13	0	5	5	0
		e	1	**53**	6	13	21	1	5
		i	0	14	0	**64**	10	0	13
		o	1	0	15	1	3	**61**	19
		u	0	0	0	1	5	49	**45**
	5yr	a	**96**	0	3	1	0	0	0
		e	1	**63**	4	6	20	3	4
		i	0	8	0	**81**	6	0	5
		o	6	1	35	0	16	**38**	3
		u	0	1	1	0	1	81	**15**

	Adults	a	**98**	0	1	0	1	0	0
		e	0	**100**	0	0	0	0	0
		i	0	1	0	**86**	11	0	1
		o	3	0	11	0	1	**81**	4
		u	1	0	0	0	1	45	**53**
j	2yr	a	**48**	2	35	0	3	12	0
		e	1	**68**	3	3	14	10	3
		i	0	10	0	**78**	6	1	5
		o	7	1	37	0	8	**43**	6
		u	0	1	0	4	29	4	**63**
	5yr	a	**70**	0	24	0	1	5	0
		e	2	**59**	5	10	20	2	3
		i	0	9	0	**84**	4	0	4
		o	18	0	22	0	3	**48**	8
		u	0	1	3	12	50	3	**32**
	Adults	a	**91**	0	6	0	1	1	0
		e	0	**67**	0	18	13	1	0
		i	0	0	0	**99**	1	0	0
		o	2	0	13	0	3	**80**	3
		u	0	8	0	16	38	0	**38**
k	2yr	a	**94**	0	4	0	1	1	0
		e	0	**87**	0	2	10	2	0
		i	0	24	0	**70**	6	0	0
		o	0	0	5	1	14	**70**	10
		u	0	1	0	1	4	10	**84**
	5yr	a	**59**	0	38	0	0	4	0
		e	6	**39**	8	0	35	6	6
		i	0	0	0	**78**	14	0	9
		o	0	0	12	0	1	**35**	52
		u	0	0	0	1	3	18	**79**
	Adults	a	**99**	0	1	0	0	0	0
		e	0	**98**	0	1	0	1	0
		i	0	1	0	**97**	1	0	0
		o	0	0	0	0	0	**78**	23
		u	0	0	0	0	0	0	**100**

CHAPTER 4: STUDY 3

A cross-linguistic developmental study of vowel spectral movement patterns

The previous chapter examined listeners' vowel goodness judgments for both native and non-native vowels produced by 2-year-olds, 5-year-olds, and adults. While the acoustic values, measured at the vowel temporal midpoint, systematically predicted adult listeners' perception of non-native point vowels, no such relationship was found for native vowels nor for non-native, non-point vowels. In addition, the effect of speaker age on listener ratings was consistently significant for non-native vowels, but this effect was minimal for the perception of native vowels. One possible explanation for this difference between native versus non-native vowel perception is that language-specific vowel spectral movement patterns also play an important role in adult listeners' goodness ratings. Some previous research has found that F1 and F2 measured from vowel onset to the offset is an important cue for listeners' vowel identification (e.g., Nearey & Assmann, 1986; Strange, Jenkins, & Johnson, 1983).

In this chapter, therefore, the spectral movement patterns of five shared vowels (vowels with the same phonemic symbols), /a/, /e/, /i/, /o/, and /u/, produced by native speakers of American English, Greek, and Korean, will be investigated to examine if vowel spectral movement patterns have language-specific characteristics, which could play an important role in distinguishing native vowels from their non-native counterparts during perception. In addition, the vowel spectral movement patterns of adults of each language will also be compared to those of 2-year-olds and 5-year-olds within a language to understand the extent to which these cross-linguistic vowel spectral movement patterns are realized in young children's speech.

INTRODUCTION

Traditionally, vowels have been characterized with "slice-in-time" formant frequency values, a so-called acoustic "target" (e.g., Hillenbrand, Getty, Clark, & Wheeler, 1995; Peterson & Barney, 1952). The assumption behind this analysis is that vowels can be characterized by "target" formant frequencies, which are typically reached at or near the midpoint in time for each vowel. Studies on vowel acoustics, however, have shown that formant frequency values corresponding to this articulatory "target" were often not attained in conversational speech (e.g. Stevens & House, 1963). Instead, these "target" formant frequencies varied as a function of the surrounding consonant context, stress pattern, and speaking rate. Despite this variable, it was found that listeners could still successfully identify intended vowels. For example, Peterson and Barney (1952) showed that the vowels in conversational speech were identified accurately 94% of the time despite considerable overlap with neighboring vowels in F1-F2 values in the vowel space. Therefore, it has been suggested that the "slice-in-time" F1 and F2 values measured at the temporal midpoint of the vowel are not the only cue that listeners rely on for vowel identification. Instead, listeners may also attend to formant transitions (Peterson & Barney, 1952; Tiffany, 1953). This claim has been studied since then employing vowel identification tasks with different types of speech stimuli, with or without time-varying spectral information present, to understand the role of formant transitions in vowel identification.

Strange, Verbrugge, Shankweiler, & Edman (1976) studied listeners' performance on vowel identification in isolation versus /pVp/. They found that vowels in CVC contexts were more successfully identified by listeners than vowels in isolation. This finding suggests that formant transitions from vowel onset to vowel nucleus and from vowel nucleus to vowel offset are more important sources for vowel identification than target formant frequencies measured at a single point in time. Jenkins, Strange & Edman (1983) and Strange, Jenkins, & Johnson (1983) further examined this finding by generating different stimulus conditions in a /bVb/ context. Conditions included one natural, unmodified vowel type, and other different types of modified vowel conditions (formants in vowel initial, center, or final position were "attenuated to silence" or the duration was modified). These studies also found that listeners had more difficulty identifying target vowels when time-varying formant frequencies were attenuated than when "target" formant frequencies were removed, supporting the claim that formant transitions are more important in vowel identification than target formant frequencies values at vowel midpoint.

The important role of time-varying spectral information has also been documented for vowels in isolation. Nearey and Assmann (1986) examined the effect of vowel-inherent spectral change on listeners' vowel identification of isolated vowels in three different conditions; natural order (nucleus-10ms of silence interval-offglide), repeated nucleus (nucleus- 10ms of silence interval -nucleus), and reverse (offglide-10ms of silence interval -nucleus). Nearey and Assmann (1986) used the term 'vowel-inherent spectral change' to indicate changes in formant frequencies specific to vowels considered as monophthongs. This study showed that the error rate increased significantly for the two last conditions, where vowel inherent spectral change in offglide is either absent or changed. This result suggests that vowel-inherent spectral change plays an important role for vowel identification, even for vowels produced in isolation. Hillenbrand & Gayvert (1993) also provided evidence demonstrating the important role of vowel-inherent spectral change for vowel identification. In this study, steady-state synthesized vowels with no formant or pitch movement, but with equal durations of 300-ms were generated based on the F0 and F1-F3 values measured in Peterson and Barney (1952). A significantly higher error rate was found for synthesized vowels than for natural vowels. This result suggests that formant and pitch movements and durational information that were eliminated in synthesized vowels were important sources of information for vowel identification. Similarly, Hillenbrand and Nearey (1999), which examined listeners' vowel identification pattern for three different vowel stimuli types found that synthetic vowels with formant movements preserved were identified with greater accuracy than those without formant movements.

Despite the increasing interest in the role of vowel spectral change on vowel identification and phonetic specification, however, there are few studies that systematically examined the spectral movement patterns of each vowel category across speakers of different ages or different languages. Hillenbrand *et al*., (1995), as well as Hillenbrand & Nearey (1999) and Nearey & Assmann, (1986) studied the F1 and F2 trajectory patterns of vowels produced by native adult speakers of two different dialects of English (Detroit and Canadian English) and showed that a fair amount of spectral change occurred in phonetically diphthongized /e/ and /o/, as well as in other monophthongal vowels in an /hVd/ context for both dialects. F1 and F2 trajectories of vowels produced by monolingual English-speaking children (aged 10 to 12, and 3-, 5-, and 7-year-olds) have also been examined (Hillenbrand & Nearey, 1999; Assmann & Katz, 2000). These studies found that children's vowels had very similar spectral movement patterns to those of adults, resulting in only a small effect of talker age-group on vowel identification. Differences in vowel spectral movement patterns across age groups were described qualitatively. In sum, there is limited research on time-varying vowel spectral patterns in languages other than English, and there are no studies that have used quantitative methods to compare formant trajectories across vowels, languages, or age groups.

In an earlier study, Chung, Kong, & Weismer (2010) examined time-varying spectral patterns between Korean and English vowels produced by adults and found statistically significant cross-linguistic differences; English vowels had relatively greater spectral movements in consistent directions, while Korean vowels had minimal spectral movement and no consistent direction of movement. This result indicated that time-varying spectral change patterns in vowels produced by speakers of different native languages could have language-specific patterns. Furthermore, in chapter 2, it was observed that vowel goodness ratings by adults of non-native and native vowels differed in ways that could not be explained solely by F1 and F2 values measured at the vowel midpoint, suggesting that listeners are sensitive to these cross-linguistic differences in formant trajectories. The focus of this study, therefore, was to examine spectral movement patterns of five common vowels, /a, i, u, e, o/, across three different languages and three different age-groups (2-year-olds, 5-year-olds, and adults).

Aims of Current Study

The specific aims of the present study are the following: 1) to assess whether the five vowels /a, i, u, e, o/, produced by native speakers of American English, Greek, and Korean, show language-specific vowel spectral movement patterns and 2) to investigate the extent to which these language-specific spectral movement patterns are realized in each of five vowels produced by 2- and 5-year-olds of American English, Greek, and Korean. Based on previous research, three predictions were made. First, it was predicted that language-specific, vowel spectral movement patterns would be observed. The results of the previous chapter showed that listeners judge native vowels as “better” exemplars than non-native vowels of the same category, despite the similar “slice-in-time” F1 and F2 values. Cross-linguistic differences in formant trajectories may explain this result, given previous research showing that .listeners attend more to time-varying spectral movements than to target formant frequencies. Second, it was predicted that vowels produced by very young children (2-year-olds) will show language-specific vowel spectral movement patterns as those of adults of their native language, based on the results of similar research on English (Hillenbrand & Nearey, 1999; Assmann & Katz, 2000).

METHOD

Participants

The participants included ten speakers (5 males and 5 females) in each of three age groups (2-year-olds, 5-year-olds, and adults) and language groups. The participants were from a larger study on cross-linguistic phonological acquisition, the *paidologos* project

(Edwards & Beckman, 2008; Edwards & Beckman, 2009). All participants were monolingual native speakers of American English, Greek, and Korean. Adult speakers' ages ranged from 18 to 30 years. Each subject was recruited from a single region of each country by a native speaker from the same dialect region. American English-speaking subjects were recorded in Columbus, Ohio; Greek-speaking subjects in Thessaloniki, Greece; and Korean-speaking subjects in Seoul, Korea. All child participants passed a hearing screening using otoacoustic emissions and had age-appropriate oromotor skills (Kaufman, 1995). All adult participants passed a hearing screening and reported no history of speech, language, or hearing problems.

Stimuli

The stimuli were /a/, /e/, /i/, /o/, and /u/ vowels in real words that began with fricative-vowel sequences. Only tokens beginning with fricatives were used in this study to remove any possible effect of different consonantal environments on vowel formant patterns. The word initial consonants were alveolar /s/ and post-alveolar /ʃ/ for American English, alveolar /s/ for Greek, and dento-alveolar /s/ for Korean. For Korean, the post-alveolar fricative [ʃ], is the allophone of /s/, which appears before /i/. Both [s] and [ʃ] were included in the analysis for Korean. For American English, /ɑ/, /i/, /u/, /e/, and /o/ vowels were included in the analysis. For both Greek and Korean, the five common vowels, /a/, /e/, /i/, /o/ and /u/, preceded by fricatives were analyzed. For English 2-year-olds, /e/ and /o/ were not included because there were no productions of /e/ and /o/ that had been transcribed as correct in word-initial /s/ or /ʃ/sequences. Similarly, for Korean 2-year-olds, /o/ was not included because many of the /o/ productions of 2-year-olds that were transcribed as correct were devoiced. The number of vowels for each consonant environment, language, and age group is provided in Table 4.1.

Table 4.1 Word initial consonant environment and number of vowels used for the analysis

English						
		/ɑ/	/e/	/i/	/o/	/u/
Adults	/s/	28	20	0	29	0
	/ʃ/	29	0	9	27	9
5yr	/s/	22	17	0	25	0
	/ʃ/	25	0	9	27	7
2yr	/s/	26	2	0	0	0
	/ʃ/	29	0	6	0	6
Greek						
		/a/	/e/	/i/	/o/	/u/
Adults	/s/	28	27	27	28	28
5yr	/s/	27	28	31	28	29
2yr	/s/	25	34	32	22	24
Korean						
		/a/	/e/	/i/	/o/	/u/
Adults	[s]	19	18	0	18	9
	[ʃ]	0	0	19	0	0
5yr	[s]	20	2	0	13	8
	[ʃ]	0	0	19	0	0
2yr	[s]	16	5	0	2	4
	[ʃ]	0	0	17	0	0

Procedures

Speech samples were collected using a word repetition task. The stimuli were familiar real words that were pictureable and culturally appropriate. Participants were asked to repeat each word after a color picture was presented on a laptop screen as an auditory recording of a target word was played. Productions were digitally recorded using a Marantz PMD 660 flashcard recorder at a sampling rate of 44,100 Hz. A trained native phonetician of each language transcribed the target vowels as correct or incorrect. Only the vowels that were judged as 'correct' were used in the subsequent acoustic analysis. The transcription procedure is described in detail in Chapter 1.

Data Analysis

a. F1 and F2 trajectory measurement

The analysis was done using the LPC solution in *Praat* (version 5.0.29) (Boersma & Weenink, 2006). The size of the analysis window was 25ms and the dynamic range was set at 30dB. First, vowel onset and offset were aligned manually for each token. Vowel onset was defined as the time at which the first clear glottal pulse was observed in both the waveform and spectrographic display, and vowel offset was defined as the final glottal pulse extending at least through F1 and F2. The first two formant frequency values (F1 and F2) were extracted sequentially from vowel onset to vowel offset, with a step size of 6ms.

b. Vowel formant transformations

The raw formant frequency values reflect more than the linguistic properties of each vowel. These values are also affected by different vocal tract sizes of speakers of different age and language groups. In order to remove this effect and make a direct comparison of vowel spectral movement patterns across different age and language groups, the data transformation procedure applied in Chapter 1 was also employed in this study. The procedure transforms the F1 and F2 values of each speaker into a speaker-specific center point in a vowel acoustic space. Since all the acoustic values have been transformed based on speaker specific centroids, we can eliminate or at least minimize the effect of vocal tract size on the formant frequency data. The data transformation procedures were as follows: first, all the raw F1 and F2 values extracted from vowel onset to vowel offset for each token were log-transformed. Second, the mean log F1 and log F2 values measured at the vowel midpoint were calculated for each of the five vowels, /a/, /i/, /u/, /e/, and /o/. Third, the mean of the mean log F1 and log F2 of these five vowels was calculated to obtain a speaker-specific centroid. Finally, the polar coordinate system was transformed to a Cartesian coordinate system so that the speaker-specific centroid was the origin (0,0). Then, each data point was transformed relative to this speaker-specific centroid (See Figure 2.1 in Chapter 2).

In order to compare formant trajectories across languages and across age groups using quantitative measures, it was necessary to normalize these trajectories across time. To do so, the duration of each vowel was divided into seven different proportional time points. The bin size differed across tokens within a vowel, across vowels within a speaker, and across speakers. F1 and F2 values were extracted from the second to the fifth time points to construct formant trajectories for each vowel for each speaker. Formant values at the first time point and the sixth and seventh time points were excluded from the analysis to minimize potential coarticulatory effects from the preceding or following consonants. Thus, the second time point was treated as the vowel "onset" and the fifth time point as the vowel "midpoint" or the endpoint of the trajectory. The focus of this study is the spectral movement pattern only from vowel onset to midpoint, as defined here.

RESULTS

Two analyses will be discussed. First, the F1 and F2 trajectories of the five vowels, /a, e, i, o, u/, produced by adults of three languages (American English, Greek, and Korean) will be examined to determine whether there are systematic cross-linguistic differences in formant trajectories. If so, then the F1 and F2 trajectories of vowels produced by adults, 5-year-olds, and 2-year-olds within each language will be analyzed in order to determine the extent to which these language-specific spectral movement patterns were produced by young children.

Vowel trajectories in an F1 by F2 space

Figure 4.1 shows the formant movement patterns of five vowels, /a/, /i/, /u/, /e/, and /o/, produced by adults, 5-year-olds, and 2-year-olds who are native speakers of American English, Greek, and Korean in a log F1 by log F2 vowel acoustic space.

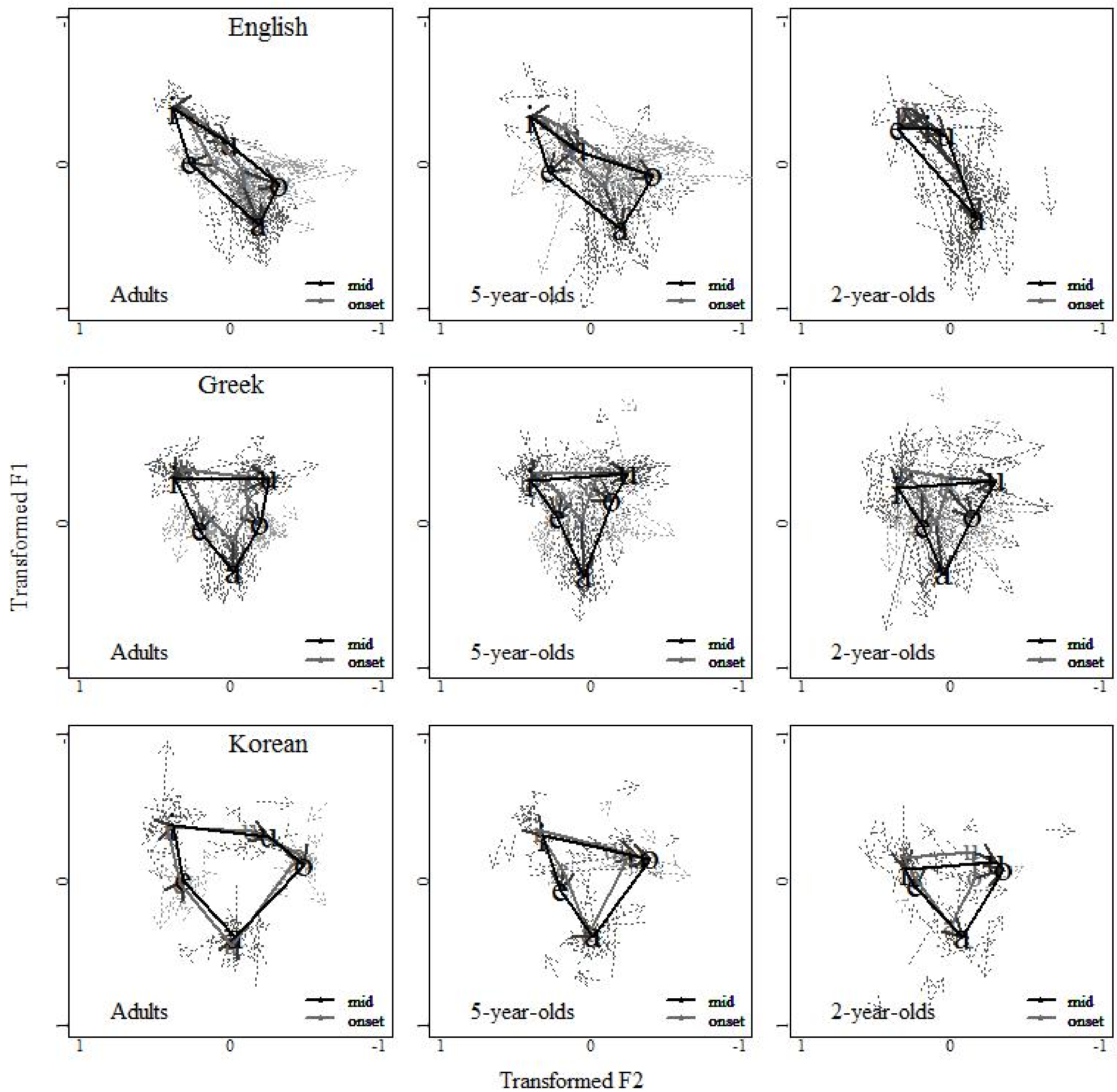

Figure 4.1 The spectral movement patterns of five vowels, /a/, /i/, /u/, /e/, and /o/, produced by adults (1st column), 5-year-olds (2nd column), and 2-year-olds (3 rd column) of American English (top), Greek (middle), and Korean (bottom) in a transformed F1 by F2 vowel acoustic space (log scale). The x-axis represents transformed F1 values and the y-axis represents transformed F2 values. Each arrow represents formant trajectory of each token. The thick black arrow indicates the averaged formant trajectories of each of five vowels. The grey vowel spaces were formed by connecting mean F1 and F2 values measured at the vowel onset and the black vowel spaces indicate those at the vowel midpoints.

It can be observed in Figure 4.1 that formant trajectories differ across the three languages for vowels produced by adults. The vowels produced by English and Greek speakers had more spectral movement than Korean vowels, and had more consistent direction of movements; all moving toward the periphery of the vowel acoustic space. This indicates that each vowel category became more acoustically distinct as target formant frequencies were reached. In contrast, vowels produced by Korean speakers showed minimal spectral change and had no consistent direction of movements across vowel types. Within vowel categories, for /i/, all languages showed minimal spectral movements and had similar F1 and F2 values. In contrast, for /a/, greater spectral movement was observed in English and Greek; the /a/ of these two languages started in a relatively similar location in the vowel space but ended up in language-specific points, while F1 and F2 values of Korean /a/ remained relatively stable from vowel onset to the midpoint. For /u/, all languages showed a language-specific pattern of movement; the /u/ of each language started from different points and moved toward a language-specific location in the vowel acoustic space. Even the /u/ produced by Korean speakers showed some movement in the F2 dimension.

The spectral movement patterns of 2- and 5-year-olds were generally very similar to those of the adults of their native languages. Like the adult patterns, vowels produced by Greek and English children showed more spectral movement than those produced by Korean children across all vowel types. Moreover, the direction of vowel spectral movements of English and Greek children patterned similarly to those of adults of their native languages, all moving toward the periphery of the vowel space. On the other hand, relatively little spectral movement was observed for vowels produced by Korean adults, 5-year-olds, and 2-year-olds, with no systematic direction of movement.

F1 and F2 trajectory analysis

a. Cross-linguistic patterns

Figures 4.2 and 4.3 show the time-normalized trajectories across languages and vowels. It can be observed that both F1 and F2 trajectories of vowels produced by English and Greek-speaking adults had systematic patterns, with either upward, downward, or minimal movement with time. In contrast, F1 and F2 trajectories of vowels produced by Korean-speaking adults were less systematic and had more variable patterns as compared to those of English and Greek adults.

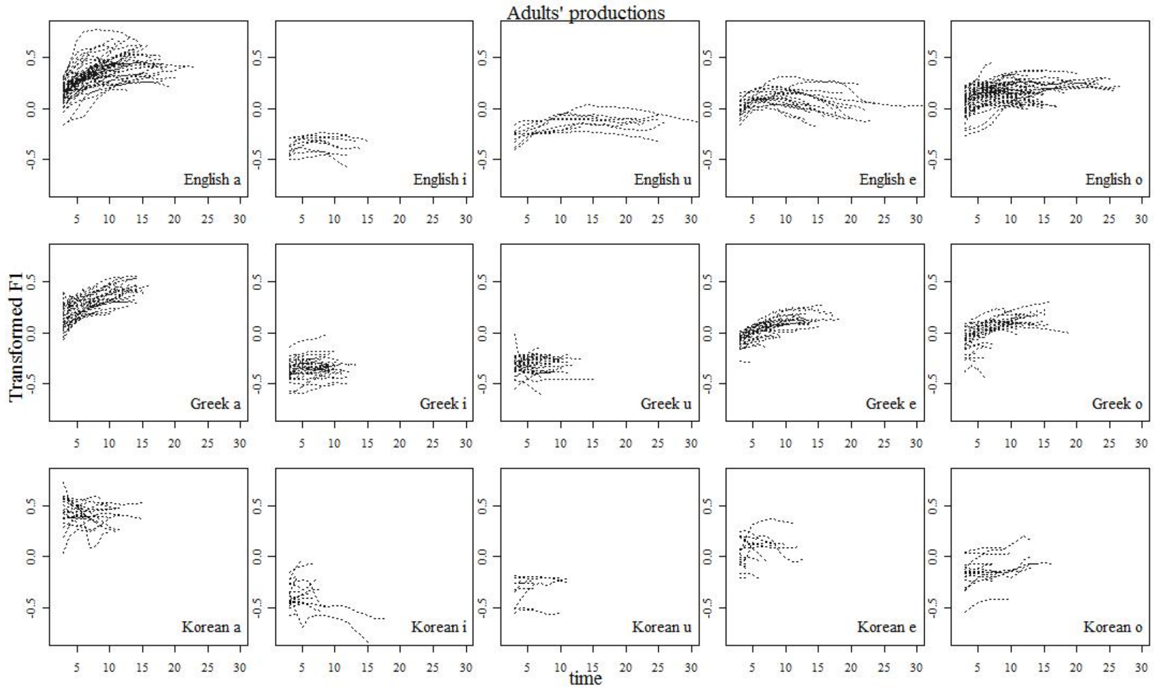

Figure 4.2 F1 trajectory patterns (raw data) as a function of vowel duration, from vowel onset to the midpoint, for each of five vowels, /a/, /i/, /u/, /e/, and /o/, produced by adult speakers of American English, Greek, and Korean. First row shows F1 trajectories of vowels produced by American English-speaking adults, the second row those of Greek-speaking adults, and the third row for those of Korean-speaking adults. The x-axis shows the time, and y-axis shows the transformed F1 values. The increase of each unit of time represents 6ms increase in time (i.e., time 1: 6ms, time 2: 12ms, time 3: 18ms...etc.).

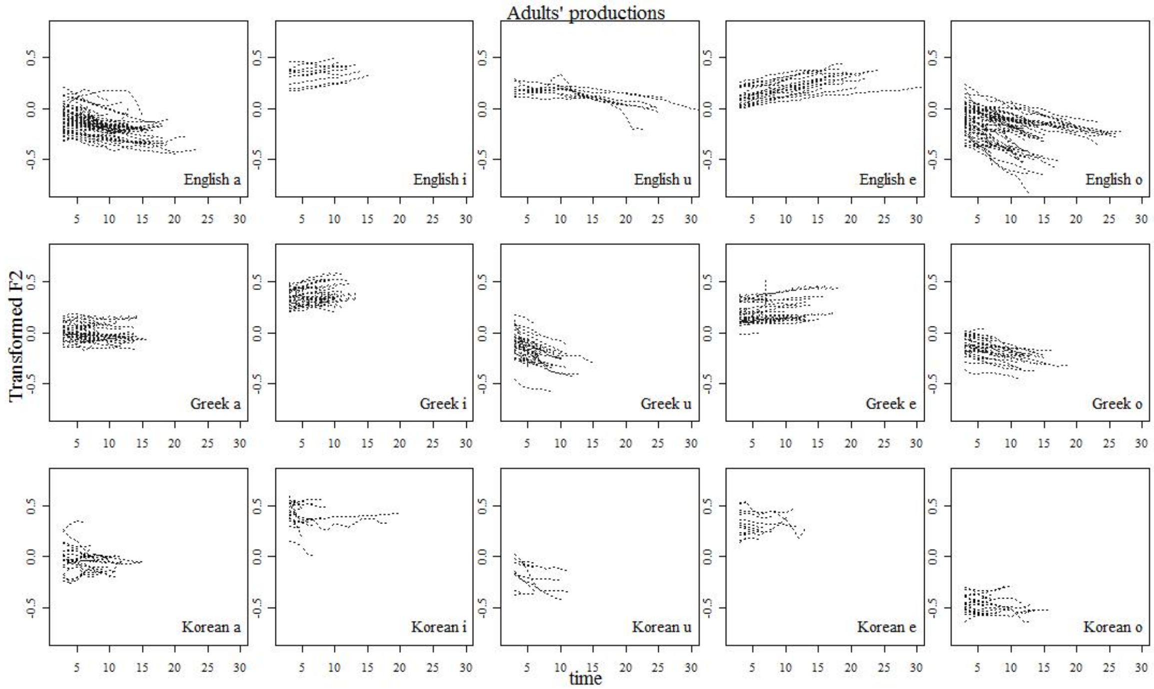

Figure 4.3 F2 trajectory patterns (raw data) of each of five vowels, /a/, /i/, /u/, /e/, and /o/, produced by adult speakers of American English, Greek, and Korean. The x-axis represents the time, with each unit of time representing 6ms increase in time from ‘0’. The y-axis represents the transformed F2 values.

In order to analyze the differences across language and age group, a series of mixed-effect models using a time-series analysis (Barr, 2008; Mirman, Dixon & Magunuson, 2008; Singer & Willett, 2003) was run. Models were run separately for each language, age group, and vowel, resulting in a total of 90 analyses (3 languages * 3 age groups * 5 vowels separately for F1 and F2). The dependent variables were always the transformed F1 and F2 values (in log scale). The independent variables were the first three orders of orthogonal polynomial time parameters. These three parameters were added to the model to capture the shape changes of the formant trajectory of each vowel as a function of time. Each order of polynomial terms describes different components of the curve shape in an independent manner. The first order of the orthogonal polynomial describes the linear slope of the curve. A negative coefficient indicates a decreasing linear trend over time, and a positive coefficient indicates an increasing trend. The second and the third orders describe the quadratic and cubic terms, respectively. The quadratic term describes the change of the curve over time as concave or convex, and the cubic term describes the symmetric or asymmetric shape of the curve. While all three orders were used in this analysis to make precise description of formant frequency trajectories, the focus will be mainly on the linear term and quadratic term. The first order parameter (linear term) was entered as a random effect, and the second and the third order parameters were entered as fixed effects. The intercept and slope were allowed to vary for each token. The analyses were done using R (R Development Core Team, 2010) and its implemented packages lme4 (Bates & Maechler, 2009).

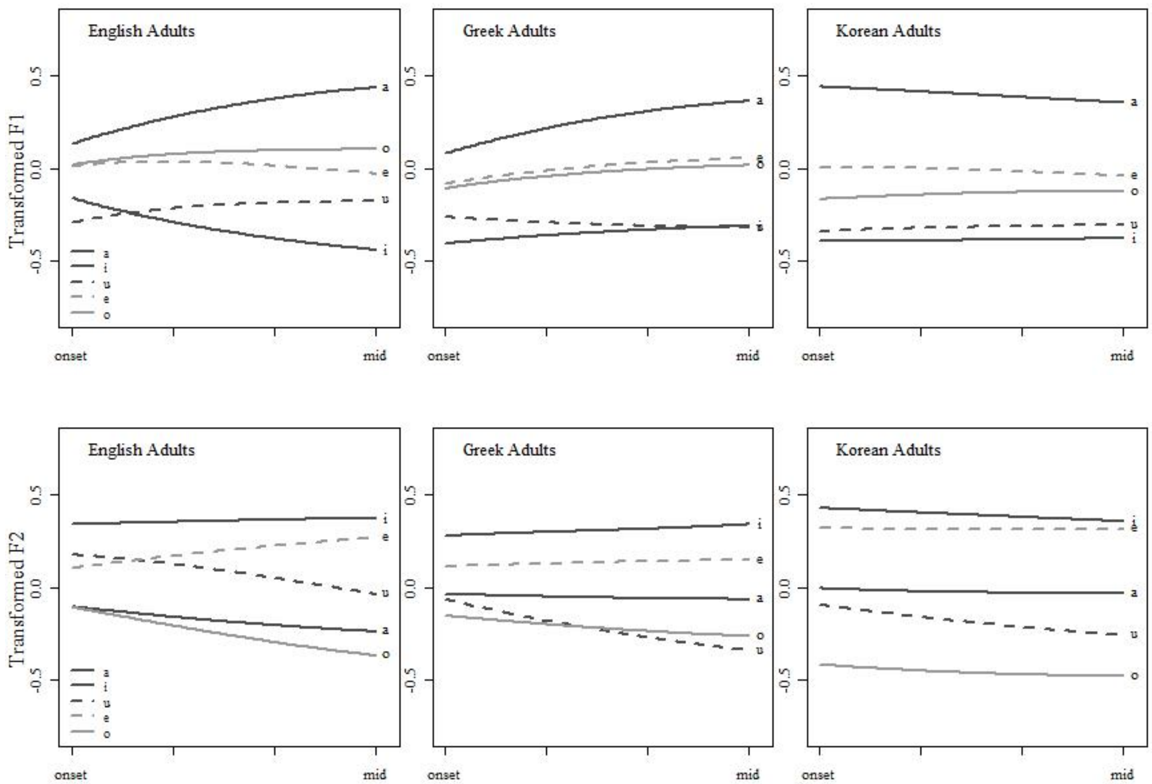

Figure 4.4 F1 and F2 trajectory patterns of each of five vowels, /a/, /i/, /u/, /e/, and /o/, produced by adult speakers of American English (1st column), Greek (2nd column), and Korean (3rd column), as estimated by the mixed-effects regression models. The x-axis represents the four proportional time, from vowel onset to the midpoint. The y-axis represents the transformed F1 (top row) and F2 (bottom row) values.

The trajectory patterns in Figure 4.4 and 4.5 and table 4.2 and 4.3 can be interpreted in the following way: for F1 trajectories, a positive linear slope indicates a movement from a relatively central position to the lower part of the vowel space, which is associated with a more open vocal tract. A negative linear slope indicates movement toward the upper part of the vowel space, which is associated with a more closed vocal tract. For F2 trajectories, a positive linear slope indicates a movement toward the front part of the vowel space, which is associated with a more front constriction of the vocal tract.A negative linear slope indicates a movement toward a more posterior part of the vowel space, which is associated with a more back constriction of the vocal tract. For the quadratic term, either positive or negative coefficients can be interpreted in terms of the shape of the trajectories: the positive term indicates concave shaped trajectories to the x-axis, while a negative quadratic term indicates convex shaped trajectories to the x-axis.

The results from the mixed-effects models for F1 trajectories for adults showed that for English, only the low vowel /a/ had significant linear and quadratic terms. The linear term had a positive slope and the quadratic term was negative, indicating F1 trajectories have a concave shape of fitted curves and F1 trajectories move toward the lower part of the vowel space as the trajectory moves from onset to midpoint. For Greek, all five vowels except /u/, had a significant positive linear slope. That is, F1 trajectories of /a/, /i/, /e/, and /o/ all showed a movement toward the lower part of the vowel space. The quadratic term was also significant (negative) for these four vowels, indicating that the fitted curves had a concave shape. For Korean, neither the linear nor the quadratic term was significant for all five vowels, indicating a relatively shallow shape of the F1 trajectories, as compared to those for English and Greek vowels.

For F2 trajectory patterns, the English back vowels /a/, /o/, and /u/ all showed significant linear movement toward the more posterior part of the vowel space. The quadratic term was also significant (negative), indicating a concave shape. For Greek, all five vowels had significant linear slopes; front vowels /i/ and /e/ had a significant positive slope, indicating movement toward the more front part of the vowel space, while the back vowels /u/ and /o/ showed a significant downward movement, indicating a movement toward more posterior part of the vowel space. The quadratic term was also significant (positive for /a/, /u/, and /o/, but negative for /i/), indicating a convex shape of the F2 trajectories of these three vowels, and a concave shape for /i/. The linear term of the low central vowel /a/ were also significant (negative), but it was only marginally significant, suggesting that there was movement toward the posterior part of the vowel space, but this movement was not as significant as those of its English counterpart. In other words, F2 trajectories of all five vowels produced by Greek adults showed a movement toward the more peripheral part of the vowel space from vowel onset to the midpoint. For Korean, the same negative significant linear slopes were found only for /i/ and /u/, indicating a significant movement toward the more posterior part of the vowel space. However, coefficients of linear terms of Korean vowels were in general smaller than those of English and Greek vowels, indicating that there is less F2 trajectory movement for Korean vowels than for English and Greek vowels. The quadratic term was not significant for any of the Korean vowels, except /o/ (which had a significant positive term, although it was only marginally significant).

Overall, the result from the mixed-effects regression models showed that F1 and F2 trajectories of vowels produced by Greek-speaking adults had the most significant extent of movement as compared to those of English and Korean. English vowels showed a similar pattern of movement as those of Greek vowels, but had less steep slopes, indicating less movement. Korean vowels, in general, showed little change in F1 or F2 values from vowel onset to midpoint, except for /u/, where a significant linear slope was observed for F2 trajectories. The coefficients of the linear and quadratic terms for F1 and F2 trajectories are summarized in Table 2.

Table 4.2 The output of the mixed-effects regression models for F1 and F2 trajectories of vowels produced by adult speakers of English, Greek, and Korean. The significant terms are in bold.

F1		English			Greek			Korean		
		Estimate	SE	t value	Estimate	SE	t value	Estimate	SE	t value
/a/	Linear	0.22	0.02	**12.81**	0.21	0.02	**9.48**	-0.07	0.04	-1.66
	Quadratic	-0.11	0.01	**-11.25**	-0.10	0.01	**-10.97**	0.00	0.02	-0.22
/i/	Linear	-0.20	0.12	-1.66	0.07	0.01	**4.81**	0.00	0.04	0.07
	Quadratic	0.09	0.06	1.50	-0.03	0.01	**-4.76**	0.01	0.03	0.32
/u/	Linear	0.06	0.03	1.98	-0.04	0.05	-0.89	0.02	0.04	0.62
	Quadratic	-0.08	0.02	**-3.71**	0.02	0.02	0.98	-0.01	0.02	-0.75
/e/	Linear	0.07	0.03	1.99	0.10	0.02	**4.73**	-0.04	0.03	-1.56
	Quadratic	-0.07	0.01	**-5.51**	-0.06	0.01	**-6.16**	-0.03	0.01	-1.72
/o/	Linear	0.05	0.03	1.66	0.08	0.02	**4.11**	0.03	0.02	1.40
	Quadratic	-0.08	0.01	**-8.03**	-0.05	0.01	**-6.83**	-0.03	0.01	-1.95
F2		English			Greek			Korean		
		Estimate	SE	t value	Estimate	SE	t value	Estimate	SE	t value
/a/	Linear	-0.10	0.01	**-9.55**	-0.02	0.01	**-2.27**	-0.02	0.03	-0.64
	Quadratic	0.03	0.01	**4.97**	0.01	0.00	**2.45**	0.02	0.01	1.96
/i/	Linear	0.03	0.03	0.99	0.04	0.01	**2.76**	-0.05	0.02	**-2.49**
	Quadratic	0.00	0.01	-0.27	-0.03	0.00	**-5.80**	0.01	0.01	0.61
/u/	Linear	-0.17	0.02	**-6.83**	-0.20	0.02	**-8.52**	-0.13	0.04	**-3.38**
	Quadratic	-0.04	0.02	**-2.43**	0.05	0.01	**4.90**	0.04	0.02	**2.02**
/e/	Linear	0.01	0.02	0.78	0.02	0.01	**2.75**	-0.01	0.01	-0.36
	Quadratic	0.00	0.00	0.44	0.00	0.00	-1.25	0.01	0.01	1.12
/o/	Linear	-0.20	0.02	**-12.82**	-0.08	0.01	**-6.04**	-0.05	0.03	-1.74
	Quadratic	0.04	0.01	**6.34**	0.02	0.01	**3.56**	0.03	0.01	**2.34**

b. Developmental patterns

For the developmental patterns, only the results from F2 trajectory patterns will be described. The patterns for F1 were very similar. Figure 4.5 shows fitted regression curves for F2 trajectories for each model, separately for each vowel, language, and age group.

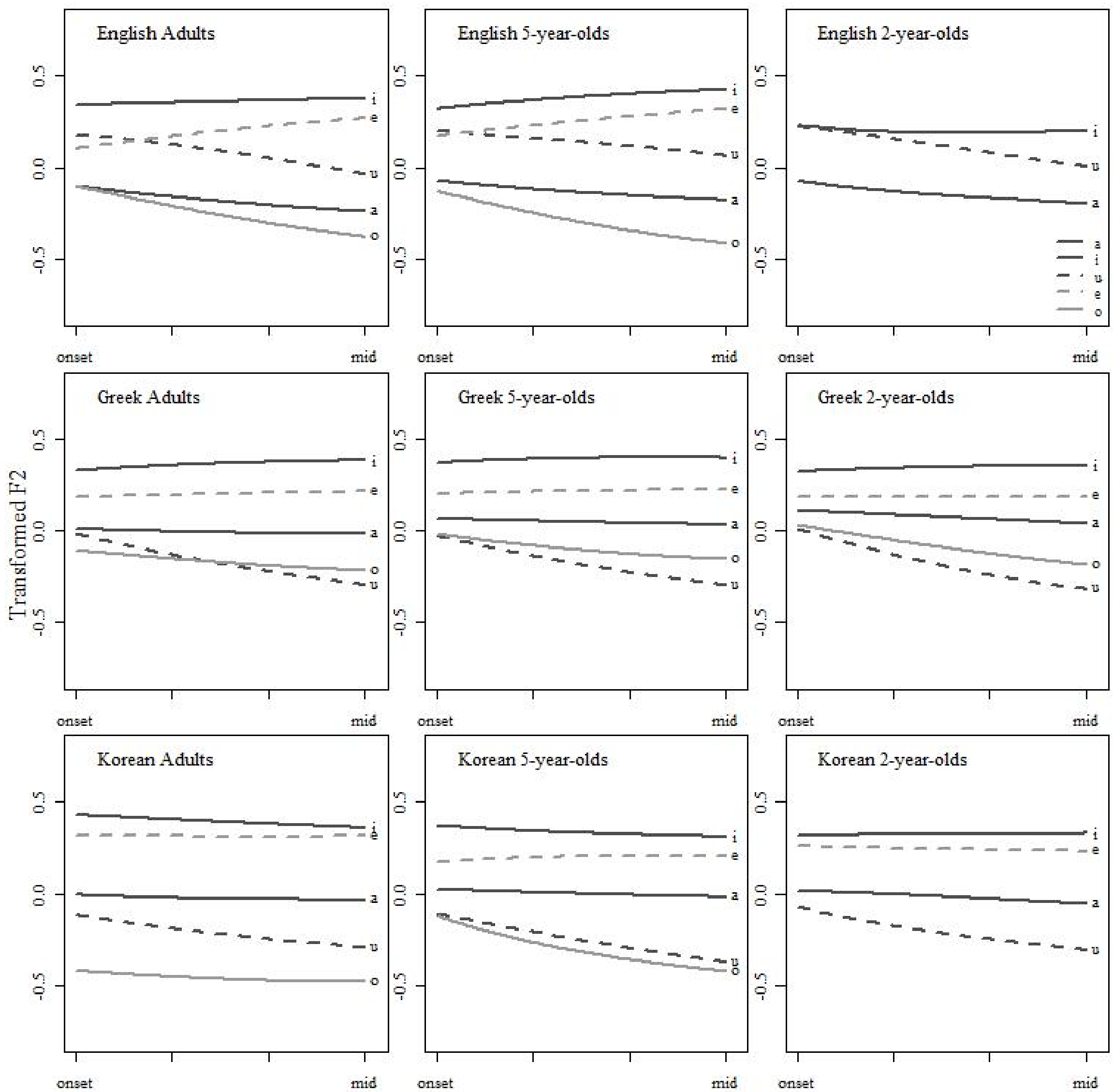

Figure 4.5 F2 trajectory patterns of each of five vowels, /a/, /i/, /u/, /e/, and /o/, produced by adults (1st column), 5-year-olds (2nd column), and 2-year-olds (3rd column) of American English (top), Greek (middle), and Korean (bottom), as estimated by the mixed-effects regression models. The x-axis represents the four proportional time, from vowel onset to the midpoint. The y-axis represents the transformed F2 values.

It can be observed that the F2 trajectories of the 5-year-olds and 2-year-olds were remarkably similar to those of adults of their native languages. In general, the linear and quadratic terms that were significant for adults' vowels were equally significant in the 2- and 5-year-olds' vowels in the same direction for both the linear and quadratic terms. For example, for /u/ vowels, F2 trajectories of /u/ produced by English-speaking adults, 5-year-olds and 2-year-olds all had a significant movement toward the more posterior part of the vowel space. The same pattern was found for /u/ produced by Greek-speaking adults, 5-year-olds and 2-year-olds. F2 trajectories of /u/ produced by Greek speakers of all three age groups had steeper downward movement than /u/ produced by English speakers of all age groups. The trajectory patterns of the vowels produced by Korean speakers were also similar across the age groups. For example, /u/ produced by all three age groups had a significant downward movement. The coefficients of both linear and quadratic terms of F2 trajectories of five vowels are summarized in Table 4.3. Those of F1 trajectories are summarized in Appendix A.

Table 4.3 The output of the mixed-effects regression models for F2 trajectories of five vowels produced by adults, 5-year-olds, and 2-year-olds of American English, Greek, and Korean. The significant terms are in bold.

			English			Greek			Korean		
			Estimate	SE	t value	Estimate	SE	t value	Estimate	SE	t value
/a/	Adults	Linear	-0.10	0.01	**-9.55**	-0.02	0.01	**-2.27**	-0.02	0.03	-0.64
		Quadratic	0.03	0.01	**4.97**	0.01	0.00	**2.45**	0.02	0.01	1.96
	5yr	Linear	-0.09	0.02	**-5.07**	-0.03	0.01	-1.73	-0.03	0.03	-1.09
		Quadratic	0.03	0.01	**4.37**	0.00	0.01	0.51	0.01	0.01	0.62
	2yr	Linear	-0.10	0.02	**-6.14**	-0.05	0.03	-1.81	-0.06	0.03	-2.13
		Quadratic	0.04	0.01	**5.50**	0.00	0.02	-0.09	-0.01	0.01	-0.93
			Estimate	SE	t value	Estimate	SE	t value	Estimate	SE	t value
/i/	Adults	Linear	0.03	0.03	0.99	0.04	0.01	**2.76**	-0.05	0.02	**-2.49**
		Quadratic	0.00	0.01	-0.27	-0.03	0.00	**-5.80**	0.01	0.01	0.61
	5yr	Linear	0.07	0.03	**2.47**	0.02	0.01	1.58	-0.01	0.02	-0.52
		Quadratic	-0.02	0.02	-1.53	-0.03	0.01	**-5.27**	0.01	0.01	1.11
	2yr	Linear	-0.04	0.07	-0.51	0.02	0.02	1.35	-0.01	0.02	-0.61
		Quadratic	0.05	0.05	0.96	-0.02	0.00	**-3.55**	0.00	0.01	0.65
			Estimate	SE	t value	Estimate	SE	t value	Estimate	SE	t value
/u/	Adults	Linear	-0.21	0.05	**-4.03**	-0.20	0.02	**-8.52**	-0.13	0.04	**-3.38**
		Quadratic	-0.05	0.01	**-4.03**	0.05	0.01	**4.90**	0.04	0.02	**2.02**
	5yr	Linear	-0.12	0.03	**-3.82**	-0.20	0.02	**-9.32**	-0.20	0.03	**-6.39**
		Quadratic	-0.02	0.01	-1.55	0.05	0.01	**4.68**	0.03	0.02	1.27
	2yr	Linear	-0.20	0.06	**-3.38**	-0.26	0.03	**-7.80**	-0.08	0.05	-1.56
		Quadratic	0.03	0.02	1.36	0.08	0.01	**6.47**	0.03	0.04	0.69
			Estimate	SE	t value	Estimate	SE	t value	Estimate	SE	t value
/e/	Adults	Linear	0.13	0.02	**8.46**	0.02	0.01	**2.75**	-0.01	0.01	-0.36
		Quadratic	-0.03	0.01	**-5.57**	0.00	0.00	-1.25	0.01	0.01	1.12
	5yr	Linear	0.07	0.04	1.69	0.02	0.01	1.22		NA	
		Quadratic	-0.05	0.02	**-2.22**	-0.01	0.00	**-2.43**			
	2yr	Linear		NA		0.00	0.01	0.04	-0.02	0.03	-0.67
		Quadratic				0.01	0.01	1.00	0.01	0.01	0.37
			Estimate	SE	t value	Estimate	SE	t value	Estimate	SE	t value
/o/	Adults	Linear	-0.21	0.02	**-10.31**	-0.08	0.01	**-6.04**	-0.05	0.03	-1.74
		Quadratic	0.03	0.01	**4.75**	0.02	0.01	**3.56**	0.03	0.01	**2.34**
	5yr	Linear	-0.23	0.03	**-9.01**	-0.10	0.02	**-4.39**	-0.18	0.04	**-4.99**
		Quadratic	0.07	0.01	**7.21**	0.04	0.01	**5.02**	0.10	0.03	3.63

2yr	Linear		NA		-0.15	0.04	**-3.58**		NA	
	Quadratic				0.02	0.02	1.56			

DISCUSSION

The purpose of this study was to examine spectral movement patterns of vowels produced by monolingual speakers of three languages with different vowel systems, but a core phonemic set of common vowels. Two predictions were made based on the literature. First, it was predicted that there would be significant cross-linguistic differences in the F1 and F2 trajectory patterns from onset to midpoint for vowels produced by native adult speakers of American English, Greek, and Korean. Second, it was predicted that, if cross-linguistic differences were observed in the trajectories of vowels produced by adults, then these differences would also be reflected in the trajectories of vowels produced by 2- and 5-year-olds. Both predictions were confirmed. First, significant cross-linguistic differences were found in adults' vowel spectral movement patterns across the three languages. It was observed that English and Greek vowels had more spectral movement and this movement was in a consistent direction, toward the periphery of the vowel space for the three point vowels, while Korean vowels did not have much movement and when movement occurred it was not in a consistent direction. These observations were generally confirmed by mixed-effects models. This finding supports the claim that spectral movement properties of vowels may differ across languages. That is, even if vowels of two languages share similar midpoint F1 and F2 values, the spectral movement patterns of the two languages may differ. For example, the result from Chapter 2 showed that /a/ of Greek and Korean have very similar temporal midpoint F1 and F2 values. The spectral movement patterns of /a/ of Greek and Korean, however, differ from each other; while F1 and F2 trajectories of Greek /a/ produced by adults had significant movement, those of Korean /a/ showed little change from vowel onset to midpoint. These language-specific, vowel spectral movement patterns also provide an explanation for the result reported in Chapter 3, which showed that naive adult listeners of Greek and Korean rated native vowels as better examples of a specific vowel category than the non-native counterparts despite the similar static-point F1 and F2 values. The differences in formant trajectories may explain why listeners gave higher goodness ratings to native vowels relative to non-native vowels. This interpretation is consistent with previous findings that showed the important role of time-varying spectral movement patterns for vowel identification (e.g., Strange *et al*. 1976).

Cross-linguistic differences with respect to vowel spectral movement patterns differed across vowels. This result is not surprising, given that Chapter 2 found language-specific differences in F1 and F2 values at vowel-midpoint for some vowels, but not for others. This study showed that the spectral movement patterns of /u/ produced by adult speakers of English had a significant movement into a more posterior part of the vowel space. The findings of previous studies, however, showed that /u/ had the least spectral movement among vowels that are normally treated as monophthongs (Hillenbrand *et al.* 1995; Nearey & Assmann, 1986). The conflicting results between the previous studies and the current study could be due to the effect of dialect. The speakers of Nearey & Assmann (1986) were from Alberta, Canada and those of Hillenbrand *et al.* (1995) were from Michigan. English speakers in the current study were from Columbus, Ohio. Recent work (Jacewicz & Fox, 2011) has found that /u/ produced by Columbus, Ohio speakers is almost as fronted as the /u/ produced in Southern American English. It is considerably more fronted than the /u/ produced by Wisconsin speakers, who are presumably similar to Michigan and Canadian speakers. Another possible explanation for these differing findings may be related to the sound change of /u/ which is currently in progress, as discussed in Chapter 2 (e.g., Clopper & Pisoni, 2006; Harrington, Kleber, & Reubold, 2008). The /u/ sounds studied in both Hillenbrand *et al.,* (1995) and Nearey & Assmann (1986) were recorded at least a decade or more before the speech samples of the current study. Studies have reported that the fronting of back vowel /u/ is still in progress across many of the regions in the U.S. and it is the younger generation, especially in California, which is leading this change (e.g., Fridland, 2008). Therefore, the different spectral movement patterns of these studies could be a reflection of sound change in progress. In contrast to /u/, the spectral movement patterns of /i/ showed very similar patterns for all speakers, regardless of study, language, and even age groups. All three studies, Hillenbrand *et al.,* (1995), Nearey & Assmann (1986), and the current study, found that /i/ had minimal spectral movement, and the current study showed that there were no significant differences across language or age-group in the formant trajectories of /i/. The stable pattern of /i/, as opposed to the dialectal and language-specific patterns of /u/, may be due to the ease of articulatory demands for /i/, which requires less precise positioning of the tongue body, as compared to /u/ which requires more precise tongue positioning as well as coordination with lip rounding (Perkell, 1996).

Future research is needed to confirm the results reported in this chapter, as the patterns observed in this study could have been influenced by other factors, such as consonantal context. Coarticulation between word-initial consonants and vowels could have affected the formant trajectory patterns of vowels, especially F2 trajectories (Stevens and House, 1963), although this study tried to minimize coarticulatory effects by excluding the first 18ms of the vowel and intervals beyond the vowel midpoint from the analysis. Korean vowels were much shorter in duration than English and Greek vowels (mean vowel duration by vowel, language, and age group is provided in Appendix B). The shorter duration of Korean vowels could result in more gestural overlap (more coarticulation) with the preceding consonant than in the case of longer vowels (Weismer & Berry, 2003). The minimal spectral movement observed in Korean vowels may reflect greater consonant-vowel coarticulation in Korean as compared to English and Greek. That is, language-specific vowel spectral movement patterns could also be reflecting language-specific degrees of coarticulation rather than language-specific vowel-inherent spectral movement patterns.

The age group comparison examined in this study also confirmed what has been reported in previous studies (Hillenbrand & Nearey, 1999; Assmann & Katz, 2000); vowels produced by young children showed similar spectral movement patterns as those of adults of their native languages. This result indicates that children as young as 2 years are capable of producing vowels in an adult-like manner, especially in the way they approach “target” formant frequencies. This developmental pattern, however, can also reflect young children’s ability to produce a similar degree of coarticulation as adult speakers (Katz, Kripke, & Tallal, 1991), as discussed above. Either way, however, the similarity between adults’ and children’s vowel spectral patterns suggests that children are attending more to the time-varying information rather than information at a single time point. The findings of Chapter 2 showed that the F1 and F2 values measured at the temporal midpoint of vowels produced by young children showed some degree of language-specificity, but their patterns were more variable than those of adults’ and some significant differences across age groups were found. For vowel spectral movement patterns, however, in general, no significant differences across age groups were found for any of the five vowels for any of the three languages. This result suggests that children can produce adult-like consonant-vowel movement patterns before they learn to refine the phonetic details of the formant targets (Hawkins, 1984).

One limitation of this study is that, as discussed above, vowel spectral movement patterns reported here could have been influenced by different factors that also influence vowel spectral movement patterns, such as the consonant context. In order to better understand the nature of vowel spectral movement patterns, therefore, analysis of vowel stimuli that are better controlled in terms of consonantal context (in a context that can minimize the effect of consonant environment on vowel spectral movement patterns, such as the /hVd/ context in English) would be helpful. Moreover, this study showed that the spectral movement patterns of vowels varied from one language to another. This implies that the relationship between time-varying spectral change and listeners' ability to identify target vowels could also differ by languages. In other words, listeners who are native speakers of languages that have minimal vowel spectral movement, such as Korean listeners, might respond differently to the perception tasks of Strange *et al.* (1976) and of Nearey & Assmann (1986).

In summary, this study contributed to the current literature on vowel spectral movement with two important findings. First, there are significant cross-linguistic differences in spectral movement patterns of shared vowels across three different languages (English, Greek, and Korean). It is hypothesized that these differences explain, at least in part, the differences in goodness ratings of adults of native and non-native vowels that were observed in Chapter 3. Also, the spectral movement patterns of vowels produced by 2- and 5-year-old children in each language did not differ significantly from those of adults.

CHAPTER 4: REFERENCES

Assmann, P. F. & Katz, W. F. (2000). Time-varying spectral change in the vowels of children and adults. *Journal of the Acoustical Society of America,* 108(4)*,* 1856-1866.

Barr, D.J. (2007). Analyzing 'visual world' eyetracking data using multilevel logistic regression, *Journal of Memory and Language,* 59, 457-474.

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

Boersma, P. & Weenink, D. (2006). Praat: doing phonetics by computer [Computer program]. Retrieved Aug, 2006, from http://www.praat.org/.

Chung, H., Kong, E-J., & Weismer, G. (2010). Vowel formant trajectory patterns for shared vowels of American English and Korean. *Journal of the Korean Society of Speech Science*, *2(4)*, 67-74.

Clopper, C., & Pisoni, D. B. (2006). The Nationwide Speech Project: A new corpus of American English dialects. *Speech Commutation.* 48, 633-644.

Fridland, V. (2008). Patterns of /uw/, /ʊ/, and /ow/ fronting in Reno, Nevada. *American Speech*, *83 (4),* 432-454.

Gay, T. (1970). A Perceptual Study of American English Diphthongs. *Language and Speech*, 13, 65-88.

Harrington, J., Kleber, F., & Reubold, U. (2008). Compensation for coarticulation, /u/-fronting, and sound change in standard southern British: An acoustic and perceptual study. *Journal of the Acoustical Society of America,*123, 2825-2835.

Hillenbrand, J. M. & Gayvert, R. T. (1993). Identification of steady-state vowels synthesized from the Peterson and Barney measurements. *Journal of the Acoustical Society of America,* 94(2)*,* 668-674.

Hillenbrand, J. M., Getty, L. A., Clark, M. J. & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America,* 97*,* 3099-3111.

Hillenbrand, J. & Nearey, T. M. (1999). Identification of resynthesized /hVd/ utterances: Effects of formant contour. *Journal of the Acoustical Society of America,* 105 (6), 3509-3523.

Jacewicz, E., & Fox, R. A. (2011). Perceptual distinctiveness of vowels in relation to dialectal sound change. Poster session presented at the *Acoustical Society of America* meeting, Seattle, WA, May 23-27.

Jenkins, J. J., Strange, W., Edman, T. R. (1983). Identification of vowels in 'vowelless' syllables*. Perception & Psychophysics.* 34 (5). 441-450.

Jun, S. A. (1993). *The phonetics and phonology of Korean prosody.* Unpublished doctoral dissertation, The Ohio State University, Columbus.

Katz, W. F., Kripke, C. & Tallal, P. (1991). Anticipatory Coarticulation in the Speech of Adults and Young Children: Acoustic, Perceptual, and Video Data. *Journal of Speech and Hearing Research*, 34, 1222-1232.

Mirman, D, Dixon, J. A., & Magnuson, J. S.(2008). Statistical and computational models of the visual world paradigm: Growth curves and individual differences, *Journal of Memory and Language,* 59, 475-494.

Nearey, T. M. & Assmann, P. F. (1986). Modeling the role of inherent spectral change in vowel identification. *Journal of the Acoustical Society of America,* 80 (5), 1297-1308.

Perkell, J. S. (1996). Properties of the tongue help to define vowel categories: hypotheses based on physiologically-oriented modeling. *Journal of Phonetics*, 24, 3–22.

R Development Core Team (2010). *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal analysis: Modeling change and event occurrence*. New York: Oxford University Press.

Stevens, K. N., & House, A. S. (1963). Perturbation of vowel articulations by consonantal context: An acoustic study. *Journal of Speech and Hearing Research*, 6, 111-128.

Strange, W., Jenkins, J. J., & Johnson, T.L. (1983). Dynamic specification of coarticulated vowels. *The Journal of the Acoustical Society of America*, *7*, 695-705.

Strange, W. (1989). Dynamic specification of coarticulated vowels spoken in sentence context. *Journal of the Acoustical Society of America, 85(5),* 2135-2153.

Tiffany, W. (1953). Vowel recognition as a function of duration, frequency modulation and phonetic context. *Journal of Speech and Hearing Disorders*, 18, 289-301.

Van Summers, W. (1987). Effects of stress and final-consonant voicing on vowel production: Articulatory and acoustic analyses. *The Journal of the Acoustical Society of America*, *82(3)*, 847-863.

Weismer, G. & Berry, J. (2003). Effects of speaking rate on second formant trajectories. *The Journal of the Acoustical Society of America*, *113(6)*, 3362-3378.

CHAPTER 4: APPENDIX

Appendix A. The output of the mixed-effects regression models for F1 trajectories of five vowels produced by adults, 5-year-olds, and 2-year-olds of English, Greek, and Korean. The significant terms are in bold.

			English			Greek			Korean		
			Estimate	SE	t value	Estimate	SE	t value	Estimate	SE	t value
/a/	Adults	Linear	0.22	0.02	**12.81**	0.21	0.02	**9.48**	-0.07	0.04	-1.66
		Quadratic	-0.11	0.01	**-11.25**	-0.10	0.01	**-10.97**	0.00	0.02	-0.22
	5yr	Linear	0.22	0.05	**4.46**	0.27	0.04	**6.20**	0.01	0.04	0.19
		Quadratic	-0.11	0.02	**-6.13**	-0.14	0.01	**-9.68**	-0.04	0.02	**-2.64**
	2yr	Linear	0.24	0.04	**6.39**	0.24	0.03	**7.24**	0.04	0.05	0.85
		Quadratic	-0.12	0.01	**-9.48**	-0.11	0.02	**-7.27**	-0.06	0.01	**-3.90**
			Estimate	SE	t value	Estimate	SE	t value	Estimate	SE	t value
/i/	Adults	Linear	-0.20	0.12	-1.66	0.07	0.01	**4.81**	0.00	0.04	0.07
		Quadratic	0.09	0.06	1.50	-0.03	0.01	**-4.76**	0.01	0.03	0.32
	5yr	Linear	-0.01	0.04	-0.37	0.03	0.02	1.72	3E-03	6E-02	0.06
		Quadratic	-0.04	0.02	**-2.77**	-0.03	0.01	-2.83	-1E-05	3E-02	0.00
	2yr	Linear	-0.02	0.08	-0.21	0.10	0.02	**6.12**	0.10	0.03	**3.33**
		Quadratic	-0.07	0.03	**-2.81**	-0.04	0.01	**-4.49**	-0.04	0.01	**-3.79**
			Estimate	SE	t value	Estimate	SE	t value	Estimate	SE	t value
/u/	Adults	Linear	0.06	0.03	1.98	-0.04	0.05	-0.89	0.02	0.04	0.62
		Quadratic	-0.08	0.02	**-3.71**	0.02	0.02	0.98	-0.01	0.02	-0.75
	5yr	Linear	0.07	0.06	1.20	0.00	0.02	0.13	0.09	0.03	**3.66**
		Quadratic	-0.02	0.02	-0.98	-0.02	0.01	-2.47	-0.06	0.01	**-5.04**

	2yr	Linear	0.11	0.06	1.84	-0.01	0.05	-0.13	-0.03	0.12	-0.22
		Quadratic	-0.11	0.04	-3.04	0.00	0.02	-0.18	0.02	0.06	0.37
			Estimate	SE	t value	Estimate	SE	t value	Estimate	SE	t value
	Adults	Linear	0.07	0.03	1.99	0.10	0.02	**4.73**	-0.04	0.03	-1.56
		Quadratic	-0.07	0.01	**-5.51**	-0.06	0.01	**-6.16**	-0.03	0.01	-1.72
/e/	5yr	Linear	0.05	0.06	0.79	0.09	0.03	**3.12**		NA	
		Quadratic	-0.09	0.03	**-3.49**	-0.07	0.01	**-5.77**			
	2yr	Linear		NA		0.16	0.04	**4.32**	0.03	0.02	1.22
		Quadratic				-0.09	0.01	**-8.35**	-0.05	0.02	-2.22
			Estimate	SE	t value	Estimate	SE	t value	Estimate	SE	t value
	Adults	Linear	0.05	0.03	1.66	0.08	0.02	**4.11**	0.03	0.02	1.40
		Quadratic	-0.08	0.01	**-8.03**	-0.05	0.01	**-6.83**	-0.03	0.01	-1.95
/o/	5yr	Linear	0.03	0.03	1.25	0.07	0.03	**2.33**	0.02	0.03	0.86
		Quadratic	-0.07	0.01	**-5.88**	-0.04	0.01	**-3.71**	-0.03	0.01	**-2.16**
	2yr	Linear		NA		0.10	0.04	**3.83**		NA	
		Quadratic				-0.09	0.02	**-5.40**			

Appendix B. The mean vowel duration (in *ms*) for each of five vowels

	2-year-olds			5-year-olds			Adults		
Vowel	English	Greek	Korean	English	Greek	Korean	English	Greek	Korean
/a/	175	228	173	149	172	132	146	128	105
/i/	144	180	101	141	135	92	131	103	69
/u/	306	175	120	255	116	84	261	90	77
/e/	165	217	130	164	168	114	196	121	64
/o/	NA	217	114	188	146	102	150	111	95
average	197.5	203.4	127.6	179.4	147.4	104.8	176.8	110.6	82

CHAPTER 5: CONCLUSION AND GENERAL DISCUSSION

This dissertation provides a comprehensive examination of the acoustic characteristics of vowels across five languages and three age groups. Vowel categorization and goodness ratings of these vowels by naïve adult listeners of three of these languages were also investigated in order to determine the relationship between these acoustic characteristics and adult vowel perception. There are three major findings from these three studies.

Study 1, a cross-linguistic vowel developmental study using the temporal midpoint values of F1 and F2, revealed that the acoustic characteristics of two of the three point vowels (/a/ and /u/) of five different languages systematically differed across languages. This result was especially apparent when acoustic characteristics of each of the three point vowels were compared in a speaker-specific vowel space. The examination of children's vowel productions in each of the five languages showed that children produced a language-specific vowel space as early as the age of 2. Evidence of ongoing phonetic refinement, however, was observed even for vowels produced by 5-year-olds. This finding suggests that, while vowels produced by 2-year-olds may be perceived as "correct" by adults, the process of reaching an adult-like, language-specific characteristic is gradual and not complete even by age 5.

Study 2, a cross-linguistic vowel perception study, revealed that adult listeners were capable of perceiving the subtle phonetic differences observed in Study 1. They rated native vowels as better examples of a vowel category than non-native counterparts. They also rated adults' vowels as better examples than children's vowels in the same phonemic category. In addition, listeners' goodness ratings of vowels was found to be highly related to the location of a vowel in a speaker-specific vowel space, especially for non-native vowels and some native vowels produced by 2-year-olds. On the other hand, this relationship was not found for native vowels produced by adults. Regardless of the acoustic values, native vowels were always rated as better examples than their non-native counterparts. This finding suggests that static F1 and F2 values measured at vowel midpoint cannot explain the variability in goodness ratings for native vowels and that additional acoustic information must be included. This additional information may also be relevant to the perceptual tendency to rate native vowels as better examples than non-native vowels that have been categorized similarly.

Finally, Study 3, a cross linguistic developmental study of vowel spectral movement patterns, revealed systematic cross-linguistic differences in vowels of the same phonemic symbols. The language-specific patterns were found even for the vowels of two different languages that had overlapping temporal midpoint F1 and F2 values (e.g., Greek /a/ and Korean /a/). This suggests that formant trajectories, together with temporal midpoint F1 and F2 values, may be important for perceptually differentiating native from non-native vowels and would be an interesting area for future research. This study also found that children produced adult-like formant trajectories as early as the age of 2. This finding implies that there is more than just the "target" information in the acoustic signal, and that children are sensitive to these fine phonetic details and reproduce them in production.

Several factors complicate the interpretation of results of this study and should be addressed in future work. The vowels analyzed in this study were always in word-initial position in CV or CVC contexts. The effect of different syllable structures on vowel formant patterns should also be examined. Furthermore, the word-initial consonants were always alveolar or velar obstruents. It is likely that word-initial labial consonants (or /h/) would have resulted in less consonant-to-vowel coarticulation and would be helpful in future research designed to tease apart vowel inherent spectral change from the effect of different consonants on vowel formants. Finally, it was not possible to distinguish between the effect of the initial consonant and the vowel on the vowel goodness ratings in Study 2. A different design (natural vowels in isolation, or presenting CV's from different languages in blocks) might be better suited to minimizing consonant effects on vowel goodness ratings.

Overall, the findings of the three studies suggest the importance of fine-grained acoustic analysis to understand cross-linguistic differences in vowel production and vowel acquisition of young children. When the only analysis method is phonetic transcription, it seems as if vowels are acquired very early. However, the acoustic analyses used in this study revealed that while F1 and F2 trajectories are adult-like by age 2, vowel formant targets are sometimes not adult-like, even by age 5. That is, transcription analysis by itself could not show how gradual and relatively complicated is the process of phonetic refinement of vowel acquisition. The result from Study 2 provided additional evidence showing the importance of fine-grained acoustic analysis in speech perception research. Listeners were sensitive to fine phonetic details, related to both language and age, in vowels produced by different groups of speakers. This suggests that subphonemic fine phonetic details shown in vowel acoustic studies are also relevant in perception, even by naïve listeners.