Working Papers of the Cornell Phonetics Laboratory
1995, vol. 10, pp. 165-189

Non-native Productions of Thai:Acoustic Measurements and Accentedness Ratings*

Ratree Wayland

In this study, productions of 10 Thai words with all five tones, namely high, mid, low, falling and rising were elicited from 3 native Thai speakers and 6 native English speakers, who had learned Thai in adulthood. A variety of acoustic measurements including voice onset time of the initial consonant, vowel length, the first and second vowel formants (F1, F2) , peak, valley and range of the fundamental frequency (F0) of all the five tones, were taken on both native and non-native productions. The two groups were found to differ in almost all acoustic parameters measured, but significantly in the F1, F2 and F0 valley values. Two native and all 6 non-native productions were rated for accentedness by three native listeners. The rating data suggested that the non-native production can be readily distinguished from the native production. Only some non-native production of some of the target words were considered to be native-like. More importantly, amount of the experience with Thai did not seem to affect the rating data. When the acoustic data were regressed on the means ratings, it was found that significant predictors varied from word to word (or from to tone).

1. Introduction

Pronunciation is an aspect that one tries to master when one engages in learning a second language (henceforth L2). In order to be understood, one would try to phonetically approximate the pronunciation of the target language. However, the empirical evidence now available suggests that only a few people who learn a second language beyond early childhood manage to pronounce L2 sounds in a native-like manner when their speech is carefully examined (McLaughin 1978). The reasons for this incompleteness of phonetic learning among adults are not well understood. The age of first exposure and the amount of second language experience have been claimed to be the two important factors; thus the focus of study, in the literature (Fathman 1975; Oyama 1982a; Ervin Tripp 1974; Cochrane and Sachs 1979).

Some researchers have tried to approach the foreign-accented phenomenon by looking at the relationship between sounds or phones in the sound systems of L1 and L2. One such approach is the Speech Learning Model developed by Flege (1987, 1988, 1991). According to this model, specific predictions can be made about which sounds or phones of L2 would pose difficulty for an L2 learner. Specifically, it is predicted that L2 phones that are 'similar' to existing L1 phones will be more difficult to learn, while 'new phones' (phones which fall outside the phonetic inventory of L1) will eventually be mastered.

* I would like to thank Professor John F. Hartmann from the Department of Foreign Languages and Literatures, Northern Illinois University at DeKalb for assisting me in getting the recordings of four non-native Thai speakers as well as conducting informal interviews to obtain their language background.

However, as pointed out by Munro (1993) "the successful evaluation of this or any other approach to the study of accented speech is contingent upon the availability of adequate descriptive data which characterize the patterns of "errors" made by L2 learners" (Munro 1993; p. 39). Information concerning which sounds are mispronounced as well as nature of the mispronunciation are equally crucial to researchers.

According to Flege (1988), there is the "phonetic norm" of a language. Unlike "pronunciation norm" which refers to "the collective judgment of native speakers concerning how a sound ought to be pronounced", "phonetic norm" is based on physical measurements of specific aspects of sound production" (p. 229). For example, a phonetic norm for stop consonants in English may be partially specified in terms of voice-onset time (henceforth VOT), that is the time interval between the release of the closure and the beginning of the vibration of the vocal folds. The VOT of English [t] in pre-stressed position might be 80 ms in a particular context, as compared to 15 ms in comparable speech material produced by monolingual native speakers of French (Flege 1988). Thus, a foreign accented speech can be defined as speech which differs acoustically (phonetic norm) from the native norms, and is auditorily detectable by native speakers (pronunciation norm). The speech productions of L2 learners can be compared to these norms to determine how native-like these productions are.

2. This Study

To my knowledge, no foreign-accented phenomenon of a tonal language has ever been investigated. Thus, this study is considered the first study involving an investigation of the production and perception of foreign-accented speech both on segmental and supra-segmental levels of a tonal language. Specifically, this study proposes to investigate the way in which two Thai vowels [a:] and [a:u],[1] the aspirated velar stop [k^h], and five tones, namely low, mid, high, falling and rising, produced by 6 male native speakers of English differ from 3 male native speakers of Thai, and how these differences influence perceived degree of accentedness by 3 female Thai native listeners.

This study consists of two experiments: the production and the perception experiments. In Experiment 1, the acoustic analysis is performed to see which acoustic parameters differ in the two sets of data. The acoustic parameters examined are vowel formant frequencies (F1, F2), vowel duration, voice-onset time of the aspirated velar stop [k^h], fundamental frequency (henceforth F0) peak, F0 valley and F0 range of all five tones. i.e. The F0 peak

[1] [a:u] can be phonologically analyzed as /a:/+final consoanant /w/; however, here it is phenetically treated as two consecutive vowel segments (diphthong). Thus, measurements were taken from both [a:] and [u] segments.

is the higest point in the fundamental frequency contour of a tone. The F0 valley is the lowest point in the fundamental frequency contour of a tone. The F0 range, is the difference between F0 peak and F0 valley.

In Experiment 2, three female native speakers of Thai listened to and rated the non-native speakers' production data on a five-point scale of accentedness. The purpose of this experiment is to investigate the relationship between acoustic parameters measured in Experiment 1 and the native judges' ratings in order to establish which properties of the non-native productions lead to a perception of accentedness. Rating of individual speakers' production are also examined to investigate individual differences in degree of accentedness.

3. Literature Review

A substantial number of researchers have investigated differences between native and non-native productions in terms of various acoustic parameters. For instance, studies by Flege (1980, 1987), Flege and Munro (1984), Flege, Munro and Skelton (1992), Caramazza, Yeni-Komshian, Zurif, and Carbone (1973), and Williams (1980) among others, have found that non-native speakers do not produce native-like voice-onset times. It was also found that second language learners of English produced very small differences in vowel durations before voiced and voiceless stops, (Flege et al 1992; Crawther and Mann 1992; Port and Mitleb 1983; Mack 1982). Some studies also compared formant frequencies of vowels produced by native and non-native speakers (Flege 1987; Flege and Hillenbrand 1984). Results indicated that the formant values of 'similar' L2 vowels produced by non-native speakers differed significantly from those produced by native speakers.

However, as pointed out by Munro (1993), one of the drawbacks of studies using only one approach to investigate the 'foreign-accented' phenomenon, i.e. only comparison of native and non-native production is that it does not reveal which characteristics of the non-native productions cause native listeners to hear them as having a foreign accent. Detailed acoustic analysis may reveal differences between native and non-native speeches on various parameters. However, these differences may influence native speaker's perception of accentedness to a varying degree. Thus, acoustic data should be related to perceptual data.

In the literature, several methods have been employed to evaluate degree of accentedness. In some studies (Flege and Hillenbrand 1984), native listeners were asked to identify a single phoneme excised from natural speech or to identify an entire word produced by non-native speakers. It was found that non-native words were less well identified in comparison to native tokens. Flege (1984) asked native speakers to judge whether utterances of various durations were produced by native or non-native speakers of English. These data indicated that the rate of correct detection of accent tended to be higher for relatively long intervals of speech (i.e. phrases) than short ones (single sounds), yet remained at significantly above-chance levels as the intervals were reduced from syllable-sized to segment-sized intervals, and finally to just the first 30 ms of /t/ (roughly, the burst portion of the sound /t/)" (Flege 1988, p.233). The capability of native listeners to detect "accentedness" in such a short interval of speech "indicates a high degree of sensitivity to divergences from native speaker norms" (Munro 1993, p. 41).

Various other techniques were also employed by researchers. In some cases, native listeners were asked to directly rate utterances on a scale of accentedness. However, no standard scale has been developed. Suter (1976) had native listeners score degree of accentedness of a two-minute utterance using a six-point scale. Snow and Hoefnagel-Höhle (1982) used a five-point scale to obtain rating of individual sounds in a list of words repeated after a recorded model. In order to obtain ratings for phones, intonation, rhythm and an overall pronunciation, Schneiderman et al (1988) used multiple five-point scales. Munro (1993) used a one hundred-point scale to obtain the accentedness rating for English vowels produced by Arabic native speakers.

The advantage or disadvantage of using one rating scale over another is not obvious in the literature. In research concerned with assessment of speech and voice quality, however, it was found that intrarater reliability was not related to the scale type or statistic used (Kreiman, Gerratt, Kempster, Erman and Berke 1993).

While the studies mentioned above used linguistically-trained judges, others have used naive native listeners (Flege and Eefting 1987; Flege 1988; Flege and Fletcher 1992). It has been shown that ratings obtained from linguistically-trained and naive listeners are fairly consistent to one another (Brennan and Brennan 1981; Cunningham-Andersson and Engstrand 1989).

4. Thai

Thai or Siamese is the national language of Thailand spoken by approximately 60 million people. The dialect spoken in the capital city of Bangkok is considered the standard one. However, outside Bangkok and the central plains, other languages of the Tai and Mon-Khmer family coexist with this standard dialect.

4.1 Consonants

An interesting feature of the Thai consonant system is the existence of three classes of stop consonants, namely voiced, voiceless aspirated and voiceless unaspirated. All consonants can occur initially, but only voiceless unaspirated stops, nasals and semi-vowels can occur finally. All final consonants are phonetically unreleased.

	Bilabial	Inter-dental	Alveolar	Palatal	Velar	Glottal
Stops						
Voiced	b		d			
Vls.unasp	p		t	c	k	
Vls.asp	p^h		t^h	c^h	k^h	
Fricatives		f	s			h
Sonorants						
Nasals	m		n		ŋ	
Laterals			l			
trill/tap			r			
Semi-vowels	w			j		

4.2 Vowels

	Front	Back-unrounded	Back-rounded
Hi	i	ɨ	u
Mid	e	ə	o
Low	ɛ	a	ɔ

Diphthongs ia ɨa ua (Hudak, 1990)

Each of these 9 vowels may occur phonemically as short or long, for example short [a] and long [a:]. Phonetically, the long vowels are approximately twice as long as the short vowels. All 18 nuclei may occur alone, with an initial consonants, with a final consonant or with both initial and final consonants.

4.3 Tones

Each syllable in Thai carries one of the five phonemic tones, namely a mid tone ([k^ha:] 'to be stuck or lodged in); a low tone ([k^hà:] 'galanga, a kind of aromatic root used in Thai cooking'); a falling tone ([k^hâ:] 'I, servant'); a high tone ([k^há:] 'to engage in trade'); a rising tone ([k^hǎ:] 'leg'). These five tones may be characterized in terms of pitch contour, pitch height, and glottalised or non-glottalised voice quality. This can be schematically presented as below (adapted from Hudak 1990, p. 34).

Tones	Tone mark	Pitch contour	Pitch Height	Voice quality
mid	unmarked	level	medium	non-glottalised
low	`	level	low	non-glottalised
falling	^	contour	low to high	glottalised
high	´	level	high	glottalised
rising	ˇ	contour	low to high	non-glottalised

5. The Experiment

5.1 Experiment 1: Acoustic Measurements

This experiment is conducted to establish the differences between native and non-native Thai productions. Acoustic parameters measured including vowel formants (F1, F2, F1-F2), VOT, vowel duration, F0 peak, F0 valley and F0 range

5.1.1 Methods

Subjects: The participants were three male native Thai speakers, all recruited from the student population of Cornell University, and 6 male native English speakers. Four were recruited from Northern Illinois University at DeKalb, and the other two from Cornell University. An informal interview was conducted to obtain information on each English-native speaker's language background. All three native Thai speakers are from Bangkok area and are between 19-24 years of age. Two have been living in the United states for the past three years, and the other for six years. They are designated as speakers 1, 2 and 3 in this study. The non-native group consisted of 6 native English speakers, between 20 and 40 years of age. They are designated as speakers 4 to 9 in this study.

5.1.2 Language Background

In this section, language background of 6 native speakers of English obtained from an informal interview is presented. They are arranged according to number of years of experience with Thai.

Speaker 6 Speaker 6 has been speaking Thai for more than 12 years. 30-40% of his daily conversation is carried out in Thai. He spends approximately 6 hours per week reading Thai. He also was a teacher of Thai, and had spent approximately 6 weeks in Thailand.

Speaker 4 Speaker 4 received language training through the U.S. Peace-Corps in Thailand, and spent an additional 3 years working there. He has been speaking Thai for the past 9 years, and approximately 20% of his daily conversation is carried out in Thai. He also understands and speaks some Lao and Khmer (Cambodian).

Speaker 7 Speaker 7 has been speaking Thai for almost 8 years. He spends approximately half an hour per day reading Thai, but currently does not have much opportunity to speak. He understands some Lao, and had lived in Thailand for two years.

Speaker 9 Speaker 9 has been speaking Thai for 6 years. His daily conversation at home is carried out in Thai. He has studied Pali, Sanskrit and Arabic. Altogether, he has spent approximately one year in Thailand.

Speaker 8 Speaker 8 has been learning and speaking Thai for three and a half years. He spends around 12 to 15 hours reading Thai, but does not have much opportunity to speak, except for 3 hours per week in the Thai class that he is currently taking. He also knows Spanish and some Tamil.

Speaker 5 Speaker 5 also received training through the U.S. Peace-Corps in Thailand, and spent approximately 2 years working there. He has also been auditing the advanced Thai class offered at Cornell University. However, he hardly has any opportunity to speak Thai on a daily basis, yet has been reading on a regular basis.

5.1.3 Materials and Procedure

Recordings for all three Thai-native speakers and two English-native speakers recruited from Cornell University were made in a sound-proof booth, using a cardioid microphone (Electrovoice, model RE 20) and high quality cassette recorder (Marantz, Model PMD 222). Subjects were asked to read a list of [KHA:U], and [NA:] words, with all five tones in a sentence frame " I said the word.....". (see word list below). Each word was repeated twice in a random fashion. The words, as well as the carrier phrase were written in Thai, using standard Thai orthography.

The recordings of four English-native speakers from Northern Illinois University at DeKalb were made at the main sound-proof studio of the University, using a cordioid microphone (Electrovoice, Model RE 20) and a cassette recorder (Tascam Model 122). The recoding process was carried out under the supervision of the studio technician. The same word list was used.

5.1.4 Word list

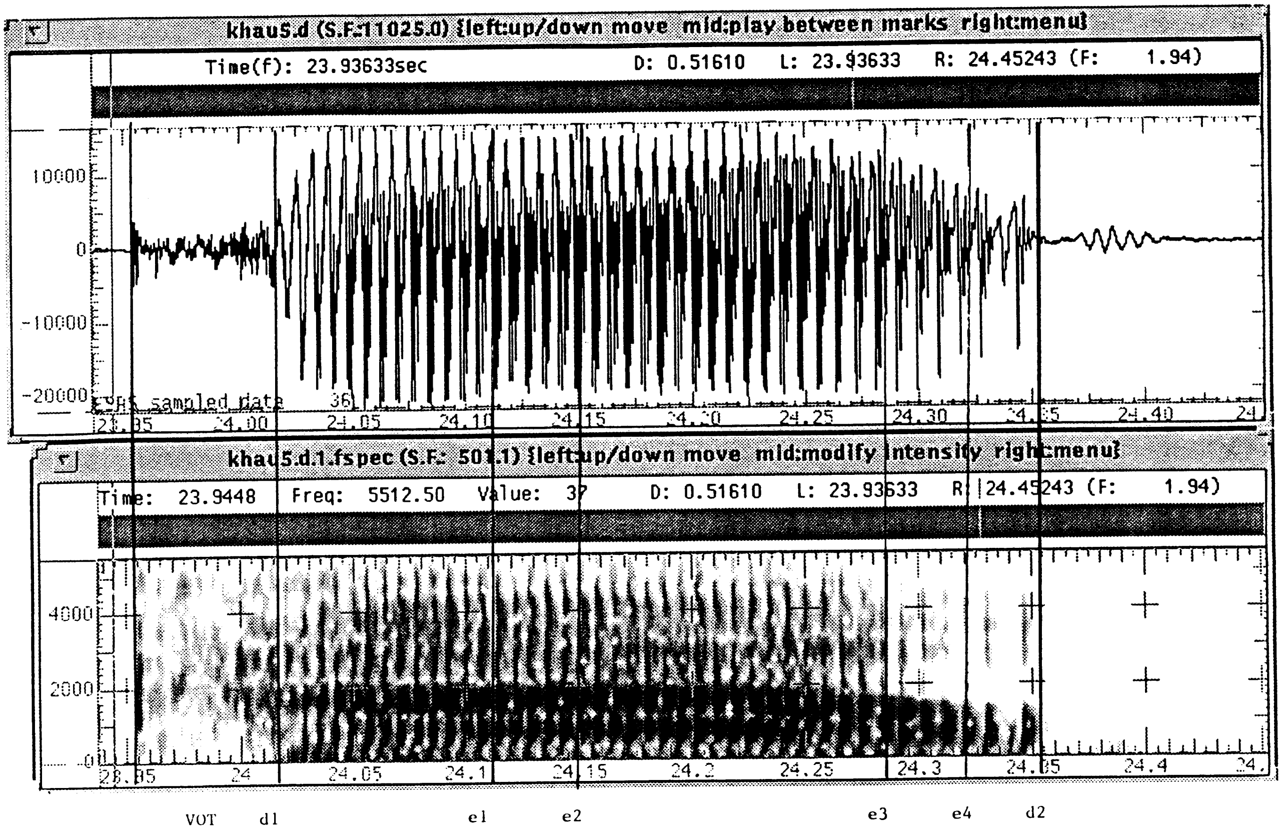
คาว	[kʰa:u]	'fishy smell'	นา	[na:]	'rice field'
ข่าว	[kʰà:u]	'news'	หน่า	[nà:]	'a nick name'
ข้าว	[kʰâ:u]	'rice'	หน้า	[nâ:]	'face'
ค้าว	[kʰá:u]	'a kind of fish'	น้า	[ná:]	'mother's younger siblings'
ขาว	[kʰǎ:u]	'white'	หนา	[nǎ:]	'thick'

The recordings were digitized on a SUN 3/160 computer at 11 kHz. with a low-pass filter setting of 6 kHz and stored as filed to be processed by the commercial software package WAVES+. This speech analysis package enabled us to simultaneously examine wave forms and (wide-band) spectrograms of each token. The target words were extracted from the carrier phrase and stored as separate files. These edited token were then submitted to LPC (Linear Predictive Coding) analysis using Hamming window of 25.6 ms, with 16 poles and preemphasis of .98. F1, F2 values were measured from the steady states of the vowels by placing the cursors on the formant tracks.

For the diphthong [a:u], F1 and F2 values were measured from these two steady state portions displayed on the wide-band spectrogram (e1, e2 and e3, e4 in Figure 1).

Vowel duration was measured by positioning the cursors at the vowel onset and offset (d1, d2). Vowel onset was taken to be the onset of the periodicity in the wave form. Vowel offset was indicated by the loss of F2 on the spectrogram.

The voice-onset time of the aspirated velar stop [kʰ] was also measured.



khau5.d (S.F.:11025.0) {left:up/down move mid:play between marks right:menu}

Time(f): 23.93633sec D: 0.51610 L: 23.93633 R: 24.45243 (F: 1.94)

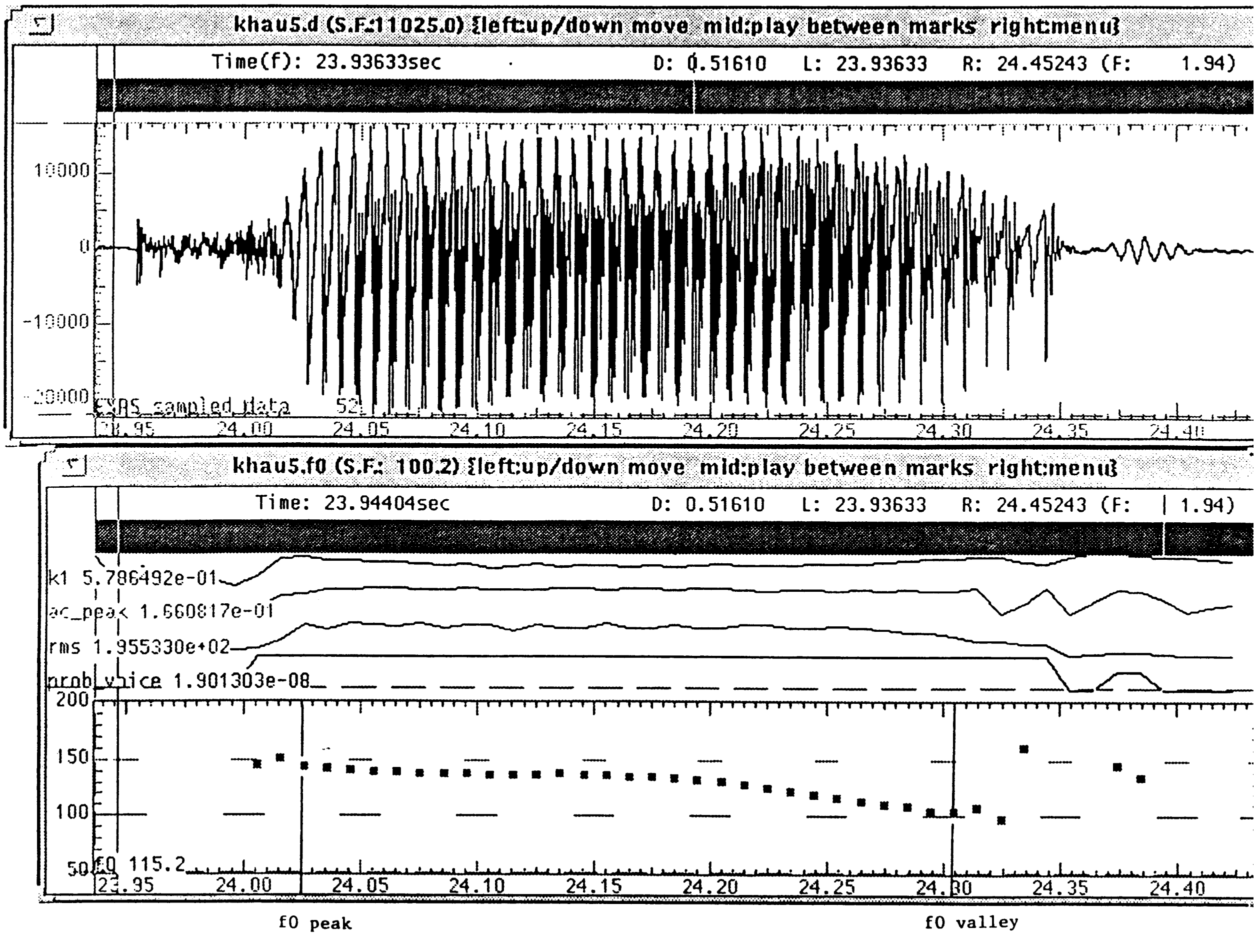
khau5.d.1.fspec (S.F.: 501.1) {left:up/down move mid:modify intensity right:menu}

Time: 23.9448 Freq: 5512.50 Value: 3 D: 0.51610 L: 23.93633 R: 24.45243 (F: 1.94)

Figure 1. Segmented and labeled waveform and wide-band spectrogram of the word [kʰâ:u] 'rice' as produced by Speaker 1. VOT is voice-onset time, d1 marks the begining of the vowel, d2 marks the end of the vowel, e1 e2 marks the steay state of [a:] and e3 e4 marks the steady state of [u].

A pitch track (a representation of fundamental frequency over time) was derived for all of the tokens, producing a data point every 5 msec with a sampling frequency of 11 kHz and an LPC order of 16. Two F0 measurements were recorded as data for each token. These included the F0 peak and F0 valley. F0 peak was the highest F0 value on the F0 contour, and F0 valley was the lowest F0 value. F0 range was the difference between F0 peak and F0 valley.

Following Cooper & Sorenson (1981), to minimize the chance of including spurious F0 estimates in the data, F0 peak and F0 valley measurements were taken in the location where two or more contiguous occurrences of the same F0 value were observed (see Figure 2).



khau5.d (S.F.:11025.0) {left:up/down move mid:play between marks right:menu}

Time(f): 23.93633sec D: 0.51610 L: 23.93633 R: 24.45243 (F: 1.94)

khau5.f0 (S.F.: 100.2) {left:up/down move mid:play between marks right:menu}

Time: 23.94404sec D: 0.51610 L: 23.93633 R: 24.45243 (F: 1.94)

Figure 2. Waveform and pitch track of the word [kʰâ:u] 'rice' as produced by Speaker 1. F0 peak marks the highest F0 value of the pitch contour (tone), and F0 valley marks the lowest f0 value

5.1.5 Results and Discussion

5.1.5.1 Duration data

For the word [KHA:U] pronounced with all five tones, both vowel and voice-onset time duration values are log-transformed and submitted to a multivariate analysis with native language as an independent variable. The analysis reveals no significant language effect for either voice-onset time duration [p=.271], or vowel duration [p=.90]. Similar procedure is carried out for the vowel duration of all five tones of the word [NA:]. The analysis reveals no significant effect for native language [p=.145]. The mean durations in msec of both vowels (i.e. [a:] in [NA:] and [a:u] in [KHA:U]), and voice-onset time durations for [kʰ] in [KHA:U] are given in Table 1 below.

Table 1. Mean vowel durations (in msec) and voice-onset time durations (in msec) for native Thai group and native English group. The means are based on 2 repetitions of each of the five tones from each speaker. The means of all five tones are shown in bold. Standard deviations are given in the parentheses.

Vowels	Tones	VOT (msec.)		Duration (msec)	
		Native Thai	Native English	Native Thai	Native English
[a:u]	[k^{h}a:u] mid	90	94	391	411
	[k^{h}à:u] low	88	89	347	386
	[k^{h}â:u] falling	68	90	374	417
	[k^{h}á:u] high	84	79	384	409
	[k^{h}ǎ:u] rising	69	86	376	427
		80 (11)	**88 (6)**	**374 (17)**	**410 (15)**
[a:]	[na:] mid			371	397
	[nà:] low			315	388
	[nâ:] falling			343	361
	[ná:] high			366	370
	[nǎ:] rising			358	401
				351 (23)	**383 (17)**

5.1.5.2 Spectral Data

All spectral data for the word [KHA:U], namely vowel formants (F1, F2 and F2-F1) for the first and second elements of the diphthong [a:u] as well as F0 peak, F0 valley and F0 range were log-transformed and submitted to a multivariate analysis of variance with native language as an independent variable. The analysis revealed a significant effect on F2 [p<.023] and F2-F1 [p<.043] of the first element, i.e. [a:] of the diphthong [a:u], F1 of the second element i.e. [u] of the diphthong [a:u] [p <.003], and F0 valley [p<.006].

To examine individual differences, Student-Newman Keuls post-hoc analysis was performed. This analysis revealed that in general three native Thai speakers, namely speakers 1, 2 and 3 formed a homogeneous group, while differences among native English speakers existed in almost all acoustic parameters. No general pattern or conclusion could be drawn from the native English speakers' production.

Similarly, all spectral data measured from all five tones of the word [NA:] were log-transformed and submitted to a multivariate analysis of variance with native language as an independent variable. The analysis revealed a significant effect on F1 [p<.009], F2 [p<.000], and valley [p<.009].

Similar to [KHA:U], when student-Newman Keuls post-hoc analysis was performed, the results suggested that while native Thai speakers may form a homogeneous group, no general pattern emerged from the native English speaker group. Their production of Thai appeared to differ from one another in nearly all acoustic parameters measured.

The mean values of F1, F2 for vowel [aː] in [NA:] are given in Table 2, and the mean values of F1, F2 of the first and second element of the diphthong [aːu] are given in Table 3. Table 4 illustrates the mean values of F0 peak, F0 valley and F0 range for both [NA:] and [KHA:U].

Table 2. Mean F1 and F2 values (in Hz) of the vowel [aː] for native Thai and native English groups. The means were based on two repetitions of each of the five tones from each speaker. Standard deviations are given in parentheses.

		Native Thai		Native English	
vowel	Tones	F1	F2	F1	F2
[aː]	[naː]	783	1536	726	1418
	[nàː]	818	1532	722	1392
	[nâː]	740	1593	706	1415
	[náː]	749	1547	675	1408
	[nǎː]	793	1503	721	1384
		777 **(32)**	**1542** **(33)**	**710** **(21)**	**1403** **(15)**

Table 3. Mean F1 and F2 values (in Hz) of the first and second element of diphthong [a:u] (i.e. [a:] and [u]) of all five tones of the word [KHA:U] for native Thai and native English groups. The means were based on two repetitions of each five tones from each speaker. Standard deviations are given in parenthesis.

	Native Thai				Native English			
	[a:]		[u]		[a:]		[u]	
Tone	F1	F2	F1	F2	F1	F2	F1	F2
mid	817	1408	575	946	785	1334	614	975
low	819	1407	608	987	796	1322	654	1057
falling	804	1489	579	1006	794	1387	679	1065
high	822	1421	456	889	744	1340	606	973
rising	827	1416	482	920	795	1326	611	960
	818 (9)	**1428** (34)	**540** (67)	**950** (48)	**783** (22)	**1342** (26)	**633** (32)	**1006** (51)

Table 4. Mean value of F0 peak, F0 valley and F0 range (in Hz) of both target words [KHA:U] and [NA:] spoken with five tones for native Thai and native English groups. The means are based on two repetitions of each of the five tones from each speaker. Standard deviations are given in parentheses.

Native Thai				Native English		
word	**Peak**	**Valley**	**Range**	**Peak**	**Valley**	**Range**
[k^{h}a:u]	128	97	31	132	109	23
[k^{h}à:u]	117	76	40	127	101	26
[k^{h}â:u]]	140	82	58	161	97	65
[k^{h}á:u]	137	113	24	162	127	39
[k^{h}ǎ:u]	147	85	62	148	102	46
	134 (12)	**91 (15)**	**43 (17)**	**146(16)**	**107(12)**	**40 (17)**
[na:]	120	99	21	130	108	21
[nà:]	114	75	38	123	100	23
[nâ:]	135	82	54	154	102	52
[ná:]	127	109	19	150	115	50
[nǎ:]	145	91	54	145	102	43
	128(12)	**91 (13)**	**37 (17)**	**140(13)**	**105 (6)**	**38 (15)**

5.1.6 Summary

In summary, the non-native production of Thai differ from the native-Thai production in several acoustic parameters measured. The average VOT of [k^{h}] in [KHA:U], and vowel duration (both [a:u] and [a:]) produced by native English speakers were longer than those produced by native Thai speakers (Table 1). Vowel formants (F1, F2) of [a:] in [NA:] produced by native English speakers were lower than those produced by native Thai (Table 2). However, when a statistical analysis was performed, it was found that for [KHA:U], the two groups differed significantly in:

1. F1 value of the first element of the diphthong [a:u] (i.e. [a:]); the native English value was lower than that of the native Thai (783 vs. 818 Hz.).

2. F1 value of the second element of the diphthong [a:u] (i.e. [u]). The native English value was higher than that of the native Thai (540 vs. 633 Hz.).

3. F2-1 (the difference between F1 and F2) of the first element, i.e. [a:] of the diphthong [a:u]. The native English value was lower than that of the native Thai (559 vs. 610 Hz.).

4. F0 valley value of all five tones. The native English values were higher that those of the native Thai (107 vs. 91 Hz.)

For [NA:] the two groups differed significantly in:

1. F1 value of [a:]. Native English's value was lower than that of the native Thai (710. vs. 770 Hz.).

2. F2 value of [a:]. Native English's value was lower than that of the native Thai (1403 vs. 1542 Hz.).

3. F0 valley values. Native English's value was higher than that of the native Thai (105 vs. 91 Hz.).

However, individual differences also existed. In general, it was found that native Thai speakers, i.e. speakers 1, 2 and 3, form a homogeneous group, while significant differences in almost all acoustic parameters measured existed among native English speakers.

5.2 Experiment 2: Accentedness Judgments

In Experiment 2, the acoustic measurements from Experiment 1 were regressed on a set of accentedness ratings assigned by three judges in order to determine the properties associated with perceived accentedness in all of the native English data. As pointed out in Munro (1993), an implicit assumption in this kind of experiment is the idea that native speakers have (implicit) knowledge of a good exemplar of all or most of acoustic parameters of the language. For example, a native Thai speaker would know what a good exemplar of a high tone should sound like, what the appropriate VOT duration for the unaspirated [k] is, and so on. When s/he assigns accentedness rating, s/he would make reference to such knowledge. The exact nature of this process, however, is not well understood. One possible explanation is that tokens being rated are compared with an abstract prototype (Flege,1984). Thus, tokens may be rated as having varying degrees of accentedness depending on the extent of the difference in all or some acoustic parameters between rated tokens and the 'prototypes'. If the prototype account of perceived accentedness is correct, the relationship between acoustics properties and rating data might be observed by using multiple regression analysis.

5.2.1 Methods

Two replications of each repetition (from experiment 1) of all 10 target words , i.e. all five tones of the words [NA:] and [KHA:U] (10 words X 2 repetitions X 2 replications of each repetition X 8 speakers =320 in total) were recorded in random order. Data from all six English-native and two Thai-native speakers, namely speaker 1 and 3 were used. Ten practice trials were also included. The stimuli were presented in a block of ten replications with inter trial interval of 3 sec and an interblock interval of 5 sec. The native data were included to ensure that the listeners could distinguish between native and non-native speakers.

5.2.2 Procedure

Stimuli were presented to three female native-Thai judges recruited from the undergraduate student population of Cornell University. None of them were linguistically-trained. The three judges were from the central, southern, and northeastern regions of Thailand respectively. All reported normal hearing. The stimuli were played on the Aiwa cassette recorder (Model AD-F400), and were presented at a comfortable listening level over the Sony Dynamic Stereo Headphone MDR V6.

All three judges were given a print out of all 320 stimuli, printed in standard Thai script. A scale of one to five was also included to the right of each stimulus in the print out. A score of one indicated strongest degree of accentedness, and a score of five indicated native like production. To assign ratings, the judges were asked to give a score to each stimulus by circling number 1 to 5 on the scale, depending on degree of perceived accentedness.

To assess whether any of the judges differed significantly from the others in their rating scores assigned to the speakers, Pearson correlation coefficients were calculated for all the ratings for all possible pairs of the judges. Since the correlations obtained were in the range of 0.97 to 0.99 and were significant at the 0.01 level, it was concluded that the judges are significantly consistent with one another.

5.2.3 Results

The rating scores for [KHA:U] and [NA:] for both groups of speakers average across 3 judges are shown in Table 5. It is organized by words and native languages. Overall mean values of all five tones of both [KHA:U] and [NA:] are also given. The mean score for the native Thai are 4.84 for both [KHA:U] and [NA:]. For the native English, on the other hand, the mean score for [KHA:U] is 3.52, and 3.97 for [NA:]. On the average, the mean score for the non-native group is around 1 point lower than the native group.

Table 5. Overall mean score of each of the target words for native Thai and native English groups by 3 judges. The mean scores across five tones of each target word are given in bold. Standard deviations are given in the parenthesis.

Word	**Native Thai**	**Native English**	**Word**	**Native Thai**	**Native English**
k^haːu	4.88	3.44	naː	4.88	4.03
k^hàːu	4.71	3.35	nàː	4.63	3.69
k^hâːu	4.92	3.63	nâː	5.00	4.32
k^háːu	4.79	3.28	náː	4.75	3.44
k^hǎːu	4.88	3.89	nǎː	4.94	4.38
mean	**4.84** (.08)	**3.52** (.25)	**mean**	**4.84** (.15)	**3.97** (.40)

Data in Table 5 also shows that mean scores of level tones, namely mid, low and high are lower than those of contour tones, i.e. falling and rising. More interestingly, for the English-native group, rating scores from low to high assigned to each individual tone are in the following order: high tone, low tone, mid tone, falling tone and rising tone. This pattern is consistent for both sets of words, i.e. [KHA:U] and [NA:].

Table 6. Overall mean score of each target words for native-Thai and native-English by 3 judges. Speaker 1 and 3 are native-Thai and speakers 4-9 are native-English. Native-English speakers are arranged according to number of years of experience with Thai.

Speakers	**1**	**3**	**6**	**4**	**7**	**9**	**8**	**5**
Words	**Native-Thai**		**Native-English**					
k^{h}a:u	5	4.74	3.25	4.9	3.15	3.48	3.33	2.48
k^{h}à:u	4.8	4.55	2.55	4.83	2.97	2.4	4.25	3.09
k^{h}â:u	4.9	4.9	2.9	5	3.34	3.15	3.98	3.34
k^{h}á:u	4.75	4.84	2.34	4.4	3.34	3.09	3.08	3.34
k^{h}ǎ:u	5	4.74	3.75	4.25	3.49	3.55	4.67	3.55
Mean	**4.89**	**4.75**	**2.96**	**4.68**	**3.26**	**3.13**	**3.86**	**3.16**
na:	5	4.75	3.65	4.24	4.03	4.55	3.4	4.24
nà:	4.67	4.65	2.99	4.59	3.55	2.9	4.7	3.33
nâ:	5	5	3.24	4.65	4.58	4.49	4.25	4.67
ná:	4.74	4.75	1.9	4.9	3.48	3.9	3.4	2.98
nǎ:	5	4.83	4.5	4.08	4.66	4.17	4.55	4.24
Mean	**4.88**	**4.80**	**3.26**	**4.49**	**4.06**	**4.00**	**4.06**	**3.89**

Table 6 shows rating scores for each individual speaker. Speakers 1 and 3 are native Thai and speakers 4 to 9 are native English. The native English speakers' scores are arranged in order of number of years of experience with Thai. The mean score of each word average across the native English group ranges from 2.96 to 4.68 for [KHA:U] and 3.26 to 4.49 for [NA:]. For the native group, it ranges from 4.55 to 5 for [KHA:U] and 4.65 to 5 for [NA:]. However, an examination of individual score reveals that for the native-Thai group, the score ranges from 4.25 to 5, and from 1.75 to 5 for the native-English group.

Since the rating scores of each native-English speaker is arranged according to the number of years of experience with Thai, It is easy to see that there does not seem to be a strong correlation between the scores received and the amount of experience with the target language. Speaker 6, for example, had the greatest amount of experience with Thai, yet his scores were much lower than every other speaker in the group.

5.2.4 Multiple Regression Analysis

5.2.4.1 Predictor Variables

All acoustic parameters of the target words measured in the Experiment1, were transformed as described below and used as predictors in the stepwise linear multiple analysis. As pointed out in Munro (1993), the simple use of raw measurement values may not yield a correlation between the acoustic predictors variables with the rating data. A solution to this problem is to quantify the difference between the non-native data and the native data. The assumption is, of course, that the native data are close to the ideal values for all parameters measured. This, however, may not be the case since only a small sample of speakers were investigated.

In computing these variables, an attempt was made to characterize how much the rated non-native tokens differed acoustically from a good exemplar of Thai tokens. Thus, all predictors were calculated by subtracting mean values obtained from the native group (3 speakers) in Experiment1 from the corresponding mean values of each of the non-native speakers. the rated tokens. The values of vowel duration predictor for the word [KHA:U], for instance, were computed by subtracting the mean value of the native Thai group from the mean value of rated token of each speaker in the non-native group.

The remaining predictors were computed in the same manner. Each of the differences was then squared to yield a positive number in order that a relatively large value of any predictor would indicate a large discrepancy between it and the native mean, regardless of direction. The values of these predictors were taken to represent an assessment of the acoustic distance between the non-native tokens and a hypothetical, ideal native-like token.

5.2.4.2 Analysis

[KHA:U]

In a stepwise linear analysis, all 8 acoustic parameters measured from the target word [KHA:U], namely F1, F2, F2-F1, vowel duration, F0 peak, F0 valley and F0 range were used as predictors and the rating score was used as a dependent variable. The results indicated that for all five tones of this target word, only F0 valley and F2 of the second element, i.e. [u] of the diphthong [a:u] were significant predictors. F0 valley accounted for an additional 26% of the variance in the rating score, and F2 accounted for an additional 17%.

It is possible, however, that predictors for each individual tone or word will vary. Thus, a separate stepwise linear analysis was also performed for each individual tone of the word [KHA:U]. It was found that there was no single significant predictor for low tone [kʰàːu] 'news', falling tone [kʰâːu] 'rice', and rising tones [kʰǎːu] 'white'.

For mid tone [kʰaːu] 'fishy smell', the results indicated that F2-F1 of the second element of the diphthong [aːu] accounted for 75% of the variance in the rating data, vowel duration accounted for an additional 20%, F0 valley 4.8%, and F2-F1 of the first element of the diphthong [aːu], i.e. [aː]accounted for an additional .2%. These four predictors accounted for 100% of all the variance in the rating data.

For high tone [kʰáːu] 'a kind of fish', the analysis revealed that F0 range accounted for 71% of the variance in the rating data, F0 valley accounted for an additional 25% and F2-F1 of the first element of the diphthong [aːu], i.e. [aː] accounted for an additional 4%. All of these three predictors accounted for 100% of the variance in the rating data.

Table 7. Results of stepwise multiple regression analyses of the words [kʰaːu] 'fishy smell', [kʰáːu], 'a kind of fish', [naː] 'rice field', [nâː] 'face' and [náː] 'a younger sibling of mother'. Predictors that accounted for a significant portion of the variance in the rating data are given. They were in the order shown in the table.

		R^2
[kʰaːu] 'fishy smell'	1. F2-F1 (of the second element of the diphthong [aːu]	.75
	2. Vowel Duration	.95
	3. F0 Valley	.998
	4. F2-F1(of the first element of the diphthong [aːu]	1.00
[kʰáːu] 'a kind of fish'	1. F0 Range	.71
	2. F0 Valley	.96
	3. F2-f1 (of the first element of the diphthong [aːu]	.997
[naː] 'rice field'	F0 Valley	.68
[nâː] 'face'	F0 Peak	.68
[náː] 'younger sibling of mother'	F2	.89

[NA:]

Similarly, all acoustic parameters measured from the target word [NA:], namely F1, F2, F2-F1, vowel duration, F0 peak, F0 valley and F0 range were used as predictors in a stepwise multiple regression analysis. For all five tones, the results suggested that F0 valley was the significant predictor. It yielded relatively strong multiple R of .81 which indicated that it accounted for 65% of the variance in the rating data. A separate stepwise linear analysis was also carried out for each individual tone. The results indicated that for mid tone [naː] 'rice field', F0 valley yielded multiple R of .82 and accounted for 68% of the variance in the rating data.

For low tone [nàː] 'a nick name' and rising tone [năː] 'thick', there is no single significant predictor. For falling tone [nâː] 'face', F0 peak yielded a moderate multiple R of .83 and accounted for 68 % of the variance in the rating data. For high tone [náː] 'younger sibling of mother', F2 yielded a strong multiple R of .95 and accounted for 89% of the variance in the rating data.

In summary, for all five tones of both [KHA:U] and [NA:] , F0 valley predictor was moderately correlated with the judges' rating. In addition, for [KHA:U], F2 of the second element of the diphthong [a:u] was also another predictor. The predictors, however, varied from word to word, or from tone to tone. For mid tone [k^{h}aːu] 'fishy smell', the F2-F1, vowel duration, F0 peak and F0 valley were significant predictors. For high tone [k^{h}áːu] 'a kind of fish', significant predictors were F0 valley, F0 range and F2-F1 of the first element of the diphthong [aːu]. For low, falling and rising tones, there was no significant predictor.

For mid tone [naː] 'rice field', F0 valley was the significant predictor. For falling tone [nâː] 'face', F0 peak was the significant predictor, and for high tone [náː] 'a young sibling of mother', F2 was the significant predictor.

6. Conclusion

Unlike most previous studies on foreign-accented speech, this study investigated this phenomenon in a tonal language, i.e. Thai. Moreover, an attempt to relate production to perception data was also an important feature of this study.

When the production of Thai by 6 native-English speakers was carefully examined along several acoustic parameters in the first experiment, it was evident that the non-native production differed from the native production in all acoustic parameters measured. However, when a statistical analysis was performed, the results revealed that only differences in spectral data, namely vowel formants (F1, F2) and fundamental frequency data between the two groups were significant. Duration data, i.e. voice-onset-time and vowel duration, on the other hand, was found to be comparable between two groups of speakers.

When the native and non-native production data were judged for degree of accentedness on a five-point scale by three female native-Thai listeners in Experiment 2, the results indicated that, on the basis of pronunciation, non-native Thai utterance can be readily distinguished from the native-Thai utterance. On the average, rating scores of the non-native group was approximately 1 point lower than that of the native group. This seemed to suggest that the non-native production was not dramatically different from the native production, and that in quite a few cases the non-native tokens were heard as native-like. A few tokens produced by speaker 4, for example, received a score of 5 indicating native-like production. The fact that only single words produced in a short carrier phrase were investigated may have accounted for this finding. An examination of longer intervals of speech, i.e., a paragraph or various durations of speech produced in a natural setting, (as opposed to a laboratory setting) may offer different results.

When the mean score of individual tokens was examined, it was found that it ranged from 4.25 to 5 for the native-Thai group, and from 1.75 to 5 for the native-English group. The wider range of scores assigned to the non-native token may lead one to hypothesize that amount of experience with the Thai language may be the source of explanation. The data on Table 6 in which the rating scores were organized according to the speakers' number of years of experience with Thai proved to the contrary. On the average, speaker 6 who had the most experience with Thai received lower scores than speaker 4 and 8. Thus, rating scores did not seem to correlate with number of years of experience with the target language. Similar observation was also found in other studies (Flege and Fletcher 1992; Flege, Munro and Fox 1993, (cited in Munro 1993); Munro 1993). The evidence in these studies suggested that "experience with a second language beyond one year does not guarantee improvement in L2 production and perception, particularly when such factors as amount and quality of the L2 are not controlled for" (Munro 1993, p. 61). Moreover, even after several years of experience, a group of Arabic-speaking learners investigated under the laboratory condition in Munro (1993)'s study were found to be unable to produce any vowel from English in a truly native-like way.

Another important finding from this present study was that five different tones in Thai seemed to pose different degrees of difficulty for the native-English speakers. In general, level tone, namely high, mid and low tones received lower scores than contour tones, namely falling and rising. The pitch contrast between the first and second portions of falling tone (from high to low) and rising tones (from low to high) may have provided non-native speakers with an internal point of reference in the process of production and perception of these two tones. It was further found that among level tones, high tone received the lowest rating score, followed by low tone and mid tone. On the other hand, rising tone received higher score than falling tone.

In order to examine which characteristic of the non-native production investigated in Experiment 1 influenced native listeners to hear them as having a foreign accent, rating data were regressed on all acoustic parameter measured. Similar to Munro (1993)'s study, this analysis was found to be moderately successful. Significant predictors were found only for mid tone and high tone of the [KHA:U] words, and only for mid, falling and high tones for the [NA:] words. The 4 significant predictors found for [k^{h}a:u] 'fishy smell', the 3 significant predictors found for [k^{h}á:u] 'a kind of fish' accounted for 100% of the variance of the rating data of both words. Only one predictor each was found for [na:] 'rice field', [nâ:] 'face' and [ná:] 'younger siblings of mother, and it accounted for 68%, 68% and 89% of the variance in the rating scores of these three words (Table 7).

Results on Table 7 also showed that significant predictors varied from word to word or from tone to tone. In general, spectral (as opposed to duration) characteristic of the non-native production, namely vowel formants and fundamental frequency were found to have greater influence on native listeners' judgment. This appears to agree with the production results (Experiment 1) whereby significant differences were found in the spectral, but not duration characteristic of the production of the two groups. From Table 7, 9 out of 10 significant predictors found for both [KHA:U] and [NA:] were spectral in nature.

It should be emphasized, however, that the process of trying to relate production and perception data in order to establish which aspects of the non-native production influenced the perceived degree of accentedness by native listeners is rather exploratory. Further investigation along this line is needed to verify the patterns of results found in this study. Furthermore, our knowledge on the differences between native and non-native speakers' production and perception, thus on the foreign-accented phenomenon will be enhanced by an investigation on the perception of both native and non-native production by *non-native* speakers.

7. References

Brennan, E., & Brennan, J. (1981) Measurement of accent and attitude toward Mexican American Speech. *Journal of Psycholinguistic Research* 10, 487-501.

Caramazza, A., Yeni-Komshian, G., Zurif, E., and Carbone, E. (1973) The acquisition of a new phonological contrast: The case of stop consonants in French-English bilinguals. *Journal of the Acoustic Society of America* 54, 421-428.

Cochrane, R., & Sachs, J. (1979) Phonological Learning by children and adults in a laboratory setting. *Language and Speech* 22,145-49.

Cooper, W.E., and Sorenson, J.M. (1981) *Fundamental Frequency in sentence production.* New York: Springer Verlag.

Crowther, C., and Mann, V. (1992) Native Language factors affecting use of vocalic cues to final consonant voicing in English. *Journal of the Acoustic Society of America* 92, 711-722.

Cunningham-Andersson, U., and Engstrand, O. (1989) Perceived strength and identity of foreign accent in Swedish. *Phonetica* 46, 138-154.

Ervin-Tripp, T. (1974) Is second language learning like the first?.*TESOL quarterly* 8, 111-127.

Fathman, A., (1975) The relationship between age and second language productive ability. *Language Learning* 25, 245-253.

Flege,J. (1980) Phonetic approximation in second language acquisition. *Language Learning* 30, 117-134.

Flege, J. (1981) Factors affecting degree of perceived foreign accent in English sentences. *Journal of the Acoustic Society of America* 91, 370-389.

Flege, J., (1984) The detection of French Accent by American listeners. *Journal of the Acoustic Society of America* 76, 692-707.

Flege, J., & Hillenbrand, J. (1984) Limits on phonetic accuracy in foreign language speech production. *Journal of the Acoustic Society of America* 76, 708-721.

Flege, J., & Eefting, W. (1987) Cross-language switching in stop consonant perception and production by Dutch speakers of English. *Speech Communication* 6, 185-202.

Flege, J. (1987) The production of "new" and "similar" phones in a foreign language: Evidences for the effect of equivalence classification. *Journal of Phonetics* 15, 47-65.

Flege, J. (1988) The production and perception of foreign languages. In H H. Winitz (ed.) *Human Communication and its Disorders* (pp. 224-401), Norwood, NJ:Ablex.

Flege, J. (1991) The interlingual identification of Spanish and English vowels:Orthographic evidence. *The quarterly journal of Experimental psychology* 43A, 701-731.

Flege, J., Munro, M., and Skelton, L. (1992) Production of the Word-final English /t/-/d/ contrast by native speakers of Mandarin and Spanish. *Journal of the Acoustic Society of America* 92, 128-143.

Flege, J. & Fletcher, K. (1992) Talker and Listener effects on degree of perceived foreign accent. *Journal of Acoustic Society of America* 91, 370-389.

Hudak, T. (1990). Thai. In B. Comrie. (ed.) *Languages of East and South-East Asia*, Routledge:London.

Kreiman, J., Gerratt, B., Kempster, G., Erman, A., and Berke, G. (1993) Perceptual evaluation of voice quality: Review, tutorial, and a framework for future research.. *Journal of Speech and Hearing Research 36*, 21-40.

Mack, M. (1982) Voicing-dependent vowel duration in English and French: Monolingual and bilingual production. *Journal of the Acoustic Society of America* 71, 172-78.

McLaughlin, B. (1978). *Second language acquisition in childhood.* Hillsdale, NJ: Erlbaum.

Munro, M. J. (1993) Production of English Vowels by Native Speakers of Arabic:acoustic measurement and accentedness ratings. *Language and speech* 36 (1), 39-66.

Oyama, S. (1982a) The sensitive period for the acquisition of a non-native phonological system. In S. Krashen, R. Scarcella, & M. Long (eds.) *Child adult differences in second language acquisitions* (pp. 20-38). Rowley, MA: Newbury House.

Port, R., and Mitleb, F.(1983) Segmental features and implementation in acquisition of English by Arabic speakers. *Journal of Phonetics* 11, 219-229.

Schneiderman, E., Bourdages, J., & Champagne, C. (1988) Second language accent: The relationship between discrimination and perception in acquisition. *Language Learning* 38, 1-19.

Suter, R.(1976) Predictors of pronunciation accuracy in second language learning. *Language Learning* 26, 233-253.

Williams, L. (1980) Phonetic Variation as a function of second-language learning. In G. Yeni-Komshian, J. Kavanagh, and C. Ferguson (eds.) *Child Psychology*, Vol. 2: Perception (pp.185-215) New York: Academic Press.