Chapter I

INTRODUCTION

1.1 INTRODUCTION

Many speech sounds occur in human languages but only a small number of them are distinctive in a particular language. Sounds which are distinctive in a language are phonemes in that language. It is common knowledge that sounds that are in contrast in one language may not be so in another language. For example, in Thai [p] and [p^h] are two phonemes; /p/ and /p^h/ and written as <ป> and <พ> such as <ปา> [pa:] and <พา> [p^ha:]. In English, however these sounds occur as two allophones [p] and [p^h] of the phoneme /p/ such as <pie> [p^haj] and <spy> [spaj]. Research on speech perception has shown that perception ability develops from birth to adulthood, and that human's perception ability may be modified by linguistic experience. As it was stated above that each language has a different phonemic structure, it is apparentthat there is an experiential effect of the listener's native language on speech perception. This thesis will explore three topics:

1. The effect of the first language experience between 6 and 8 years on children's discrimination ability.

2. The effect of second language learning between 6 and 8 years on children's non-phonemic discrimination ability.

3. Phonetic and phonemic factors in speech perception.

1.2 AIMS OF THE STUDY

This experiment examines the perception of English speech sounds by Thai children of 6 and 8 years old. It aims to investigate:

1. the effect of the first language experience on speech perception;

2. the effect of the second language learning on speech perception;

3. the effect of phonetic and phonemic factors in speech perception by variation of Interstimulus Interval (ISI).

1.3 SCOPE OF THE STUDY

The subjects are Thai children of 2 different age groups, 6 and 8 years old. Half of the children have both English and Thai language experience (children age 6 years old are exposed to English language for 1 year and children age 8 years old are exposed to English language for 3 years), and half of them have only Thai language experience. The English voiceless fricative sounds i.e. [f], [s], [θ], and [ʃ] will be used as the stimuli. The first two sounds are also phonemic or are significant sounds in Thai while the other two are non-phonemic sounds.

1.4 HYPOTHESES OF THE STUDY

According to the speech perception studies (William, 1974 quoted in Strange & Jenkins, 1978; Werker & Tees, 1984) which examined subjects who have second language experience showed improved discrimination ability as a function of age and amount of the exposure to a second language. It is expected that the phonemic sound pair should be discriminated by both subject groups. However, there may be difficulty in discriminating non-phonemic sound pair. Specifically, it is hypothesized here that discrimination ability should be best for the Thai children who are exposed to both English and Thai followed by the Thai children who are exposed to only Thai and within each age group the children age 8 years old should be better than children age 6 years old.

The two Interstimulus Intervals (ISI): 500 ms and 1500 ms are used to examine the effect of ISI on the phonemic and phonetic processing of speech perception. It is expected that the discrimination ability at the ISI 500 ms level will be better than the ISI 1500 ms level.

1.5 LITERATURE REVIEW

1.5.1 Perception of Speech Sounds

In the early studies of speech perception that used speech stimuli such as consonant vowel series which vary in equal steps along an acoustic continuum, results showed that humans tend to perceive speech sounds categorically (Liberman, Harris, Hoffman & Griffith, 1957; Mattingly, Liberman, Syrdal & Halwes, 1971; Pisoni, 1973, quoted in Strange & Jenkins, 1978). Liberman et al. (1957) used thirteen two-formant synthetic consonant-vowel (CV) syllables. The syllables constructed differed only in the direction and extent of the initial second formant (F2) transition into the vowel (F2 is a cue for place of articulation). The starting frequency of F2 differed in equal frequency intervals for the perception of stop consonants: /ba/, /da/, /ga/. It appeared that, across this continuously varying physical series, the listeners discriminated the synthetic syllables into three categories by changing from 'ba' to 'da' and to 'ga' abrubtly. The categorical perception phenomenon was also confirmed by the inability of subjects to discriminate one stimulus from another when they were both drawn from the same phoneme category, even though they were able to discriminate sound pairs which drawn from separate phoneme categories. The categorical perception phenomenon has been observed in many studies. In the subsequent studies which used the F2 and F3 transitions varied in appropriate way to cue place of articulation (Mattingly, Liberman, Syrdal & Halwes, 1971; Pisoni, 1973.) and the study which used voicing in stop consonants as stimuli i.e. /b/-/p/, /d/-/t/ (Liberman, Harris, Kinney & Lane, 1961a; Liberman, Harris, Eimas, Lisker & Bastian, 1961b; Lisker and Abramson, 1970) have also shown the categorical perception in human being.

The study with infants showed that the categorical perception develop at a very early age (Eimas, Siqueland, Jusczyk & Vigorito, 1971). The investigators found that two- and three-month-old infants can discriminate synthetic speech sounds varying in VOT* categorically in a manner comparable to that of English speaking adults, thus it has been assumed that infants are responding linguistically relevant to speech sounds. However, this assumption was argued by the findings of other cross-language perception studies with infants (Lasky, Syrdal & Klein, 1975; Streeter, 1976) and the perception study with non-human (Kuhl and Miller, 1975 a,b).

---

*VOT or voice onset time which has been termed by Lisker and Abramson refers to the relative timing of glottal and supraglottal events or the relative timing of stop release and onset of voicing.

Lasky, Syrdal & Klein (1975) investigated four- to six-month-old Spanish infants on discriminating VOT* and found three categories which those do not coincide with phoneme boundary found in their language. As similar to this study, in the study with Kikuyu infants also showed three categories of voicing for labial stops (Streeter, 1976) but there is no voicing contrasts for labial stops in Kikuyu language, thus it is possible that infants may not be responding linguistically to speech sounds. This was confirmed by the investigation with non-human. (See discussion in Kuhl and Miller below.)

---

* Generally, there are three conditions of VOT,

1. voicing begins before the release of the stop (voice unaspirated),

2. voicing after the release of the stop (voiceless unaspirated),

3. voice onset lags behind the release (voiceless aspirated).

VOT was assigned zero-time to the point of release thus measurements of voice onset time are negative numbers which referred to the voicing before release called <u>voicing lead</u> and positive numbers referred to the voicing after the release called <u>voicing lag</u>.

Kuhl and Miller (1975 a,b) tested chinchillas' ability to differentiate stimuli along a VOT dimension that distinguished /ba/ from /pa/, /da/ from /ta/ and /ga/ from /ka/ for humans. It is apparent that chinchillas can differentiate these speech sounds and showed boundaries in the same location as humans. The other evidence came from the study with monkeys by Kuhl and Padden (1982). The trained monkeys also distinguished voiced and voiceless differences in VOT on each of the three continua tested (/ba/-/pa/, /da/-/ta/ and /ga/-/ka/) on a same-different task. There were best performances for between-category pairs of stimuli and poor performances for within-category pairs of stimuli. Both studies with infants and animals suggested that categorical perception ability is an innate auditory mechanism of both human and non-human which such results are not relevant with a particular language. The studies reviewed above tested only speech sounds stimuli especially VOT, consequently there was an early conclusion that categorical perception is unigue to speech. However, this was modified by the finding of categorical perception in non-speech sounds.

Pisoni (1977) demonstrated the categorical perception in non-speech by using stimuli varied over a range from -50 ms where the lower tone leads the higher tone, through simultaneity, to +50 ms where the lower tone lags behind the higher tone. The results showed an abrubt boundary between categories in the identification task, and the discrimination task also showed categorical peaks of discrimination for pairs drawn from different identification categories and poor discrimination for pairs within the same identification categories. The investigating with infants also showed the categorical perception of non-speech sounds (Jusczyk, Rosner, Cutting, Foard & Smith, 1977).

From all of the studies reviewed above, it can be concluded here that the categorical perception phenomenon is a process of the auditory system that can also be used to differentiate non-speech sounds.

1.5.2 The Modification of the Speech Perception Ability in the Perception Development: Experiential Effects

In the study of speech stimuli such as consonant vowel series which vary in equal steps along an acoustic continuum, results showed that humans and non-humans perceive speech categorically (Eimas, 1975; Kulh and Miller, 1975 a,b; Kulh and Padden, 1982; Liberman, Cooper, Shankweiler & Studdert-Kennedy, 1967), and this implied that categorical perception was natural for both human and non-human, and it was not relevant to a particular language. However, it is apparent that those categorical boundaries can be modified by an experience of a specific language.

1.5.2.1 Adults

In the perception of adult subjects, those categories tend to correspond to the phonemic structure of the listener's language. In the investigation of voicing distinction in stop consonants (VOT) showed that adults are most sensitive to differences in VOT at just those places along the dimension where their language places a phonemic boundary. Lisker and Abramson (1970) investigated the effect of linguistic experience on the VOT dimension. They compared performances on identification and discrimination tasks with speakers from different languages, Thai and American English. Thai listeners separated the labial VOT series into three categories as /b/ /p/ /p^h/ while American English listeners divided this continuum into two categories, /b/ and /p^h/. With an apical VOT series similar results were found: Thai listeners divided the continuum into three phonemic categories /d/ /t/ /t^h/, whereas American English listeners divided it into phonemic categories, /d/ /t^h/. In a subsequent study, it was very interesting that subjects who were familiar with a particular contrast in one context could not perceive the contrast in an unfamiliar context. In Thai, three categories of stop consonants which contrast both voicing and aspiration are made only in labial and apical articulation. In velars, there is no voiced stop thus the investigators (Lisker & Abramson, 1970) tested Thai subjects on velar VOT series. The subjects produced accurate discrimination only for the voiceless unaspirated and aspirated boundary. That is the perception is specific to the phonemic structure of their language. A study of Spanish and English speakers also showed a similar effect of linguistic experience (William, quoted in Strange & Jenkins, 1978). Both subject groups were most accurate around their native language's phonemic boundaries. It is not only the first language learning modifies speech perception but the second language learning and extra training also modify speech perception.

It is apparent that the amount of exposure to a particular language is an important factor which has an effect on discrimination ability. The more the language experience the subjects have, the more the accuracy they have on discriminating native speech contrasts. On the other hand, the more the subjects are exposed with their native language experience, the more difficulty they have in discriminating non-native contrasts. Studies of speech perception which examined subjects of different age groups i.e. infants, children, and adults showed not only that infants are able to perceive non-native speech sounds, but they also showed the variation of perception ability in children of different age groups and in adults.

1.5.2.2 <u>Developing Children and Adults</u>

It has been found that there is a modification of the development of discrimination ability of humans especially in childhood. Some non-native contrasts which infants can discriminate, older children cannot discriminate. However, the ability can be resurgent sometimes. Burnham (1986) proposed that there are two periods of loss of perception ability. Firstly, loss occurs in early life around 6-12 months and secondly loss occurs around 4-6 years. The early loss of speech perception has been found especially with the sounds which are rare in human languages and phonologically irrelevant in the listeners native language while the late loss is found with more common sounds and which are phonologically irrelevant in native language. However, some studies attempted to change this decline in ability by using training method. Especially, the studies of voice onset time (VOT) showed that after training, the discrimination ability of adult English speakers could be improved (Pisoni et al., quoted in Werker & Tees, 1984; Strange, quoted in Strange & Jenkins, 1978).

In addition to training methods, it was apparent that second language learning can also improve discrimination ability. William (William, quoted in Strange & Jenkins, 1978) examined Spanish-speaking children who learn English language as a second language to investigate their perception of labial sound stimuli varying in VOT. The subjects differed in age (8-10 years, 12-14 years) and amount of exposure to English (0-6 months, 1.5-2 years, 3-3.5 years). On the identification task, all groups showed crossover boundaries between Spanish (-4 VOT) and English (+25 VOT) and the longer the exposure to the English language the more there was a shift in boundary towards the

English postvoiced location.

Streeter & Landauer (1976) also investigated an improvement of perception ability as a function of second language learning with Kikuyu children (7.5, 10, 13, 15 years) on their perception of labial stops with different VOT. The results showed that the children could discriminate the contrasts which are non-native and the ability improved with their age and the exposure time with English language. There are some recent studies which investigated both the effect of training and exposure to the non-native language. Werker & Tees and their colleagues (1981, 1983, 1984a, 1984) investigated the ability to perceive sounds in Hindi [ṭ]-[t̪], [tʰ]-[dʰ] and sounds in Thomson [k]-[q] across age groups and across languages. It is apparent that young English infants (6-8 months) could discriminate these contrasts as well as native speaking adults but that the adults have little or no discrimination ability, even the adults who had extra training. With the 4-, 8-, 12-year-old, it appeared that half of them could discriminate the voicing contrast [tʰ]-[dʰ] but that most of them could not discriminate the place of articulation contrast [ṭ]-[t̪].

Although the training method was not successful in improving adult's discrimination ability, it is remarkable from those studies that subjects who had more experience with Hindi language (5 years) and early exposure to that language had more discrimination ability than the other subjects.

According to all of the studies reviewed above, it is apparent that speech perception development can be modified by linguistic experience with both native and non-native languages.

1.5.3 Age and Speech Perception

As revealed in 1.5.2.2 that there is a loss of perception ability in the perception of non-phonemic sounds. The two periods of loss are: loss in the first year of life and the loss at 6 years (Burnham, 1986). Burnham (1986) claimed that loss in the infancy occurs as a result of lack of exposure to specific sounds. And the late loss is due to the children's experience with particular sounds and lack of experience with others. The perception studies which investigated the perception ability of English subjects age 4, 6 and 8 years on the perception of bilabial stops (Burnham, O'Connor, Clark, & Earnshaw, 1985; Burnham, O'Connor, & Earnshaw, 1986), were found that children age 6 years paid relatively less attention to the non-phonemic sounds than the children age 4 and 8 years. They referred this phenomenon as phonological bias and also proposed that when children begin to learn language formally at around age 6 years, they will learn to segment speech sounds into phonemes and associate those phonemes with graphemes. The competence in language and segmentation skills will enable, and possibly forces, children around this age to use a phonemic processing strategy when listening to speech and have little capacity for phonetic processing, thus the perception ability of non-phonemic sounds were reduced. However, as children become more experienced at phonemic processing, it becomes more autometic and requires less attentional capacity. Thus, it was found an increase perception ability in the children age 8 years.

1.5.4 Phonemic and Phonetic Factors in Speech Perception

Several studies examining the perception of non-native (non-phonemic) sounds found that the interstimulus interval (ISI) (the time between the first and the second sounds presented in a discrimination task given in duration terms i.e. milisecond) can influence the subject's perception of speech sounds (Werker & Tees, 1984; Werker & Logan, 1985). It was evident that in the shorter ISIs the subject could discriminate more contrasts than in the longer ISIs. Werker and Tees (1984) investigated the perception of Hindi contrast with different ISIs, 500 ms and 1500 ms. The results showed better discrimination ability with 500 ms ISI than with 1500 ms. The investigators claimed that the 500 ms leads to phonetic processing whereas the longer time leads to phonemic processing. The shorter ISIs allows the subjects to relinguish the phonemic processing stategy which they always use to perceive speech sounds in their language, and use a phonetic mode to perceive speech sounds thus allowing irrelevant sounds to be perceived. Werker and Tees stated that

> ... a memory trace is available following the 500 ms which has decayed after 1500 msec. This memory trace may enable the subjects to relinguish an exclusively phonemic processing strategy, and detect differences within phonemic categories... At a phonemic level, stimuli are treated as meaningful events that can be efficiently encoded and represented in memory for a long time. At a phonetic level, the sound may be perceived as synthesized (nonmeaningful) percepts, and thus may be retained in memory for a short duration, but have a more rapid decay period than phonemic stimuli...
>
> (Werker & Tees, 1984, pp. 1875 and 1876).

Thus, it is possible to say that allowing the subject to relinguish a phonemic processing strategy is a way to improve perception of non-phonemic speech sounds.

1.5.5 Perception of Fricative Sounds

1.5.5.1 Fricative Sounds

Fricatives are characterised by a turbulent airstream which occurs when air is channelled through a narrow constriction, possibly striking an object at high velocity (the upper teeth in the case of [s]). The acoustic consequence of turbulence is a noise source; the spectra of fricatives can be considered as the product of the noise source located supraglottally (except for [h]), which is modified by a resonator. In the case of voiced fricatives, a second source due to the vibrating vocal folds may also be present in the spectrum (Harrington, 1988). The place at which the constriction is formed, the shape and size of the front cavity between the constriction, and the air outside the lips have a frequency-filtering effect on the source sound in much the same way that vowel shape acts as a filter to form the vowel spectra from the source sound produced by the glottis. The back cavity, behind the constriction, does not strongly affect the fricative spectrum (Pickett, 1980). Fricatives are used in languages and can be made at every place from bilabial to glottal (O'Connor, 1973). Those sounds were presented in table 1.1 page 17.

Table 1.1 Fricative Sounds

Voiceless	Voiced	Place of Articulation
ɸ	β	Bilabial fricative
f	v	Labio-dental fricative
θ	ð	Dental fricative
s	z	Alveolar fricative
ʂ	ʐ	Retroflex fricative
ʃ	ʒ	Palato-alveolar fricative
ç	ʝ	Palatal fricative
x	ɣ	Velar fricative
ʍ	w	Labial-velar fricative
χ	ʁ	Uvular fricative
ħ	ʕ	Pharyngeal fricative
h	ɦ	Glottal fricative

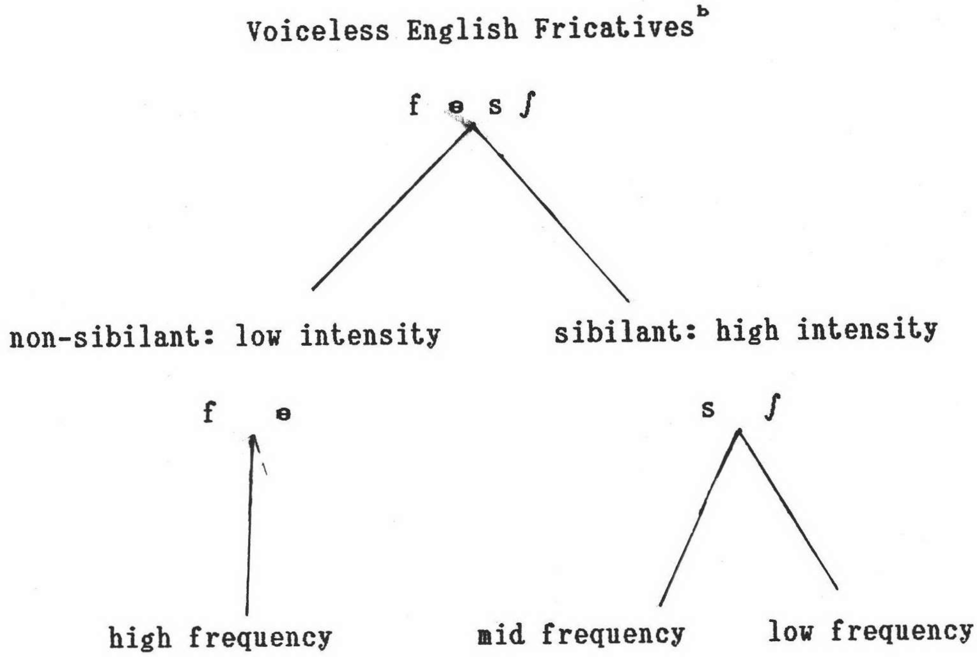
The present study used only voiceless fricative sounds which are phonemic sounds in English language namely [f], [θ], [s] and [ʃ]. These four sounds can be separated into 2 groups by their intensity friction, called sibilant and non-sibilant, [s] and [ʃ] have greater intensity friction are sibilants contrast with non-sibilant [f] and [θ] (Ladefoged, 1975; Strevens, 1960). Another acoustic characteristic of sibilants is a concentration of energy in a 3- 4 kHz band in the 1.5-7 kHz region. Non-sibilants exhibit an approximately equal concentration of energy in the 1-8 kHz range and have intensification in the 8-16 kHz range due to the very small front cavity resonator which is considered to be characteristic of [θ] and [f] (Harrington, 1988).

1.5.5.2 Cues to Fricative Place of Articulation

Heinz and Stevens (1961) suggested that frequency of fricative sounds are related to the size of the front cavity. As the front cavity becomes smaller, the position of the resonance moves upward in frequency. Within the 4 English fricative sounds ([f], [θ], [s] and [ʃ]) used in the present study, the sound [ʃ] has the largest front cavity due to its sound source which is at the palatal and thus has the lowest frequency of the four fricatives, followed by [s], the next is [θ] and [f] which have highest frequency in the set (Pickette, 1980). And these reflected in the perception of fricatives. The [s] and [ʃ] which are in the mid and low frequency range between 2.5 to 6 kHz and high intensity friction are well differentiate, whereas [f] and [θ] which have higher frequency but low intensity friction are more difficult to differentiate. The study by Harris (Harris, 1958) indicated that [s] and [ʃ] were perceived correctedly independent of information from the transition to the adjacent vowel, but [f] and [θ] were confused, the listeners need both the friction cues and the transitions into neighboring vowels to determine the place of articulation of [f] and [θ]. The low intensity friction of [f] and [θ] accounts for the difficulties that listeners have in identifying them out of the context. In the perception experiment of Heinz and Stevens (1961) which varied the resonant frequency of the single-pole circuit through several values from 2-8 kHz. Identification of [f] and [θ] usually occured at 6.5 kHz or 8 kHz ; most [ʃ] responses were obtained when the resonance was around 2.5 kHz while most [s] responses were elicited for resonances around 5 kHz.

A summary of the cues which could be used for the discrimination within the fricative sounds used in this experiment is shown in the form of a discrimination tree in figure 1.1

Figure 1.1 Acoustic Features of Fricative Sounds [a]



[a] Adapted from Harrington, 1988.

[b] voiceless English fricatives used in this study except for [h]

The relationship between the sounds and the acoustic features in the Figure is given as follow:

1. non sibilant: low intensity friction and low energy for (f-θ) and sibilant: high intensity friction and high energy for (s,ʃ).

2. high frequency for (f-θ), mid frequency for (s) and low frequency for (ʃ).

1.5.5.3 The Studies of Fricative Perception

There have been a number of studies on the perception of fricative contrasts. Some studies which used discrimination task have shown that although fricative contrasts are difficult for young infants to discriminate (Eilers, Wilson, & Moore, 1977; Morgan & Kuhl, 1977), English speaking infants can perceive these fricative contrasts when maximal acoustic cues are provided (Eilers, 1977). In a cross-language study, Spanish infants were found to perceive these contrasts as well as English infants (Eilers, Gavin, & Oller, 1982). This is consistent with the universal perception theory (Ingram, 1989) which proposes that infants can perceive a variety of speech contrasts, even those contrasts that are not relevant in their language. In the present study, the perception of English voiceless fricatives by Thai children will be investigated. It is expected that Thai children are able to discriminate these sound contrasts.

1.6 DEFINITION OF TERMS AND ABBREVIATIONS USED IN THE STUDY

1. Subject Group: 48 Thai children age 6 and 8 years old half of whom have English language experience, and the rest have only Thai language experience.

a. T- 6 group: the subjects age 6 years old which are exposed to only Thai language.

b. T- 8 group: the subjects age 8 years old which are exposed to only Thai language.

c. TE- 6 group: the subjects age 6 years old which are exposed to both Thai and English language.

d. TE- 8 group: the subjects age 8 years old which are exposed to both Thai and English language.

2. Sound pairs: the speech stimuli consist of six pairs of English voiceless fricative contrasts.

a. P- sound pair or Phonemic sound pair: the sound pair consists of 2 English fricative sounds (f-s) which are also phonemically significant in Thai language i.e. both sounds are phonemic sounds in Thai language.

b. NP-sound pair or Non-phonemic sound pair: the sound pair consists of 2 English fricative sounds (θ-ʃ) which do not occur as phonemic sounds in Thai language i.e. both sounds are non-phonemic sounds in Thai language.

c. PNP-sound pair or Phonemic and Non-phonemic sound pair: there are 4 sound pairs which consist of both phonemic and non-phonemic sounds. They are (f-θ), (f-ʃ), (s-θ), and (s-ʃ).

d. Sound pair : sp1 = (f-s), sp2 = (f-θ), sp3 = (f-ʃ), sp4 = (s-θ), sp5 = (s-ʃ), sp6 = (θ-ʃ)

3. Interstimulus Interval (ISI): the time between the first and the second sounds presented in a discrimination task given in duration terms i.e. milisecond.

4. <u>Discrimination Index (DI)</u>: the measurement which is used to measure the subjects' ability to discriminate whether two sounds are different. It is given by the number of correct responses on different trials minus the number of incorrect responses on same trials, divided by the number of different trials.

5. <u>F Value</u>: the obtained F value is the result of ANOVA calculation. The obtained F value will be compare with the F value in the F table. (The F table is a two-dimentional table and we look up the F value in the table in terms of its degrees of freedom (df) for numerator (which is the df for the mean square between the groups) and its degrees of freedom for demoninator (the df for the meansquare within the groups.) There are the 0.5 level of the significant and the 0.1 level of the significant.) The obtained F which exceeds the F value in the F table is a significant F value.