PERCEPTION OF VOICING CONTRAST AND ASPIRATION IN WORD-INITIAL STOPS : A CROSS-LANGUAGE STUDY.

Registration No. 05SLP002
ANJALI .G

A Dissertation Submitted in part fulfillment of
Master's Degree (Speech Language Pathology),
University of Mysore,
Mysore.

ALL INDIA INSTITUTE OF SPEECH AND HEARING
MANASAGANGOTHRI
MYSORE - 570 006

APRIL - 2007

Dedicated to Achan, Amma,
Kichu n Manu….

Certificate

This is to certify that this Dissertation entitled **"Perception of voicing contrast and aspiration in word-initial stops : A cross-language study"** is a bonafide work in part fulfillment for the degree of master of (Speech Language Pathology) of the student with Registration No.05SLP002. This has been carried out under the guidance of a faculty of this institute and has not been submitted earlier to any other University for the award of any other Diploma or Degree.

Mysore

April, 2007

Dr. Vijayalakshmi Basavaraj
Director
All India Institute of Speech and Hearing
Naimisham Campus
Manasagangothri
Mysore-570 006.

Certificate

This is to certify that this Dissertation entitled **"Perception of voicing contrast and aspiration in word-initial stops : A cross-language study"** is a bonafide work in part fulfillment for the degree of master of (Speech Language Pathology) of the student with Registration No.05SLP002. This has been carried out under the guidance of a faculty of this institute and has not been submitted earlier to any other University for the award of any other Diploma or Degree.

Savithri. SR. 30.4.07.

Guide

Mysore

April, 2007

Prof. S. R. Savithri
Professor and Head,
Department of Speech-Language Sciences
All India Institute of Speech and Hearing
Naimisham Campus
Manasagangothri
Mysore-570 006.

Declaration

This Dissertation entitled **"Perception of voicing contrast and aspiration in word-initial stops : A cross-language study"** is the result of my own study under the guidance of Prof. S. R. Savithri, Professor and Head, Department of Speech-Language Sciences, All India Institute of Speech and Hearing, Mysore, and has not been submitted earlier in any other University for the award of any Diploma or Degree.

Mysore

April, 2007 Register No. 05SLP002

Acknowledgements

"Good teaching is more a giving of right questions than a giving of right answers".

My first thoughts of appreciation go to my guide, ***Prof. S. R. Savithri.*** *Dear Ma'm, a million thanx will not suffice for your contributions by way of reviews, comments, support and encouragement while planning and executing the study, which gave me the confidence to do what I did.*

I would like to thank Dr. Vijayalakshmi Basavaraj, Director, All India Institute of Speech and Hearing, Mysore, for allowing me to carry out this study.

ToAchan and Amma . . . . you have raised me to feel that I can accomplish any goal or task I decide upon and in that you have showered upon me the greatest of all blessings. I owe you everything.

Dearest Kichu, my brother... We know each other as we always were. We know each other's hearts. You are the keeper of my identity, the only person with the keys to my unfettered, more fundamental self. Just to let you know that it was nice growing up with someone like you - someone to lean on, someone to count on... Love you...

Manu *. . . I sought my soul, but my soul I could not see. I sought my God, but my God eluded me. I sought you and I found all three. You are more than special to me - you are a part of me. You touch my heart in a way I never knew before. Being deeply loved by someone gives you strength, while loving someone deeply gives you courage... you have given me both ... My life... Love u Kutta ...*

"When it hurts to look back, and you 're scared to look ahead, you can look beside you and your best friend will be there."

*Dearest **Ani**... you are like a hand that is always holding mine, no matter how close or far apart we may be. You are one of the best things that's ever happened to me. Thanx Erumsfor being therefor me, when I needed u the most.*

"A true friend is someone who thinks that you are a good egg even though she knows that you are slightly cracked".

*Dearest **Savi**... through thick and thin u have stood beside me and you bring out the best in me. Thanx Pans for all the love and support u had endowed upon me.*

*Nambi..u are a wonderful friend.. The word thanks is too less for all your help & support when I was in need of it. Sheetz... u r a very special friend to me, truly cherish the days we were together. **Radz and Suma** ...thanx a lot for the times we spent together... **Priya Rahan n Johns...** Nice having friends like you... **Suppi n neethu...** thanx for all the gud times we had... **Pooja n Shaloo...** thanx for being such great room mates.*

*To all my B.Sc & M.Sc **classmates** for the fun and laughter we all have shared and the sweet memories we've made. A big YO to AIISH PEARLS!!!*

*My sincere thanks to **Vasanthalakshmi ma'm** for her constant guidance with statistics. My special thanks to all the **Library staffs** and **Xerox***

*I thank **Almighty** for all the wonderful opportunities He has given me in life and His support through the ordeals in my life.*

Mistakes are one of the few things we can truly call our own and if we cannot learn from them, who can...?

Table of contents

List of tables

List of figures

Chapter I

Introduction

Speech may be defined as a form of oral communication in which transformation of information takes place by means of speech waves which are in the form of acoustic energy (Fant, 1960). Speech is a fascinating human attribute that can be analyzed, synthesized and recognized. Communication by speech is the transmission of thoughts or feelings from the mind of a speaker to the mind of a listener. The concepts and attitudes that the speaker intends to express are embodied in a linguistic frame and rendered audible. The listener hears the speech signal and interprets its meaning. Audition is the process of registering the sounds in the brain of the hearer. Speech perception is the process of decoding a message from the stream of sounds coming from the speaker.

The human auditory mechanism analyzes sound according to changes in frequency and intensity across time. The sounds of speech change not only in amplitude but in their mode of transmission as they travel through the outer ear, middle ear, cochlea, and auditory nerve to the brain. The pressure waves of speech are usually disturbances in air and thus they continue in the outer ear. In the middle ear, however, they are converted from pressure waves to mechanical vibrations by a series of small bones leading to the cochlea of the inner ear. In the cochlea, a snail-shaped cavity within the temporal bone of the skull, these vibrations are again transformed to vibrations in fluid, since the cochlea is filled with fluid. Finally, the nerve endings in the cochlea act to transform the hydraulic vibrations into electrochemical changes that are sent to the brain in the form of nerve impulses.

The human auditory system is especially tuned for speech, or, our speaking mechanisms and auditory mechanisms have developed together, so that we are best at hearing speech sounds. From the perspective of historical linguistics, it can be considered that the languages of the Earth may have developed as they have by taking advantage of, and at the same time being constrained by, the speech mechanisms and auditory mechanisms of the human being.

The processing of speech signals within the existing constraints of natural language has interested and excited scientists for many years. Speech sounds are varied and have numerous acoustic cues like the formants, their bandwidths and levels, fundamental frequency, energy, duration of closure, preceding vowel duration, burst energy, voice onset time etc. It appears that the auditory system depends on some of the acoustic cues of the speech sounds to identify and thus to perceive it.

There are many acoustic dimensions distinctive in speech that can be systematically varied by speech synthesis in the construction of tests for the identification and discrimination of speech sounds. Acoustic cues important to manner of articulation, such as duration of F2 transition or intensity rise time of noise have been varied along a continuum. When transition durations are increased in equal steps, listeners categorically report /ba/, then /wa/, and finally /ua/, and the discrimination peaks correspond to the boundaries between the different manners of speech sounds. Noise patches representing friction can be arranged to continuously vary from sudden rise time to gradual rise time, with listeners reporting a sudden perceptual shift from /tʃ/ as in 'chop' to /ʃ/ as in 'shop'. Acoustic cues important to place perception have also been systematically varied. E.g., changes in the direction of F2 transition reveal the categorical perception of /b/ */d/* and /g/. Like categorical perception of changes in manner and place, responses to continua of smooth variations in voice onset time (VOT) demonstrate categorical perception of voicing. By increasing VOT in equal steps, continua can be synthesized which are perceived as going from /ba/ to /pa/, from /da/ to /ta/, or from /ga/ to /ka/. Again, listeners group the stimuli into voiced and voiceless categories and perceive and are sensitive to differences between stimuli at the voiced-voiceless boundary, while they are relatively insensitive to equal VOT differences within categories.

The phenomenon of categorical perception for consonants has been demonstrated for manner, place and voicing contrasts. For vowels, results are a bit different. Fry, Abramson, Eimas and Liberman (1962) showed that changing continua from /l/ to /£/ to */x/* do not show the same close relationship between identification and discrimination functions. When the vowels are shorter and embedded in CVC contexts, however, as reported by Stevens et al. (1969), the discrimination-identification relationship is more like that for consonants.

Cross language speech perception represents a dynamic research area in which the ideas have grown and changed over years. Cross language perception in simpler terms refers to the perception of non-native contrasts by native speakers. Cross linguistic research is one way of investigating the effects of exposure to a given phonological system on the perception and production of phones from another phonological system.

Languages across the world vary in their phonological and phonetic inventories. In a particular language (L1), 2 phones may occur, while in another language (L2), the phones may not occur at all, or L1 and L2 may share 2 phones, but in LI the phones may be phonologically contrastive, while in L2, they may occur in contextual or free variation rather than being used to distinguish meaning. Because of this variation across languages, several questions have been asked about the potential role of linguistic experience in the perception of phonological categories. Investigators have tried answering questions like do individuals exhibit universality in perception of speech sounds or does the native language restrict the range of speech sounds which the native language users can perceive. If so, does adequate exposure or training help to perceive the non-native contrasts? What are the developmental changes occurring in cross language perception? And so on.

The pattern of results in cross-language research has led investigators to propose that exposure to specific phonetic contrasts during an early critical period is needed to maintain the neural elements that are innately tuned to the phonetic features involved and, conversely, that lack of exposure to particular contrasts results in attrition of the associated neural elements (Eimas, 1975; Aslin & Pisoni, 1980).

Existing empirical research has indicated that young infants could discriminate between native and non-native phonetic contrast (Lasky, Syrdal, Lasky & Klein, 1975; Streeter & Landauer, 1976; Trehub, 1976; Aslin, Pisoni, Hennessy & Perey, 1981) but that adults and children often have difficulty discriminating non-native contrast (Singh & Black, 1966; Lisker & Abramson, 1970; Goto, 1971; Miyawaki, Strange, Verbrugge, Liberman, Jenkins & Fujimura, 1975; Snow & Hoejnagel-Hohle, 1978, Trehub, 1976; Sheldon & Strange, 1982). Several studies have also reported that infants and adults are able to distinguish some non-native contrast (Best, 1995). A few studies in Indian languages also support the above findings (Vinay, 1990; Savithri & Sridevi, 1991; Jessy, 2003). A great deal of research has focused on perception of non-native contrasts by individuals, but most of these studies have used meaningful words as stimuli which could have affected the perception.

Initial research (Caramazza, Yeni-Komshian, Zurif and Carbone, 1973; Lisker & Abramson, 1970) on cross- language perception was with respect to the various contrasts which could be discriminated by non-native listeners. The focus gradually shifted over the efficacy of using various training strategies to sensitize these individuals to non-native contrasts. These studies (Strange & Jenkins, 1978; Werker & Tees, 1984; Pisoni, Aslin, Perey & Hennessy, 1982; Flege & Wang, 1989; Polka, 1991; Pruitt, 1993) led to the conclusion that inability to discriminate the non-native contrasts does not imply a sensory loss, rather a perceptual reorganization resulting from linguistic exposure (Werker, Gilbert, Humphrey & Tees, 1981; Best, 1990;).

India, a country with diverse linguistic groups, offers greater scope for research in cross language speech perception. This would help in understanding the perceptual skills of bilinguals and multilinguals, the extent to which the native language and subsequent exposure to other languages would re-organize their perceptual skills. This has implications in teaching second language to adults and children.

Some studies have been conducted on Indian languages too. These include perception of vowels by Bengali and Hindi speakers (Himanshu Khanna, 1996), discrimination of stop consonants by Malayalam and Tamil speakers (Sreedivya, 1997), consonant perception by Malayalam and Hindi speakers (Jessy George, 2003) and discrimination of Malayalam laterals by Malayalam and Hindi speakers (Agarwal & Savithri, 2005). All these studies are on perception of speech sounds by speakers of two languages. Most of these studies use an AX design, where A is the first target phoneme and X is the second target phoneme. A "same', 'different' response or a multidimensional status is used in these studies. In the study by Agarwal & Savithri, the stimuli were presented only once. Thus, bias may be involved in the response. In order to remove the response bias multiple presentations of the stimuli is required. Also, most of these studies have not taken care of the context effects as they have used words as stimuli. In this context, the present study was planned. The objective of the present study was to investigate the native and non-native speakers' ability to discriminate voiced - unvoiced, aspirated — unaspirated, and voiced - voiced murmur stops.

The parameters cueing voicing differ depending on the position of plosives in a language. In the word-initial position, voice onset time and Fl transition cue voicing. Voice onset time (VOT) is the time difference between the onset of articulatory release and the onset of voicing and is considered a major cue for differentiation of prevocalic stops along the voicing dimension (Lisker & Abramson, 1964; Abramson & Lisker, 1965).The differences in VOT have been termed lead vs. short lag for voiced and unvoiced, respectively (Lisker & Abramson, 1964; Keating, Mikos & Ganong, 1981). Across languages, Lisker & Abramson (1964, 1967), indicated a fairly consistent 60 ms minimum difference in VOT between voiced and unvoiced stops. Data from studies with synthetic speech have suggested that the acoustic characteristics providing the simplest and the most direct indication of whether a stop consonant is voiced or unvoiced is VOT. Voiced stops have a well defined transition in the first formant of the following vowel, while Fl transition of unvoiced plosives is essentially non-existent after the onset of voicing (Sussman & Carney, 1989, among others). This lack of formant transitions after voicing onset for the stops indicates that the rapid movements of the supraglottal articulators are essentially complete before the vocal cords are in configuration for the onset of voicing. Based on synthesis experiments, it is known that the duration of these transitions is in the order of 40 ms or less.

Plosives can also be classified based on aspiration as unvoiced unaspirated, unvoiced aspirated, voiced unaspirated and voiced murmur. Aspiration is cued by aspiration duration, but not much of research has been focused on this aspect.

Malayalam has 21 stop consonants, of which 10 are grouped into five pairs each with voiced and voiceless counterparts (Table 1). The alveolar stop has no voiced counterpart. It has only a voiced allophone which occurs after homorganic nasal. There are 3 lax plosives. Regarding aspiration, in Malayalam, there is the distinction between aspirated and unaspirated sounds. The aspirated sounds occur in Sanskrit loans. But these are not given independent phonemic status considering the principle of economy. They can be considered as sequences of a stop +h. This is possible because there is no such cluster of a stop +h that will be in contrast with this aspirated stop in Malayalam. Tamil has 8 stop consonants (Table 1). The voiceless stop phonemes /p, t, t, and k/ have allophones /b, d, d and g/ which occur initially in the

Chapter II

Review of Literature

Research in speech perception is concerned with several general issues surrounding perceptual learning. Cross-language perception throws light on the perception on non-native contrast as well as the role of linguistic environment in speech perception in infants, children and adults. Over the years cross-language differences in perception of non-native contrasts, consonants, vowels and suprasegmental aspects are studied and theoretical explanation are provided for the existing differences.

The review of literature will be elaborated under the following headings:

1. Cross-language perception of consonants in adults
2. Variables in cross-language studies on consonants
3. Theoretical explanation for cross-language differences in speech perception.

1. Cross-language perception of consonants in adults

a. Perception of voicing

Lisker and Abramson (1970) carried out perceptual studies using synthetic stimuli varying in VOT. The stimuli consisted of labial, apical and velar stop consonants. English, Thai, and Spanish natives showed that their identification and discrimination function and boundaries were determined by their native language. Goto (1971) studied adult learners and found that adult bilinguals are often quite insensitive to perceptual distinctions in their non-native language, even if they can produce them. Japanese speakers judged by Americans to be making the English /r/ /l/ distinction appropriately as they produced words such as 'lead', 'read', nevertheless found difficulty in perceiving the distinctions. In another study, Lisker and Abramson (1972) attempted to train native speakers of Russian to distinguish between the voiceless unaspirated and voiceless aspirated stop. This voicing distinction is in English but not in Russian. Although Russian subjects learned to identify the end point stimuli (i.e., +10ms and +60ms in VOT) slightly better than chance, their performance was not same for both the stimuli. Though they could use two discrete labeling responses their performance on the task was neither consistent nor reliable. Since no immediate feedback were provided after training trial, they probably had more difficulty in determining which specific acoustic attribute of the stimuli to attend selectively.

VOT as a linguistic cue to separate word-initial stop consonants in three groups of subjects viz., Unilingual Canadian, French and English and bilingual French-English speakers were studied by Caramazza, Yeni-Komshian, Zurif and Carbone (1973). Identification tasks were used. Unilingual English speakers had sharp monotic slopes in their perceptual functions, this being absent for unilingual French subjects. Bilinguals had steeper slopes but non-monotic, cross-over points in the intermediate position. This lack of monotonicity indicated that the first learned language interfered in the perception.

Williams (1979) established the monolingual identification boundaries for /b/ and /p/ and the discrimination peaks for both English and Spanish adult speakers. She found that their production of /b/ and /p/ in word-initial position corresponded to their perception. English speakers separated the phonemes at about +25 VOT, while Spanish speakers put the boundary at about -4 VOT. Spanish speakers learning English varied more than monolinguals in the crossover points of their identification functions, and the discrimination peaks spanned both the monolingual Spanish and English boundaries.

In a second study, Williams tracked changes in production and perception in young Puerto Rican Spanish-speaking children who were learning English. She found the labeling crossover point gradually shifting toward the English boundary as exposure to English increased. In production, the children were using VOT patterns closer to English, both in their English and Spanish words.

Miyawaki et al., 1975, assessed perception of synthetic /r/-/l/ continua by native Japanese, native bilingual Japanese speakers with English as their 2nd learned language in adulthood and native American English speakers. The continuum varied on F3 formant considered as primary cue for contrast in English. The results showed that whereas English listeners were successful in categorical perception, the Japanese listeners both monolinguals and bilinguals were unable to do so and discrimination was nearly random.

Elman, Diehl, and Buchwald (1977) found that the language set the listeners have when making decisions about speech sound identity is capable of changing the boundary between categories. Bilingual subjects divide stimuli according to the phonemic contrasts of the particular language they are using immediately before each stimulus.

Strange and Jenkins (1978) attempted to modify voicing perception in adults. A small number of college-age students were trained to identify and discriminate differences in the lead region of VOT continuum in which Thai voiced / voiceless unaspirate boundary occurs. Oddity discrimination task with immediate feedback, identification paradigm with delayed feedback and a scaling procedure was used. They found only a small change in the perceptual categories with performance improving in the target region of VOT continua where identification and scaling procedure was used. Subjects failed to generalize from one VOT series to another and the results were marked by high variability.

Pisoni, Aslin, Perey and Hennessy (1982) studied the perception of stops differing in VOT in two different conditions by two groups of native subjects. In the first condition, they used two response categories, corresponding to phonemes /b/ and /p/. In the second condition, they were provided with three response alternatives /b/, /p/, /p^h/. They showed reliable 2-category and 3-category identification function. Two additional groups of subjects were studied in another experiment using the same stimuli and in addition to identification function, discrimination (AXB) function was also carried out. They could discriminate stimuli in the voicing lead region despite being identified as belonging to same perceptual category. A discrimination training procedure with immediate feedback was carried out. Training was presented in a predictable order using only three stimuli, one from each of the three voicing types (-70ms, 0ms, +10ms). After training, those who met a predetermined criteria were tested for identification and discrimination function. The subjects were highly consistent in labeling the sounds in the voicing lead region of the continuum. Steeper slopes in identification function were obtained. Thus, within a short period of time, native English speaking adults can reacquire the non-native contrast in voicing using simple laboratory training.

Flege and Hillenbrand (1987) studied the differential effect of release burst on stop voicing judgement of native French and English listeners. When the release burst was removed from the word final English /g/ tokens, the voicing judgements of native French speakers were influenced but no effect was seen in native English subject. French natives gave more emphasis on release burst, as the stops are consistently released in French.

Flege and Wang (1989) studied Chinese subjects' perception of English word final *IXl* - /d/ from which word final closure voicing and release burst cues were removed. Performance of three Chinese groups - Cantonese, Mandarin and Shanghainese were compared before, during and after the feedback training Cantonese subjects were expected to perform best because their L1 permits unreleased /p, t, k/ in word-final positions. Mandarin subjects were expected to perform poorly because their L1 permits no word-final obstruents. Shanghainese subjects were expected to perform at intermediate level. Results showed an increased sensitivity to these contrasts as a result of training. Specifically, Cantonese subjects focused greater attention on end of CVC stimuli than Mandarin subjects which enabled them to use other acoustic cues, Fl offset frequency for *ltl - /d/* contrasts.

Flege, Bohn and Jang (1997) studied language set effects in Spanish and English monolinguals and bilinguals. The stimuli used were short-lag Spanish /t/, Spanish /d/ with lead VOT, short-lag English /d/ and long-lag English /t/. A small phonetic context effect was observed for Spanish and English monolinguals and bilinguals. That is, Spanish /t/ tokens were presented in Spanish and English perceptual sets. A small 'language set' effect was observed for the monolingual and bilingual listeners. Since both the groups showed the same effects, a post-perceptual language independent decision strategy based on their language-dependent perceptual processing could have been employed.

b. Perception of Place of Articulation

A number of cross-language studies have inferred greater perceptual difficulty for some nonphonemic contrasts than others. In particular the non-native contrasts involving phones that subjects experience as allophones of the native language phones are less difficult to differentiate than the ones involving place of articulation contrasts which are unlikely to have experienced as allophones.

Werker & Tees, 1981, compared English speaking adults and Hindi speaking adults on their ability to discriminate the English & Hindi voiced bilabial versus alveolar contrast /ba/-/da/ and 2 non English, Hindi contrasts. The Hindi place of articulation distinction between retroflex and dental voiceless stop consonants /ta/-/ta/ was selected as a potentially difficult non English contrast and the Hindi voicing distinction between breathy voiced & voiceless aspirated dental stops /dha/-/tha/. The retro flex/dental contrast was acoustically less salient than those in voicing contrast. Results of the study indicated that while both English & Hindi listeners could detect English-Hindi /ba/-/da/ contrasts, the Hindi listeners scored 100% on both native contrasts & the English listeners scored 40% on potentially easy contrast & 10% on potentially difficult contrasts. They concluded that there is an effect of experience on speech perception. Also, experiential effects are more pronounced on some contrasts than others.

Yoshida & Hirasaka (1983) investigated the identification of minimal-pair contrasts between English /b/ - /v/, *Ixl* - /I/, and /s/ - *lei.* Some pairs consisted of real words and some consisted of nonwords. The results obtained from 96 Japanese listeners indicated that, overall, the rate of errors decreased from /r/ - / l / (27%) to / b/ - /v/ (23%) to /s/ - *le/* (16%). An effect of lexical status was also observed in that more errors were found for consonants in word level than real words.

Werker & Tees (1984) studied English speaking adults' ability to discriminate the Hindi contrast voiceless aspirate /t^h/ vs breathy voiced /d^h/.

They could not discriminate this contrast, but limited training of 25 trials was sufficient to facilitate the discrimination of this voicing contrast. This VOT boundary cross is distinctive in English and hence it could be easily recovered.

Werker & Tees (1984) conducted four different experiments to determine whether the decline in non-native speech perception represents a sensori-neural loss, a shift in attentional focus, or the use of language specific processing strategies. In experiment 1 authors investigated whether the ontogenetic reorganization in cross-language speech perception resulted in a decline in sensorineural responsivity (i.e., neural atrophy) or result of a change in attentional mechanisms and or processing strategies? Also if a decline cannot be accounted for by a neural atrophy, will an attentional manipulation enable adult English subjects to discriminate the non-English phonetic contrasts? Eight adult English-speaking subjects in the age range of 18 to 40 years were tested. The material consisted of two non-English place of articulation contrasts. The first contrast was Hindi voiceless, unaspirated retroflex versus dental place of articulation, /t.a/ - /ta/, distinction. The second stimulus pair was Thompson glottalized velar with a glottalized uvular (or post-velar) place of articulation sounds, /k i/ - /q i /. The subjects were tested on their ability to discriminate multiple natural exemplars of the two non-native sound contrasts under four different perceptual set conditions i.e., (1) a native-language phonemic, (2) an acoustic categorizing, (3) a single-token phonetic, or (4) a within-category vocalic perceptual strategy. The results suggested that subjects have the sensori-neural ability to discriminate the acoustic parameters of the non-native phonetic contrasts, but they do not use that ability when required to discriminate full syllables. Thus it can be concluded that a simple neural atrophy explanation is not sufficient. Second, the results lead us to suspect that the ontogenetic decline cannot be explained by a shift in attentional focus since the perceptual set manipulation used in this experiment did not improve performance. Instead, these findings suggest that when subjects hear speech like sounds, a phonemic processing strategy is elicited, and that a more powerful manipulation that the attentional set induction used in to enable subjects to shift to different processing strategies.

Experiment 2 was designed to see if subjects could discriminate the full syllables if tested in a more sensitive testing procedure. Ten adult English-speaking subjects in the range of 18-35 were utilized. The same material as that of experiment 1 was utilized. Subjects were tested for AX discrimination task, by giving a form of training during the first block of 34 trials, followed by testing on three blocks of 34 trials. The results of this experiment supported the notion that the inability to discriminate non-native contrast cannot be explained by neural atrophy, but rather is best accounted for by the use of different processing strategies.

Experiment 3 was designed to determine if the familiarization (training) procedure used in experiment 2 facilitated accesses to a nonphonemic processing strategy. Ten naive, English-speaking subjects within the age range of 18-35 years were used. The same materials as of previous experiments were utilized. Subjects were tested in the AX procedure, similar to that of experiment 2, but no feedback was given during the first block of 34 trials. The results of this experiment indicated that subjects can discriminate the full syllable non-English contrasts when tested in an AX procedure even without being given any familiarization trials. It can be concluded that training is not the experimental variable to adopt a nonphonemic processing strategy, although training may improve performance.

In experiment 4, authors tested whether the evidence of discrimination obtained for full non-native syllables in experiments 2 and 3 was a function of the shorter ISI used in the AX than in the button-press task, rather than being the result of different testing procedures. Ten naive subjects between the ages of 18 and 35 were tested on both speech contrasts in the AX procedure without being given training. Testing conditions were identical to those described before, except the ISI between stimuli within each trial was 1500 ms rather than 500 ms, and the interval between trials was 3000 ms rather than 2500 ms. Results indicated that there is a difference between the results obtained for full syllable discrimination using the AX procedure with a short (500 ms) and a long (1500 ms) ISI. Apparently a memory trace is available following the 500 ms delay, which has decayed after 1500 ms. This memory trace enables the subjects to relinquish an exclusively phonemic processing strategy, and detect differences within phonemic categories.

In summary, these results indicate that the previously observed age-related decline in cross-language speech performance is not the consequences of a neural loss. It was found that under some circumstances adults discriminate speech sounds according to the phonemic categories of their native language, and under other circumstances discriminate the same sounds according to phonetically relevant category boundaries used in another, but not their native language. The ineffectiveness of the attentional manipulation used in experiment suggests that a model of attentional allocation may not be adequate to explain these results and that an explanation based on task-invoked processing strategies may be more appropriate. Further research using additional perceptual set manipulations is required to disambiguate these two possible explanations, however.

Morosan & Jamieson (1989) used both natural *Id/* and */6/* speech tokens and a synthesized /d/-/d/ continuum to study the Francophone perception of this stop-fricative contrast. Their results indicate that the Francophones had difficulty identifying the natural tokens correctly. In particular, they had problems identifying the /S/ items with the shortest frication durations, which suggest that they were relying only on the frication duration due to differentiate this contrast. Their identification of the /d/-/5/ continuum also failed to show a clear categorization of the series.

Polka (1991) studied English listeners' perception of the retro flex vs. dental place distinction in Hindi in four different voicing contrasts - prevoiced /d/ vs. /d/, voiceless unaspirated /t/ vs. /t/, voiceless aspirated /t^h / vs. /t^h / and breathy voiced /d^h / vs. /d^h /. Percentage of errors in the AX discrimination task for the four contrasts increased in the order of voiceless unaspirated, breathy voiced, voiceless aspirated and prevoiced. Differences in assimilation strategy (which take both phonemic and articulatory phonetic factors) could amount for the variability in the perceptual difficulty among the four contrasts. Acoustic-phonetic factors also play an important role in the perception of both assimilated and non assimilated contrasts.

Best & Strange (1992) found that Japanese listeners correctly labeled and discriminated English /w/- */r/* at higher rates than /r/ - */l/* Listeners with more English language experience responded to both contrasts more like native English speakers than those with less experience.

Yamada & Tohkura (1992) have explored interactions among stimulus materials, task variables, and individual differences in a set of studies of Japanese living in Japan, and Pisoni and his colleagues (Logan, Pisoni & Lively, 1991) have reported some results of a /r/- */l/* training study with Japanese in the USA. Three conclusions were drawn from the results of these and earlier studies.

(1) Phonotactic Context Effect: The /r/ - */l/* contrast in prevocalic positions (syllable-initial and syllable-initial clusters) is much more difficult for Japanese learners to differentiate perceptually than /r/ - */l/* in postvocalic positions (Mochizuki, 1981; Sheldon & Strange, 1982). Japanese have difficulty with post-vocalic /r/ - /l/ This is probably due to articulatory and acoustic differences between the allophones produced in prevocalic and postvocalic contexts. Improvement in /r/ -/l/ perception with training is also variable across phonotactic contexts; the perception of initial /r/ -

/l/ may be particularly difficult to alter (Strange & Dittmann 1984; Logan etal., 1991).

(2) Multiple Acoustic Cues: While F3 onset and transition differences in prevocalic /r/ - *l\l* are sufficient "cues" for native English speakers to distinguish phoneme categories, studies using synthetic speech materials indicate that Japanese learners of English rely more heavily on F2 transition information and temporal parameters of Fl which also vary systematically in natural speech (Underbakke et al., 1988, Yamada & Tohkura 1992). However, when synthetic speech containing multiple acoustic cues is used to assess perception consistency of identification and categoricalness of discrimination correlate well with Japanese ability to identify natural speech exemplars of the phonetic categories (Yamada & Tohkura 1992).

(3) Relation of American English Liquids and Glides to Native Japanese Categories: American English liquids /r/ and /l/ are phonetically dissimilar from Japanese *Ixl,* the latter being phonetically realized most often as an alveolar flap *Ixl.* Acoustically and perceptually, American English *Ixl* (and perhaps /l/) may be more similar to the Japanese (unrounded) glide /w/. This conclusion is supported by research using a synthetic /r*l* - */l/* continuum in which Japanese subjects labeled stimuli with intermediate F2 and F3 values as "w" when allowed this response alternative (Yamada & Tohkura 1992).

Pruitt (1993) studied American-English listeners' identification of Hindi retroflex and dental stops in varying voicing conditions. Results showed ind/l/idual variability in the difficulty to discriminate the contrasts. This again depends on the speaker and vowel contexts. It was also viewed that increasing the stimulus variability did not affect the transfer to new stimuli.

Thus, the above studies show that some of the non-nat/l/e contrasts, though not readily discriminated by adults can be easily taught in laboratory (e.g. voicing), while certain other contrasts (e.g. place) are more difficult to discriminate and require elaborate training procedure.

Many studies have examined Japanese learner's acquisition of English */r/* and /l/. Adult learners have great difficulty in distinguishing /r/ from /l/ (Goto, 1971). This is because Japanese does not have a */rl - /l/* contrast and neither English liquid is phonetically similar to any Japanese consonant. At a more abstract phonological level, the single liquid consonant found in Japanese might be considered similar to both English /r/ and */l/*. The perceptual relationship between English */r/* and */l/* and the Japanese */r/* is uncertain. It appears that English */l/* is perce/l/ed as being phonetically more similar to Japanese */r/* than English */r/*.

Takagi (1993) found that inexperienced Japanese listeners identified English word-initial */r/* and */l/* tokens as instances of Japanese */r/*. However in a rating experiment, the English */l/* tokens were judged to be more similar to Japanese */r/* than the English */r/* token were.

Sekiyama & Tohkura (1993) found that Japanese listeners identified word initial tokens of English */r/* most often as English *l\l* but also as Japanese /r/, /w/ (an unrounded velar approximant) and /*gl*. Conversely, English listeners identified syllable initial tokens of Japanese */r/* most often as */l/*

A few studies have examined other consonants as well in Japanese learners of English. Japanese learners of English are reported to produce English *lei* as /*s/* (Lado, 1957; Ritchie, 1968).

Lambacher, Marten, Nelson & Berman (1997) found that Japanese listeners had the most difficulty distinguishing between /e/ as /s/. When presented with a syllable containing *lei*, 28% of the participants chose /s/ and when presented with a syllable containing /s/, nearly 25% chose *lei*. The number of *lf/* for *le/* response was also quite high (13%).

Srid/l/ya (1997) investigated cross-language difference in the perception of stop consonants in Tamil and Malayalam by Tamil and Malayalam monolinguals and bilinguals. The two languages differ in voicing and aspiration. While both are phonemic in Malayalam, they are not so in Tamil. 90 subjects chosen for the study constituted three groups with 15 males and 15 females in each group - Tamil monolinguals, Malayalam monolinguals and bilinguals with Tamil as nat/l/e language and Malayalam as second language. The audio recorded words wee subjected to acoustic analysis. The first three terminal frequencies (frequency at the onset of the first, second and third formants of the following vowel), VOT (time difference between the onset of the vocal fold vibration and the articulatory release) and closure duration (time difference between the offset of vocal fold vibration and the onset of the burst) were measured using the waveform in the DSP Sonograph 5500. The results of the analysis showed that perceptual data correlated with the production data. Tamil monolinguals showed the poorest performance on the voicing contrasts, which was expected as voicing is not phonemic in Tamil. Aspiration contrast was well discriminated by Tamil monolinguals as they could form strong category goodness contrast and hence easily discriminable. Performance of the subjects improved when a combination of cues (aspiration and place; aspiration, voicing and place; place and voicing) was used than each of these in isolation. Bilinguals performed better in voicing and aspiration contrasts which are phonemic in Malayalam. They performed closer to the Malayalam monolinguals, reflecting the second language influence. This study implied that learning the second language widens the perceptual dimensions for stops, making them closer to the monolinguals. Further, it was speculated that contextual cues would play a role in discriminating the contrasts.

Guion, Flege, Yamada & Pruitt (2000) reported two experiments with nat/l/e speakers of Japanese. In experiment I, monolingual Japanese listeners identified English and Japanese consonants in terms of a Japanese category, and then rated the identifications for goodness of-fit to that Japanese category. Experiment II used the same set of stimuli in a categorical discrimination test. Three groups of Japanese speakers varying in English language experience and one group of nat/l/e English speakers participated. Contrast pairs composed of two English consonants, two Japanese consonants, and one English and one Japanese consonant were tested. The results indicated that the perce/l/ed phonetic distance of second language (L2) consonants from the closest first language (L1) consonant predicted the discrimination of L2 sounds. This study also investigated the role of experience in learning sounds in a second language. Some of the consonant contrasts showed significantly higher scores for the experienced than the relat/l/ely inexperienced Japanese groups. The perce/l/ed phonetic distance of L1 and L2 sound was found to predict learning effects in discrimination of L1 and L2 sounds in some case. In summary, the results obtained here indicated that certain English consonant contrasts are more difficult for Japanese adults to discriminate than others. The degree of perceptual difficulty seemed to depend on the extent to which the two members of a consonant contrast would be identified as instances of a single Japanese consonant category.

Polka, Colantonio & Sundara (2001) designed a study to examine the effects of age and language experience on the perception of a stop - fricat/l/e contrast *(Id* - 5/) by French and English speaking adults and French and English learning infants at 6-8 and 10-12 months of age. This contrast is phonemic in English but not in French, which has a dental voiced stop /d/, but lacks either *lei* or / 5/ phoneme. They observed a robust effect of language experience in the adults. French adults were consistently less accurate and showed greater inter subject variability compared to English adults. With respect to the 6-8 months olds, there was no evidence of language effect. These findings were similar to previous studies. Contrary to earlier studies, 10 and 12 months olds failed to show an effect of language experience. Thus, a robust effect of language experience was observed in adults but not in infants. The results were also discussed with respect to developmental patterns. For English listeners, comparable performance levels were maintained across both infant ages, whereas there was clear evidence of an increase in performance between 10-12 months and adulthood. These results indicate that for native listeners, perceptual differentiation of /d - 5/ improve with increasing age and language experience after 12 months of age. For French listeners, there were no significant changes in differentiation of /d - 6/ with increasing age, suggesting that, for French listeners the level of perceptual differentiation observed in infancy is maintained across development. The result of/b-v/ contrast showed the same pattern of age differences (adults greater than 6-8, and 6-8 equal to 10-12 months) for both English and French subjects. This rinding further confirms that the divergent patterns observed in French and English subjects' differentiation of/d - ð/ are attributable to differences in language experience. In summary, this study has shown that perceptual differentiation of /d - ð/ is not influenced by language experience in the first year of life, but is clearly affected by language experience by adulthood.

In a cross-language speech perception study done by Harnsberger (2001) on nasals, the abstract units such as phonemes or their context-dependent variant allophones were evaluated in a perceptual similarity test, employing a broad range of non-native stimuli and listener groups. In this study, Malayalam nasal consonants varying in place of articulation (bilabial, interdental, alveolar, retroflex, palatal, velar) were administered to three sets of listener groups with similar coronal nasal consonant inventories : dental-retroflex (Marathi, Punjabi), alveolar - retroflex (Tamil, Oriya) and alveolar (Bengali, American English). A multidimensional scaling analysis of the similarity scores revealed language specific differences that were not predictable from the listener groups' nasal consonant inventories, as represented by either phonemes or allophones. The dental-retroflex and alveolar groups showed intra-group differences in their clustering of nasal consonants, resulting in language specific perceptual spaces. Only the alveolar-retroflex group spaces were similarly organized. The results demonstrated that nat/l/e perceptual categories must be embedded at a "sub allophonic" level of description.

Overall, the results of this experiment point to the need for richer and more detailed descriptions of the perceptual categories of listener groups at sub allophonic levels of analysis. However, the results do not differentiate between candidate sub-allophonic units, such as weighted features or episodic distributions. To test a feature - weighting model of language would first have to be determined by acoustic analyses of appropriate sets of stimulus materials. From such analyses, a large set of candidate cues would then need to be evaluated in a series of perception tests using edited natural stimuli or synthetic stimuli. The purpose of such tests would be the measurement of the weights placed on each cue in the identification of the phoneme by nat/l/e listeners. If the cue weights for "equ/l/alent" phonemes or contrasts in two languages were adequately described, a feature-weighting model of cross-language speech perception could then be tested by presenting non-nat/l/e stimuli that are phonetically similar to these phonemes to nat/l/e speakers of both languages.

Jessy (2003) studied cross-language differences in the perception of Malayalam consonants by nat/l/e, non-nat/l/e and bilingual speakers. Malayalam consonants were studied, as Malayalam language is one of the languages, which has maximum number of consonants. Malayalam differs from Hindi in having lax consonants, more than one trill/lateral and dental vs. alveolar distribution. Ninety subjects were taken up constituting three groups, nat/l/e Malayalam speakers, nat/l/e Hindi speakers and bilinguals with Malayalam as the nat/l/e language and Hindi as the second language. Each group consisted of 10 subjects in the age range of 5-6 years, 9-10years and 18-35years. The audio-recorded stimulus was presented binaurally, and subjects were instructed to indicate if the two words in a pair were the 'same' or 'different'. The results indicated that Malayalam monolinguals scored higher than Hindi monolinguals and Malayalam bilinguals. Among the nat/l/e and bilingual listeners refinement in perceptual ability was seen across age group, suggesting that modification of the perceptual ability as a function of experience with the particular language. Also Malayalam monolinguals and bilinguals scored higher on perception of stop consonants and affricates compared to Hindi monolinguals. Hindi monolinguals perception of lax consonants was poorer compared to Malayalam monolinguals and bilinguals. This study implied that perceptual ability undergoes refinement with language experience. The authors speculated that contextual cues play a role in discriminating few contrasts for monolinguals and bilinguals.

Agarwal & Savithri (2005) investigated the Hindi listeners' ability to discriminate Malayalam lateral contrasts and the differences between adults and children in the ability to discriminate such contrasts. Malayalam language has 3 laterals - alveolar, retroflex and palatal - and Hindi has only the alveolar lateral. Forty subjects chosen for the study consisted of four groups with 10 subjects in each group. Group I and III consisted of nat/l/e Malayalam and nat/l/e Hindi speaking children in the age range of 4-6years. Group II and /l/ consisted of nat/l/e Malayalam and nat/l/e Hindi speaking adults in the age range of 19-21years. Two sets of stimuli with word pairs contrasting in laterals were prepared. The stimuli were audio-presented binaurally and the responses were recorded. Results indicated that nat/l/e Malayalam speakers discriminated palatal-retro flex contrast better than retro flex-alveolar contrast. The non-nat/l/e speakers discriminated retroflex-alveolar contrast better than palatal-retroflex contrast. Nat/l/e Malayalam speakers discriminated both laterals better than non-nat/l/e Hindi speakers and adults performed better than children. The results of this study do not support the Un/l/ersal theory and appears that a new model or theory of cross-language perception needs to be proposed. Non-nat/l/e consonant contrasts have been shown to vary in their discriminability from chance to near-nat/l/e level performance (Best, Me Roberts & Sithole, 1988; Polka, 1991; Pruitt, 1995). Moreover, listener groups varying in nat/l/e language but sharing a similar phonemic inventory can differ in the extent to which they find a g/l/en non-nat/l/e contrast difficult to discriminate. For instances, Japanese, Cantonese and Korean listeners have been shown to differ in their ability to identify and / or discriminate natural American English or Australian (l-r), despite the fact that all three listener languages have only one nat/l/e liquid phoneme (Henly & Sheldon, 1986; Ingram & Park, 1998).

Maharani (2006) investigated the cross-language differences in the perception of Tamil laterals and trills by nat/l/e (Tamil) and non-nat/l/e (Hindi) speakers. Tamil laterals and trills were studied as Tamil language has three laterals (alveolar /I/, retroflex /I./ and retroflexed palatal / 17) and two trills (flap */r/* and a trill /r/). Hindi has only one lateral */l/* and one flap */r/*. The material consisted of 80 Tamil meaningful mono/bi/tri syllabic words (minimal pairs) with laterals and trills. These words were spoken by a nat/l/e Tamil female speaker aged 22 years, which were audio-recorded. The words were paired with the same words (same) and words (different) that differed in laterals/trills. The word pairs were contrasting in word-medial and word-final positions. Four sets of word pairs with 20 word pairs each were prepared. In set I, each word pair had minimal pairs contrasting alveolar */l/* and retroflex /I./, in set II the contrast was between retroflex /I./ and retroflexed palatal */l/,* set /l/ had contrast between alveolar */l/* and retroflexed palatal /!/, and set /l/ had minimal pairs contrasting alveolar flap */r/* and alveolar trill /*xl.* These word pairs were used to test the discrimination ability. In addition, each word was paired with itself forming 40 word pairs in each set. These word pairs were used as catch trials. The word pairs were randomized in their corresponding sets and iterated thrice. The inter-stimulus interval used was 3 sec. Thus a total of 720 word pairs formed the material with each set consisting of 180 word pairs.

Two groups of subjects participated in the experiment. Group I consisted of 20 nat/l/e Tamil speakers and group II consisted of 20 nat/l/e Hindi speakers in the age range of 18-25 years. There were 10 males and 10 females in each group. All subjects had normal speech and hearing and had no neurological, organic, or psychological problems as reported. Subjects were tested ind/l/idually. They were audio-presented the stimulus through headphones and were instructed to record whether the two words in a pair were the 'same' or 'different' on a force choice response sheet provided to them. Percent 'same' and 'different' responses were calculated and group as well as gender related differences were analyzed. In general, results indicated that nat/l/e Tamil speakers discriminated significantly better than non-nat/l/e Hindi speakers on all four sets.

In adult speech perception the hypothesis that has become increasingly accepted is that adults may have difficulty discriminating phonetic contrasts that are not used in their nat/l/e language. The understanding of this area of research has changed considerably over the years. On the basis of research published in the 1960s and early 1970s, it appeared that adults might simply be unable to perce/l/e non-nat/l/e phonetic contrasts. Subsequent research showed that although adults have difficulty discriminating non-nat/l/e distinctions, some ability is still present, for e.g., adult discrimination of non-nat/l/e contrast can improve with training and even without training adults show a latent sensit/l/ity to non-nat/l/e phones if tested in sensit/l/e enough procedures. Recently, it has been shown that non-nat/l/e phonetic distinctions differ in their perceptual difficulty. Adults can discriminate some phonemes with virtually no training, whereas, hundreds of training trials are required to show only small improvement in adults' discrimination of other contrasts. Even for perceptually 'easy' contrasts, adults do not perform as well as nat/l/e speakers. Thus, there is a considerable support for the hypothesis that adults have more difficulty perce/l/ing non-nat/l/e than nat/l/e speech contrasts. Overall, the data on adult cross-language perception is consistent with a conceptualization of adult speech perception as being organized to process the nat/l/e language with the least effort and he greatest efficiency.

2. Variables in cross-language studies on consonants

The review reveals that the results depend on several factors like the environment, stimuli, interstimulus interval, perceptual paradigm and contextual cues. There is a Tetrahedral Model adapted to investigate the factors that influence perception of non-nat/l/e contrasts. This Tetrahedral Model was first proposed by Jenkins (1979) to describe memory phenomena and states that the outcome of an experiment in any cognit/l/e domain is a complex interaction of four variables: (1) subjects (abilities, interests, knowledge, purposes...), (2) orienting tasks (instructions, act/l/ities, apparatus...), (3) criterial tasks (recognition, problem solving, performance...), and (4) materials (sensory mode, physical structure, psychological organization and sequencing...). Table 2 shows some of the variables of each type that have been shown to influence the outcome of cross-language studies of speech.

Subject Variables	Nat/l/e Language (L1) Experience Second Language (L2) Experience Age (Critical or Sensit/l/e Periods) "Talent" for Language Learning (Ind/l/idual differences)
Orienting (Training) Task Variables	L2 Instruction (Usage vs.. perception/production drills) Laboratory Procedures Discriminating tasks vs. identification tasks Physical identity discrimination vs. Categorial (name identity) discrimination Prototype vs. gradient (fading) techniques One vs. many contexts; one vs. many speakers Blocked vs. mixed contexts and speakers
Criterial Task Variables	Laboratory Procedures (as above) Memory Load Inter-stimulus interval (ISI) Stimulus uncertainty Transfer Tasks (Tests of generalization of training) "Novel" stimuli: New contexts, new speakers Perception vs. production performance "Nat/l/e-Like" Performance Criteria
Stimulus Variables	Type of Contrast Vowel vs. consonants; voice vs. place contrasts Acoustic salience of information (temporal vs. spectral cues) Relation of Non-nat/l/e Contrast to Nat/l/e Categories Phonetic and Phonotactic Context Type of Stimulus Materials (Training and Testing) Synthetic speech (single vs. multiple acoustic cues) Natural speech (speaker intelligibility) Modified natural speech (e.g., truncation)

Table 2: Four types of variables that interact to determine the outcome of cross-language perception studies.

The Tetrahedral Model emphasizes that all four variables interact in complex ways to determine performance. Thus, there are no simple (or general) answers to questions concerning which non-nat/l/e contrasts will present the most difficulty for second language learners or to questions about which training procedures will be most efficacious in modifying phonetic perception. Answers to these questions will depend upon the subjects' linguistic experience and "talent" for language learning; the acoustic and articulatory structure of the phonetic categories to be learned and their relation to nat/l/e language categories; and how the criteria for perceptual learning and mastery are defined and performance is assessed. Which stimulus materials and which tasks are most effect/l/e in training language learners to perce/l/e non-nat/l/e contrasts will depend upon the subjects' initial level of performance, the contrasts to be learned, and the performance goals by which progress is measured.

The choice of stimulus materials and tasks also depends upon the experimenters' goals in conducting cross-language experiments. For instance, if the experimenter's purpose is an analytical assessment of whether nat/l/e and non-nat/l/e perce/l/ers attend to different sources of acoustic information, then synthetic speech materials are ideal (Underbakke, Polka, Gottfried & Strange, 1988;Yamada & Tohkura 1992).

Alternat/l/ely, if the primary concern is to train subjects to form new phonetic equ/l/alence categories, then natural speech materials containing the appropriate kinds of variation across speakers and contexts may be preferable (Logan, L/l/ely, & Pisoni, 1991).

Several factors affect the results of the studies also. These include the following:

(1) Environment of testing: There is some recent research showing that non-nat/l/e listeners show more difficulty perce/l/ing even relat/l/ely easy phones than do nat/l/e listeners under certain testing conditions. Takata & Nabelek (1990) compared nat/l/e English speakers to nat/l/e Japanese speakers on their performance in the modified Rhyme test. Results indicated that though the two groups performed significantly more poorly than the nat/l/e English speakers in conditions of noise and or reverberation. Not surprisingly, one of the more common errors for nat/l/e Japanese listeners was r/1 confusion.

(2) Stimulus used: Stimuli used in the perceptual studies of vowels has been of two types: (i) Natural/synthetic, (ii) Isolated vowels/vowels containing syllables i.e., as the CV or CVC. Therefore overall stimuli becomes of 4 types: (a) Natural isolated vowels (Fischer-Jorgenson, 1973), (b) Synthetic isolated vowels (Vinegard, 1970), (c) Natural vowels containing syllable (Flege, 1990), and (d) Synthetic vowels containing syllables.

All the four kinds of stimuli are used and it is still a matter of controversy. Whereas argument in favor of use of synthetic vowels is that they are speaker independent, but the same thing can act as disadvantage as synthetic vowels do not take into account normalization aspects (Verbrugge & Rakerd, 1985). Advantage of using isolated vowels is in its pure form and does not have coarticulatory effects. Whereas CVC minimal pairs add more meaning to it by adding phonetic context and making vowel perception of ambiguous vowels more categorical (Rakerd, 1984; Vinegard, 1970) and obviously CVC syllables make the coarticulatory and contextual variations near to constant (Rakerd, 1984).

(3) Inter-Stimulus Interval (ISI): Inter-Stimulus Interval has been found to affect level of linguistic participation (Werker & Tees, 1984). Werker & Logan (1985) studied using Hindi and retroflex/dental stimuli in English speakers. They tested subjects for f/l/e blocks of trials on three ISI conditions, 1.5 sec, 0.5 sec and 0.25 sec. Results indicated sensit/l/ity to non-nat/l/e phonetic contrast in shorter ISI conditions, as subjects could discriminate non-nat/l/e

phonetic cues within retro flex or dental category at 500 ms ISI, whereas in ISI above 1500 ms, subjects used phonemic cues. Flege, Munro & Fox, (1994) suggest use of 1 sec to 1.2 sec as ISI so that subject is able to retrieve phonetic cues from memory.

(4) Perceptual study tasks or paradigms used: Specific paradigms are used for specific research needs in cross-language vowel perception studies. Following paradigms have been used in the research reviewed (Flege et al., 1994; Werker & Tees, 1984; Fischer-Jorgenson, 1973).

a) Identification tasks
b) AX or similar/different or discrimination tasks
c) AXB method
d) ABX task
e) Oddity task
f) Rating procedures
g) Multi dimensional scaling

Identification task involve identification of the stimulus by the subject in the stimulus presented. This is easier than the other tasks and memory requirements are low.

AX or similar/different or discrimination task is also one of the simple tasks. In this, subject has to indicate whether X i.e., the target phase is similar to A, i.e., reference phase or different. In this task also, memory demands are less and is almost appropriate to test sensit/l/ity to the contrasts.

AXB task has A, X and B i.e., three sounds are represented success/l/ely to the subjects. A and B are standard stimuli, and X is the target stimulus. The subjects are required to judge whether X is more similar to A or to B. This is usually used to study assimilation and other processes.

ABX task has three sounds A, B and X which are presented success/l/ely to the subjects. A and B are the standard stimuli and X is the target stimulus. The subjects are required to confirm X to either A category or B category. This is used in categorical perception.

In the oddity discrimination, the subjects has to identify the odd item out of the three stimulus presented success/l/ely (triad) and encircle it. In case of ambiguity he/she is required to guess. It assesses identification indirectly and has high memory demands.

In the rating procedures, dissimilarity between two stimuli is rated on rating on rating scales, e.g., Flege et al., (1994) used a nine point scale with (1) as 'very similar' and (2) as 'very dissimilar'. This dissimilarity is rated on predetermined dimensions. Correlational analysis may be done which are helpful in obtaining weight age g/l/en to different dimensions and know which contrast is more readily discriminated. This procedure places high memory demands.

Multi-dimensional Scaling Analysis (MDS): The ratings are obtained and examined using MDS analysis. This technique is used to account for the perce/l/ed difference between pairs of stimuli by locating the stimuli within an 'n' dimensional perception space. The listeners mean ratings are entered into symmetrical matrices and then analyzed using ALSCAL, a program that assumes dissimilarity judgment for any of stimuli reflects underlying perceptual distance between them (Takane, Young & De Leruw, 1976). MDS are more sensit/l/e to acoustic differences.

The tasks discussed above are used in different kinds of research requirements and shown to g/l/e variations in results. Therefore a proper method should be selected to meet the investigators requirement.

(5) Contextual cues: Contextual cues can be of two types and increases identification of vowels (Rakerd, 1984). The contexts can be phonetic, phonological and acoustic (Werker & Logan, 1985). There are also some cues known as linguistic sentence context cues. House & Fairbanks (1953) show that vowel perception varies depending upon the identity of the consonant that precedes or follow it. Rakerd (1984) performed ind/l/idual scaling analysis and the study revealed two ways in which vowels in consonantial context can be said to have been perce/l/ed more linguistically meaningful dimensions of vowels were more integrated in perception when vowels were in context. According to Flanagan (1972), experiments have demonstrated that intelligibility of words (vowels) is substantially higher in grammatically correct meaningful sentences than when words are presented

randomly in isolation. The sentence context reduces number of alternat/l/e words among which listeners may decide. Centmayer (1973) presented synthetic vowel sound in isolation as well as within certain spoken linguistic environment to study the effect of linguistic context on vowel perception. They found that a change from isolated vowel sound to vowels within spoken words reduces the region of physical ambiguity, i.e., discrimination becomes more categorical. They also concluded that subject's vowel boundary is not fixed but varies within a certain range. The contextual cues can completely over ride the instantaneous boundary.

3. Theoretical explanations for cross-language differences in speech perception

Theoretically based explanations specifying which non-nat/l/e contrast would be easy or difficult to discriminate have been proposed by many researchers. Burnham (1986) suggested that there might be both fragile and robust non-nat/l/e contrasts. Fragile refers to phonetic contrasts that are both rare across the world's languages and of particular importance, are acoustically similar and it is due to the loss of these cues that difficulties arise in perception of non-nat/l/e contrasts in adults. Robust refers to contrasts that are widely distributed across the world's language and are acoustically less similar.

Flege (1990) proposed the Speech Learning Model, and hypothesized that phonetic similarity between LI and L2 phonetic categories is predict/l/e of L2 learning difficulty. He defines L2 phones as old, new, or similar, based on their (acoustic) phonetic "distance" from LI phonetic categories. Old categories are

L2 phone classes with boundaries that are nearly identical to LI categories and present few problems perceptually or product/l/ely; language learners continue to use LI equ/l/alence classification strategies to process these L2 phones. New phones are dissimilar from all LI categories and thus present perception and production problems initially. However, with experience, learners establish new equ/l/alence classes, which result in accurate perception and unaccented production. Similar L2 phones are assimilated to LI categories, which may facilitate perception and production initially. However, to the extent that there is a mismatch between LI and L2 phonetic boundaries, perceptual confusions and "accented" production may persist.

Polka (1991, 1992) has highlighted at least three independent factors that need to be considered when making predictions concerting the discriminability of non-nat/l/e contrasts among adults. These are functional phonetic status (phonemic contrast), substant/l/e phonetic status (phonetic variation), and acoustic differences (the absolute amount of measurable acoustic difference between members of non-nat/l/e contrast irrespect/l/e of phonetic status). She emphasized that all three of these factors need to be considered in assessing the discriminability of a non-nat/l/e contrast for subjects of any age.

Kuhl (1992) proposed the Nat/l/e Language Magnet Theory (NLM) to describe how the innate factors and experience with a specific language interact in the development of speech perception. Exposure to language results in the formation of language specific magnets. Thus, the difficulty in discriminating the non- nat/l/e sounds depends on their proximity to nat/l/e language magnet, that is the nearest it is to the nat/l/e language magnet; the more it will be assimilated by it, making it indistinguishable from the nat/l/e language sound. The second factor accounts that L2 learners can establish new L2 phonetic category, if they detect phonetic difference between an L2 sound and nearest L1 sound. SLM predicts that greater the perce/l/ed phonetic distance between an L2 sound and the closest L1 sound is, the more likely it is that phonetic difference between the sounds will be detected and a phonetic category eventually established. The acquisition of phonetic categories is thought to make L2 segmental perception more nat/l/e like because it enables the learner to base perception on L2 phonetic input without interference from prior learning.

Best (1995) has proposed that, for adults, language effects are evident for some contrast but not others because there are differences in the way in which the contrasting non-nat/l/e phones relate to the nat/l/e phonology. She has developed a perceptual assimilation model (PAM) that is able to account for some differences in an adult differentiation based on assimilation patterns. (Best 1990, 1993, 1994a). PAM indicates that phonological status is the predictor of the discriminability of non-nat/l/e contrast. Best, MacRoberts & Sithole (1988) have proposed that there are at least four kinds of non-native contrasts in terms of phonological status; they are as follows:

a) Assimilable,

b) Non-assimilable,

c) Category goodness and

d) Two category.

Assimilable contrasts are those in which each member of the contrasts can be assimilated to an intermediate phone in a nat/l/e language. These should be most difficult to discriminate (Example, glottal stop of Urdu to velar stop of Hindi).

Non-assimilable contrasts include phones that do not even sound at all like any possible phone from nat/l/e language. These contrasts are predicted to be most easily discriminable. (Example, /f/ in English and its absence in Tamil).

Category goodness refers to a non-nat/l/e contrast whose member can each be assimilated to an intermediate phoneme in the nat/l/e language, as in assimilable, but one which will stand out as clearly a better instance of that category than the others (Example, 1 of Tamil can be assimilated to '1.' in Kannada, but never to '1').

Two category refers to a non-nat/l/e contrast that consists of two non-nat/l/e phones each of which is assimilable to a contrasting phonemic category in the nat/l/e language (Example, /ph/ and /bh/ of Hindi can be assimilated to /p/ and *Pol* of English). According to ease of discrimination the four kinds of contrasts can be arranged as, Two category > Non-assimilable > Category goodness > Assimilable.

A Connectionist Model recently proposed by Behnke (1998) explains that language effects emerge later in development for some phonetic contrasts than for others. According to Behnke delays may occur either because general limitations in auditory processing during infancy make it difficult for infants to differentiate certain phonetic contrasts. Example, contrasts involving brief or low amplitude phones and or because differentiation of some contrasts may remain difficult until the child has gained lexical knowledge that serves to fine-tune phonetic processing.

For long cross language differences in perception of non-nat/l/e contrasts, consonants, vowels and suprasegmental aspects are studied and theoretical explanations are provided for the existing differences. The results of the cross language studies have indicated that adults have difficulty in discriminating non-nat/l/e distinctions. While virtually all cross-language phonetic perception studies, and, by extension all cross-language speech models assume some kind of phonetic similarity metric relating non-nat/l/e stimuli to nat/l/e categories, no current model currently incorporates an explicit, language general metric of similarity. Such a metric has been difficult to devise since its primit/l/e units are unclear. The simplest candidate primit/l/e would be the phoneme. There are, however, several problems with a phoneme based approach. While the phoneme contrasts in the language undoubtedly color our perception of most non-nat/l/e sounds, listeners in a number of studies have shown sensit/l/ity to greater phonetic detail in non-nat/l/e stimuli than can be captured in phonemic representations (Polka, 1991; Pruitt, 1995). This detail may correspond to the ind/l/idual gestural or acoustical cues that play a predominant role in signaling the identity of a nat/l/e sound in question. Adult listeners of different languages have been to shown to weigh these cues in a language-specific manner (Terbeek, 1977; Rochet, 1991).

India, a country with diverse linguistic groups, offers greater scope for research in cross-language speech perception. This would help in understanding the perceptual skills of bilinguals & multilinguals, the extent to which the native language and subsequent exposure to other languages would re-organize their perceptual skills. This has implications in teaching second language to adults & children.

Tamil and Malayalam are two Indian languages that differ in voicing and aspiration. While both are phonemic in Malayalam, they are not in Tamil. These differences were used to investigate the cross-language differences with perception of stop consonants. Stop consonants depicting voicing and aspiration differences as in the syllable initial position were used. Responses of Tamil and Malayalam speakers were elicited. If native language affects the perception, it would imply that Tamil speakers would perform poorly on voicing and aspiration contrasts and Malayalam speakers would perform better on the same. Therefore, it would be interesting to investigate perception of voicing and aspiration by Malayalam (native) and Tamil (non-native) speakers. In this context, the objectives of the present study were two fold and as follows:

1. To determine the perception of voicing in native and non-native speakers.
2. To determine the perception of aspiration/ murmur in native and non-native speakers.

Chapter iv

Method

In order to meet the objectives of the study, three experiments were carried out.

Experiment I: Perception of voicing in native and non-native speakers

As voicing is not phonemic in Tamil, and is phonemic in Malayalam, it is expected that the response in Tamil speakers to voiced-unvoiced continuum will not have a fixed pattern while that in Malayalam speakers will have a fixed pattern. Table 3 shows stop consonants in Malayalam and Tamil.

Place of articulation	Malayalam					Tamil				
	U		V			U		V		
	UA	A	UA	M	L	UA	A	UA^	M	L
Velar	k	kh	g	gh	K	k		g		K
Retroflex	t	th	d	d.h		t.		d		
Alveolar	t	th								
Dental	L	t.h	d.	d.h	T.	t.		d.		T.
Bilabial	p	ph	b	bh	P	p		b		P

Table 3: Plosives in Malayalam and Tamil (U - Unvoiced, V - Voiced, UA - Unaspirated, A - Aspirated, M - Murmured, L - Lax).

Subjects: Ten normal adults, in the age-range of 18-25 years, with mother tongue Malayalam and ten with mother tongue Tamil participated in this study. None of the subjects had any speech/ hearing problems as tested by the experimenter.

Material: Three voiced stop consonants, i.e. velar /g/, dental *IdJ* and bilabial /b/, and their counterpart unvoiced cognates were selected. The study used nonsense syllables, which will discard the probability of high response by native speaker. Three syllables with these voiced stops in the initial position and vowel /a/ in the final position were used. Syllables as uttered five times by a native Malayalam adult female speaker were recorded using a microphone, kept at a distance of 10 cm from the speaker's mouth, in the Speech Science Laboratory. These were recorded on a data acquisition system and were then digitized using a 12 Bit A/D converter, with a sampling rate of 12,000 Hz and stored on to the computer memory. The digitized waveform was displayed on the screen of the computer using the program DISPLAY on the SSL Pro2V2. (Voice and Speech Systems, Bangalore). The original VOT was measured using waveform display of SSL, for each of the stop consonants (figure 1). A continuum of voiced-unvoiced (-90 to +30 ms) was generated using Waveform Editor of SSL Pro2V2. The original lead VOT was truncated in steps of three pitch periods, from a continuum between lead VOT to 0 ms, i.e. till the burst was reached (figure 2). When VOT was '0', synthetic tokens with lag VOT were generated by inserting silence in steps of 10 ms between the burst and the following vowel, till 30 ms was reached (figuire 3). The tokens thus generated for each cognate were iterated thrice, randomized and recorded onto a CD with an inter-stimulus interval of three seconds. A total of 126 (14x3x3) tokens formed the material.

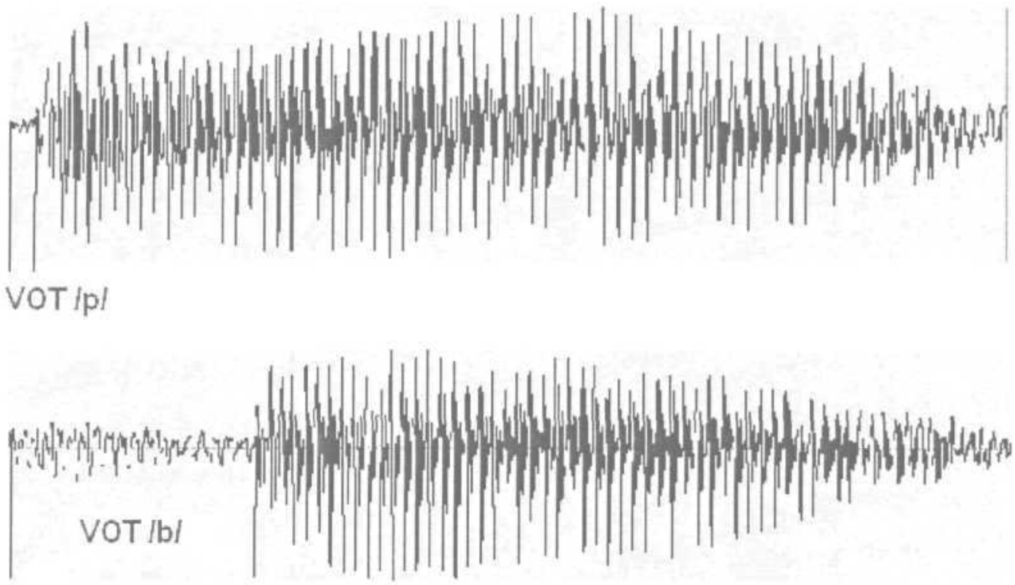

Figure 1: Illustration of measurement of original VOT using SSL Pro2V2.

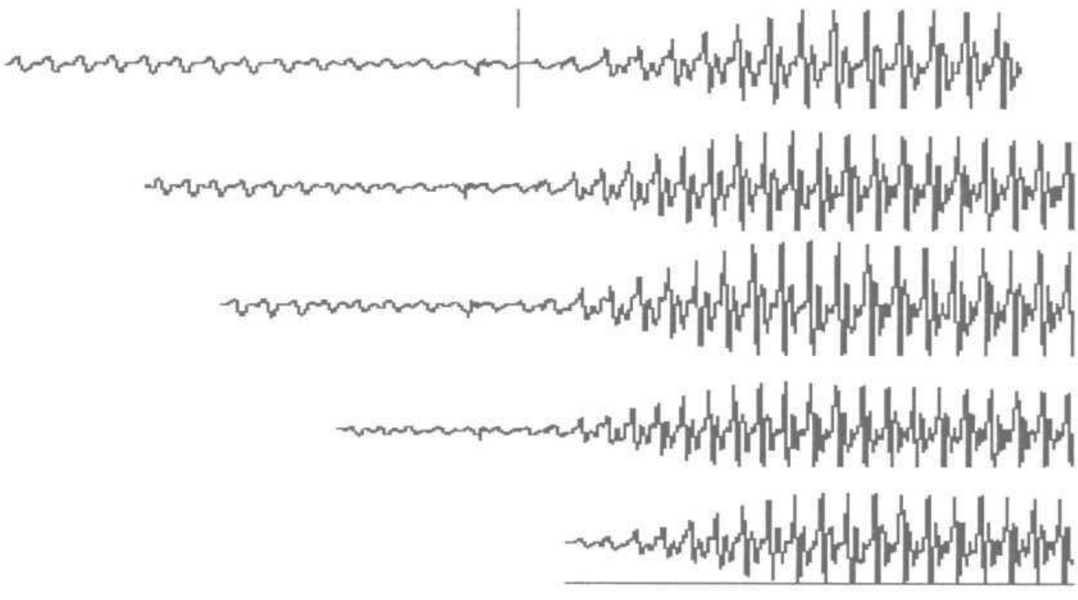

Figure 2: Illustration of truncation of original VOT in *Pol*.

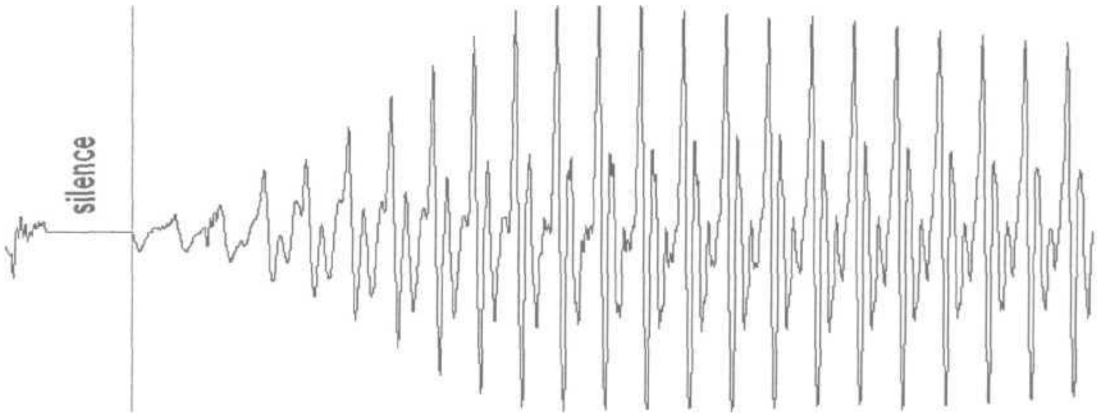

Figure 3:Illustration of insertion of silence between burst and following vowel.

Procedure: The material was audio-presented through headphones at comfortable listening levels. A binary forced-choice response was used. Subjects were instructed to carefully listen to each token and identify it as either voiced or unvoiced (for example: /k/ or ***Igl)*** and record their response on the sheet provided. The percent response for each token was calculated. Four measurements were calculated for each phoneme in two languages, which were then compared. They are as follows:

1. **50% cross over:** It is that point on the graph which was the actual or interpolated point about the acoustic cue continuum for which 50% of the subject's response corresponds to the voiced (unvoiced) category.

2. **Lower limit of Phoneme Boundary Width:** It is defined as that point along the acoustic cue continuum where an individual identified voiced (unvoiced) stop 75% of the time.

3. **Upper limit of Phoneme Boundary Width:** It is defined as the corresponding point for the identification of the (voiced) unvoiced cognate 75% of the time.

4. **Phoneme Boundary Width:** Between voicing and category phoneme boundary width is defined as the arc boundary cross point along the acoustic cue continuum and is determined by subtracting the lower limit from upper limit. Figure 4 illustrates these four measurements.

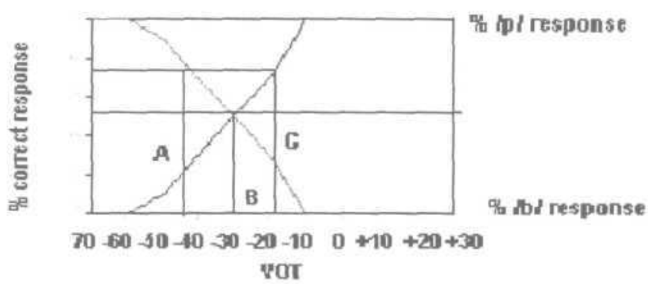

Figure 4: Illustration of 50% cross over (B), lower limit (A) and upper limit (C).

Experiment II: Perception of aspiration by native and non-native speakers.

As aspiration / murmur is not phonemic in Tamil and is phonemic in Malayalam, it is expected that the Tamil speakers' perception may not change from aspirated/ murmured to unaspirated and that the percept will change in Malayalam speakers.

Subjects: The same subjects as in experiment I participated in this experiment.

Material: Three syllables with aspirated stop consonants, i.e. velar /k^h/, dental /L^h/ and bilabial /p^h/, in the initial position and vowel /a/ in the final position were selected. Native speaker's productions of these syllables (thrice) were audio recorded and digitized and stored onto the computer memory. Using the Waveform Display, aspiration duration was measured for each syllable. The average aspiration duration was calculated for each syllable. The syllables with this average aspiration duration, were then presented to ten native speakers for perceptual analysis. If there was 100% identification of aspiration by native speakers, then that value of aspiration duration was considered for the study. A continuum of aspirated-unaspirated cognate (mean aspiration duration to '0' aspiration) was generated by truncating aspiration noise in steps of 10 ms till '0' aspiration duration was reached. The tokens for each cognate were iterated thrice, randomized and recorded onto a CD with an inter-stimulus interval of three seconds. A total of 78 (velar - 11*3 = 33, dental - 7*3 = 21, bilabial - 8*3 = 24) tokens formed the material. Figure 5 illustrates the original and truncated tokens of aspiration.

Truncated token

Figure 5: Illustration of truncation of aspiration in /ph/.

Procedure: The procedure was the same as in experiment **I.**

Experiment iv: Perception of murmured stops by native and non-native speakers.

Subjects: The same subjects as in experiment I and II participated in this experiment.

Material: Three syllables with voiced murmured stop consonants, i.e. velar /g^h/, dental /d^h/ and bilabial /b^h/, in the initial position and vowel /a/ in the final position were selected. As in experiment II, the mean murmur duration was calculated for each syllable. Syllables as uttered by a native Malayalam speaker (iterated thrice) were recorded and stored onto the computer memory and a continuum of voiced murmured-voiced unaspirated cognate (mean murmur duration to '0' aspiration) was generated, as in experiment II. The tokens for each cognate were iterated thrice, randomized and recorded onto a CD with an inter-stimulus interval of three seconds. A total of 60 tokens (velar - 7*3 = 21, dental - 7*3 =21, bilabial - 6*3 = 18) formed the material. Figure 6 illustrates the original and truncated tokens.

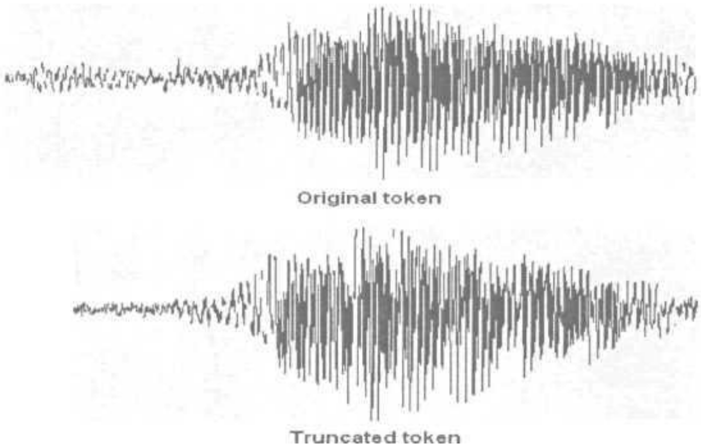

Figure 6: Illustration of truncation of murmur in /bh/.

Analysis: Statistical analysis was performed using a commercially available SPSS package (Version 10.0). One-way repeated measure ANOVA with groups as independent variable (mixed ANOVA) was carried out find out interaction between language group and place of articulation condition. Independent t test was done to compare language groups within place of articulation conditions. Finally, to compare place of articulation conditions within each language group, one -way repeated measures ANOVA was carried out. In ANOVA, if significant difference was observed between variables, Bonferroni's multiple comparison will be carried out for pair wise comparison.

Chapter iv

Results and Discussion

Results

Experiment I: Perception of voicing in native and non-native speakers

One-way repeated measure ANOVA with languages as independent variable (mixed ANOVA) showed no significant difference between native and non-native speakers in their response to the VOT continuum {$F(1, 13) = 0.559$, $p>0.05$}. Also, there was no significant difference between the three places of articulation in Malayalam {$F(2, 14) = 2.20$, $p>0.05$} and Tamil {$F(2, 12) = 2.213$, $p>0.05$}. Independent t test did not show any significant difference between places of articulation across languages. Table 4 shows the mean 50% crossover, lower limit, upper limit and phoneme boundary width for native (Malayalam) and non-native (Tamil) speakers for the VOT continuum and table *6* shows the mean 50 % cross over for 3 places of articulation in two languages.

	50% Cross over	LL	UL	PBW
Malayalam	5.3	0	7 (only for t-d)	7 (t-d)
Tamil	6	-5	29 (only for b-p)	34 (b-p)

Table 4: Mean 50% crossover, LL, UL and PBW (in ms) for VOT.

	Malayalam	Tamil
Velar	8	-
Dental	3	7
Bilabial	5	5

Table 5: Mean of 50% crossover (ms) for 3 places of articulation.

Figures 7 and 8 show identification functions for VOT in Malayalam and Tamil speakers. Malayalam speakers shifted their percept from voiced to unvoiced in all places of articulation. However, Tamil speakers did not shift their percept from voiced to unvoiced in velar place of articulation. Also, the shift was not above 60% for d-t continuum in Tamil speakers.

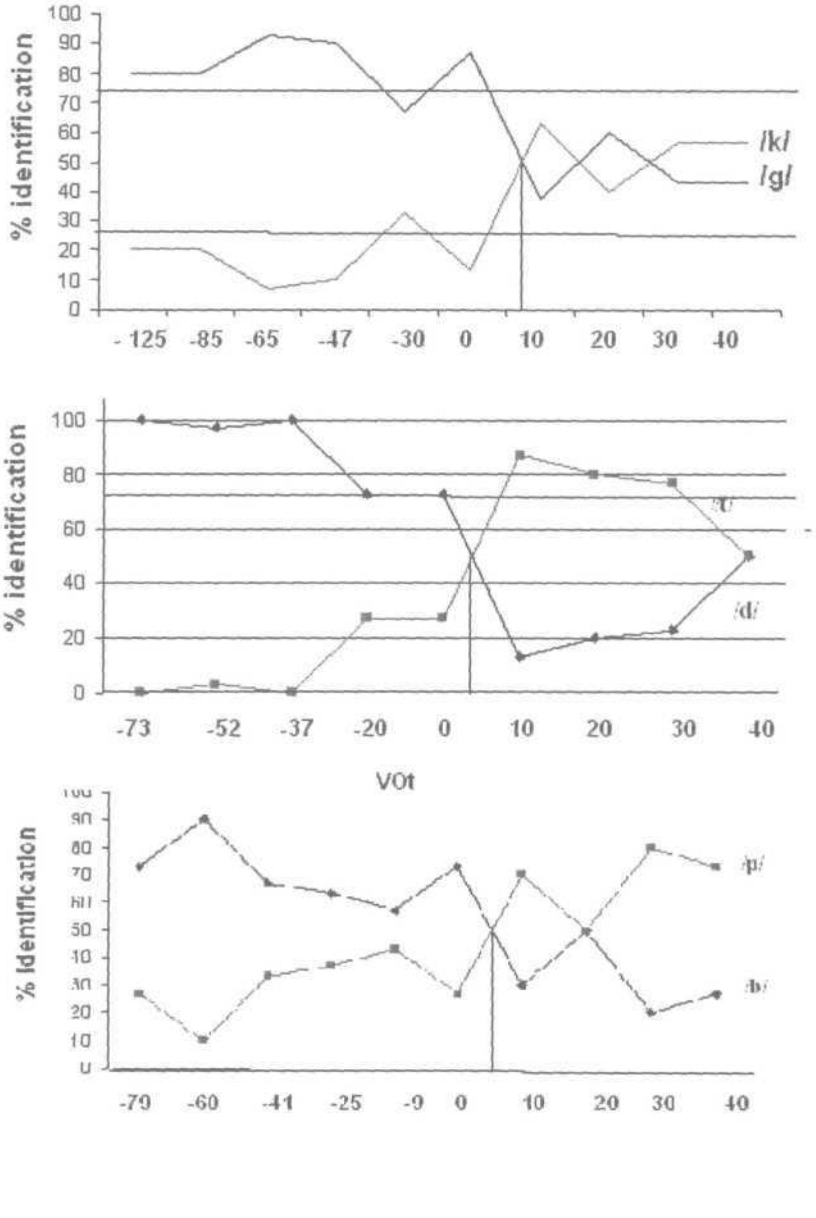

Figure 7: Identification function for VOT in Malayalam speakers.

In general, 50% crossover was observed in the positive region of VOT in both the languages. However, Malayalam had 50% cross over at shorter VOT compared to Tamil. Lower limit occurred at lag VOT in Malayalam compared to Tamil in which it occurred in lead VOT region.. While Malayalam speakers shifted their percept in lag VOT region, Tamil speakers had the shift in lead VOT region for bilabials.

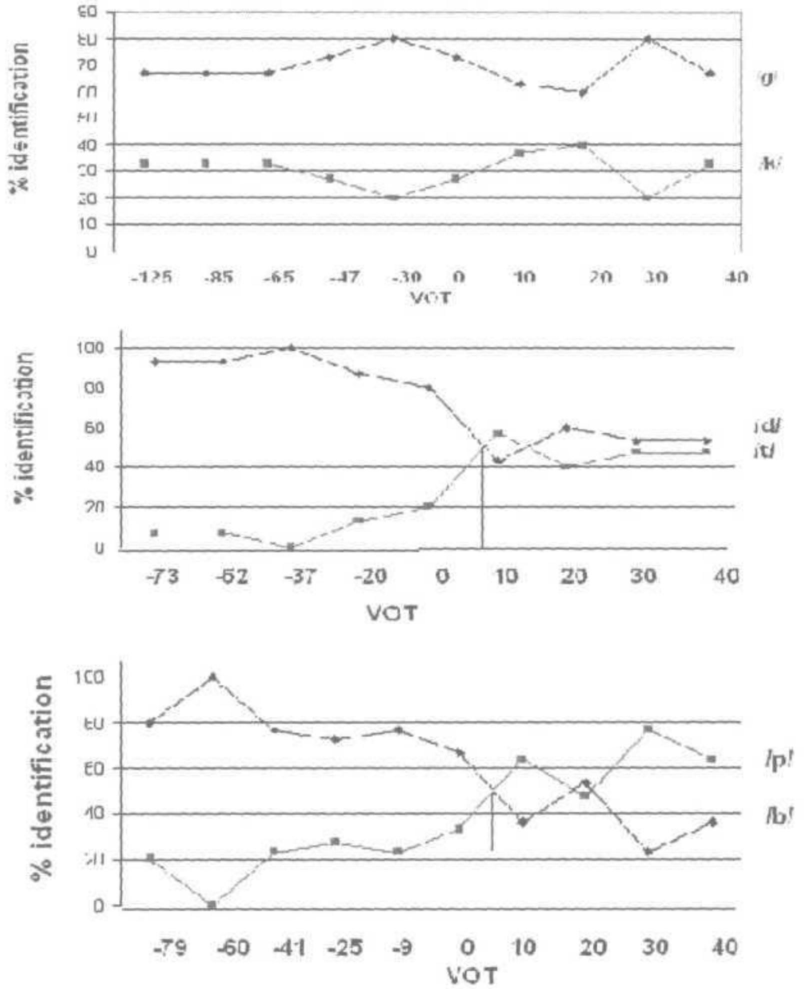

Figure 8: Identification function for VOT in Tamil speakers.

/g-k/ continuum: 50% cross over occurred in the lag VOT region in Malayalam. However, subjects in both the languages could not perceive unvoiced |k| 75% of times. Hence UL and PB W were not measured. Table 6 shows all measures for /g-k/ continuum.

	50% cross over	LL	UL	PBW
Malayalam	8	-	-	-
Tamil	-	-	-	-

Table 6: 50% crossover, LL, UL and PBW (in ms) for /g-k/ continuum.

Id. **- *t* /continuum:** 50% cross over occurred in the lag VOT region in both the languages. Malayalam had 50% cross over at a shorter VOT compared to Tamil. While Malayalam had lower limit in the lead VOT region, Tamil did not have a cross over. Though Tamil had 50% cross over, subjects could not identify the unvoiced plosive for 75% of the times. Hence no upper limit of PBW or PBW could be calculated in Tamil. Table 7 shows the measures for /d. -1. /continuum.

	50% cross over	LL	UL	PBW
Malayalam	3	0	7	7
Tamil	7	-	-	-

Table 7: 50% crossover, LL, UL and PBW (in ms) for *Id.* - t. /continuum.

/b-p/ continuum: 50% cross over occurred in the lag VOT region in both the languages at 5 ms. The lower limit was not obtained in Malayalam; while Tamil had lower limit in the lead VOT region. Table 8 shows all four measures in both languages.

	50% cross over	LL	UL	PBW
Malayalam	5	-	-	-
Tamil	5	-5	29	34

Table 8: Average 50% crossover, LL, UL and PBW (in ms) for /b-p/ continuum.

Experiment II: Perception of aspiration by native and non-native speakers.

One-way repeated measure ANOVA with groups as independent variable (mixed ANOVA) revealed that there was no significant difference between the two languages on aspiration duration $\{F(I, 18) = 0.450, p > 0.05\}$. A significant difference was observed between the places of articulation, within the Malayalam group $\{F(2, 18) = 3.102, 0.05 < p < 0.1\}$. Bonferroni's multiple comparison showed that a significant difference between bilabial and velar places of articulation ($p < 0.01$). However, there was no significant difference between the bilabial and dental, and dental and velar places of articulation within the Tamil language group $\{F(2, 18) = 1.144, p > 0.05\}$. Table 9 shows the mean 50% crossover, lower limit, upper limit and phoneme boundary width for native (Malayalam) and non-native (Tamil) speakers for the aspiration continuum and table 10 shows the mean 50 % cross over for 3 places of articulation in two languages.

	50% cross over	LL	UL	PBW
Malayalam	21.66	16.66	38.33	21.66
Tamil	36.33	10.0	64.66	33.33

Table 9: Average 50% crossover, LL, UL and PBW (in ms) for Aspiration duration.

	Malayalam	Tamil
Velar	31(8)	26(19)
Dental	24(13)	25(9)
Bilabial	21(6)	33(13)

Table 10: Mean and SD of 50% crossover (ms) for 3 places of articulation.

Independent t test showed significant difference between languages ($p<0.05$) on the bilabial condition. But, there was no significant difference between the groups in the dental and velar conditions.

In general, the 50% cross over occurred earlier in Malayalam compared to Tamil. Also, the phoneme boundary width was shorter in Malayalam compared to Tamil. Figures 9 and 10 show the identification functions in 2 languages for aspiration continuum.

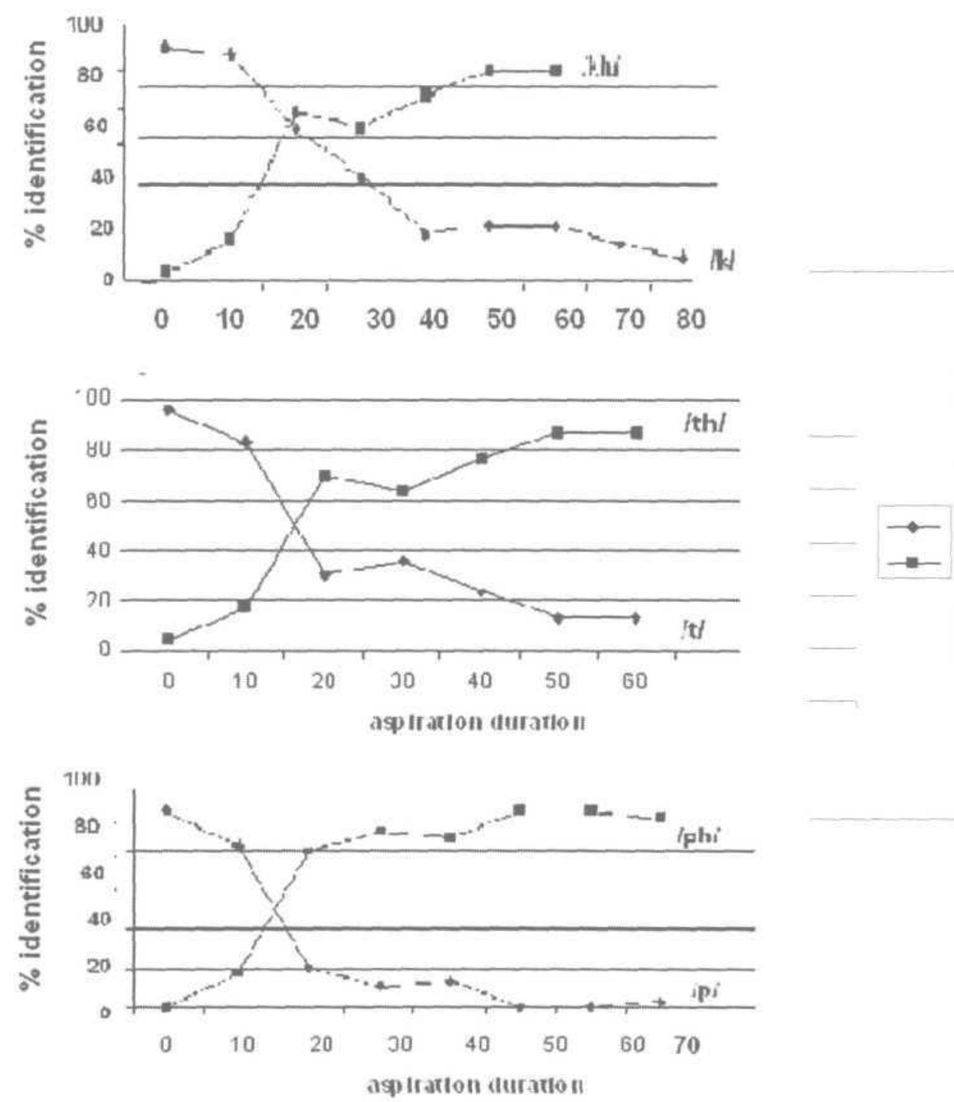

Figure 9: Identification function in Malayalam speakers for all continuums.

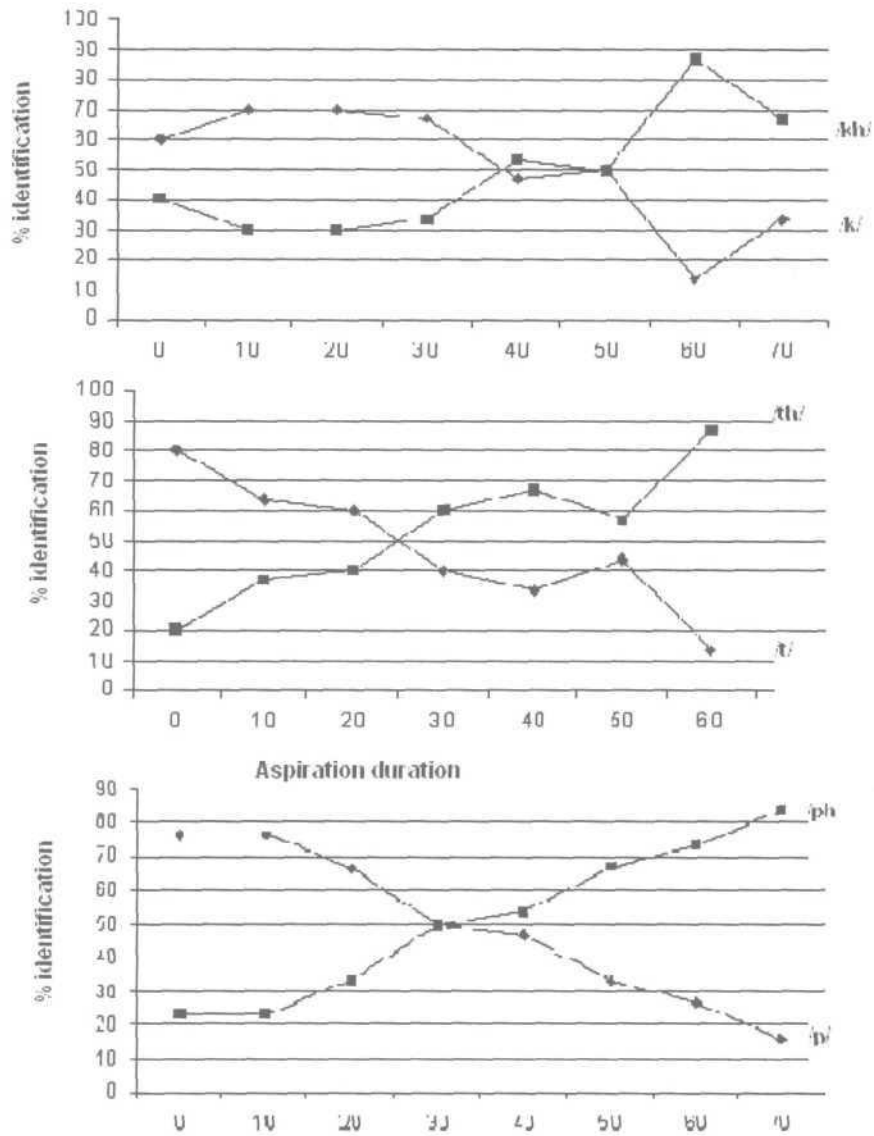

Figure 10: Identification function in Tamil speakers for all continuums.

Both language speakers shifted their percept from aspirated to unaspirated. 50% crossover occurred at shorter aspiration duration in Malayalam compared to Tamil. Lower limit occurred at shorter aspiration duration, and upper limit occurred at longer aspiration duration in Tamil compared to Malayalam. Hence, the phoneme boundary width was wider in Tamil than in Malayalam.

I p - p^h I continuum: 50% crossover occurred at shorter aspiration duration in Malayalam compared to Tamil. Lower limit and upper limit occurred at shorter aspiration duration in Malayalam, compared to Tamil. PBW was found to be wider in Tamil than in Malayalam. Table 11 shows all measures for | p - p^h | continuum.

	50% cross over	LL	UL	PBW
Malayalam	20	15	25	10
Tamil	35	20	70	50

Table 11: Average 50% crossover, LL, UL and PBW (in ms) for | p - p | continuum.

| **t - t^h** | **continuum:** 50% crossover occurred at shorter aspiration duration in Malayalam when compared to Tamil. Lower limit occurred at shorter aspiration duration, and upper limit occurred at longer aspiration duration in Tamil compared to Malayalam. Hence, the phoneme boundary width was wider in Tamil than in Malayalam. Table 12 shows all measures for /t - t^h / continuum.

	50% cross over	LL	UL	PBW
Malayalam	20	15	45	30
Tamil	30	to	60	50

Table 12: Average 50% crossover, LL, UL and PBW (in ms) for 11 - t^h | continuum.

| **k - k^h** | **continuum:** 50% crossover occurred at shorter aspiration duration in Malayalam when compared to Tamil. Though Tamil had 50% cross over, subjects could not judge the aspirated |k^h| continuum for 75% of the times. Hence no lower limit of PBW or PBW could be calculated in Tamil. Upper limit occurred at longer aspiration duration in Tamil compared to Malayalam. Table 13 shows all measures for |k - k^h | | continuum.

	50% cross over	LL	UL	PBW
Malayalam	25	20	45	30
Tamil	44	-	64	-

Table 13: Average 50% crossover, LL, UL and PBW (in ms) for | k - k^h | continuum.

Experiment iv: **Perception of murmured stops by native and non-native speakers.**

One-way repeated measure ANOVA with groups as independent variable (mixed ANOVA) indicated a significant difference between languages {$F (1, 18) = 3.592$, $p< 0.1$} for 50% crossover. A significant difference was observed between the places of articulation, within the Malayalam group {$F (2, 18) = 19.749$, $p<0.001$}. Bonferroni's multiple comparison showed that all three places of articulation were significantly different from each other. Bilabial and dental places of articulation were significantly different at $p< 0.05$, bilabial and velar places of articulation were significantly different at $p< 0.001$, and dental and velar places of articulation differed significantly at $p< 0.01$. A significant difference was also observed between the places of articulation, within the Tamil group {$F (2, 18) = 10.478$, $p<0.001$}. Bonferroni's multiple comparison showed that significant difference existed between bilabial and velar places of articulation at $p< 0.001$, and dental and velar places of articulation at $p< 0.05$, whereas no significant difference was obtained between bilabial and dental places of articulation. Table 14 shows the 50% cross over, lower and upper limits and PBW in both languages and table 15 shows the 50% for 3 place of articulation in 2 languages.

	50% cross over	LL	UL	PBW
Malayalam	26.0	18.0	37.66	19.66
Tamil	23.0	11.0	45.0	34.0

Table 14: 50 % cross over, lower limit, upper limit and PBW (ms) in both languages for murmur continuum.

	Malayalam	Tamil
Velar	36(10)	29(6)
Dental	24(5)	21 (10)
Bilabial	16(7)	13(8)

Table 15: Mean and SD of 50% crossover (ms) for 3 places of articulation.

Independent t test showed no significant difference between the languages in each of the three places of articulation. In general, 50% crossover for Malayalam speakers occurred at slightly longer murmur duration when compared to Tamil speakers. Lower limit occurred at shorter murmur duration, and upper limit occurred at longer murmur duration in Tamil compared to Malayalam. Hence, the phoneme boundary width was wider in Tamil than in Malayalam. Figures 11 and 12 show percent identification functions in both languages for all places of articulation.

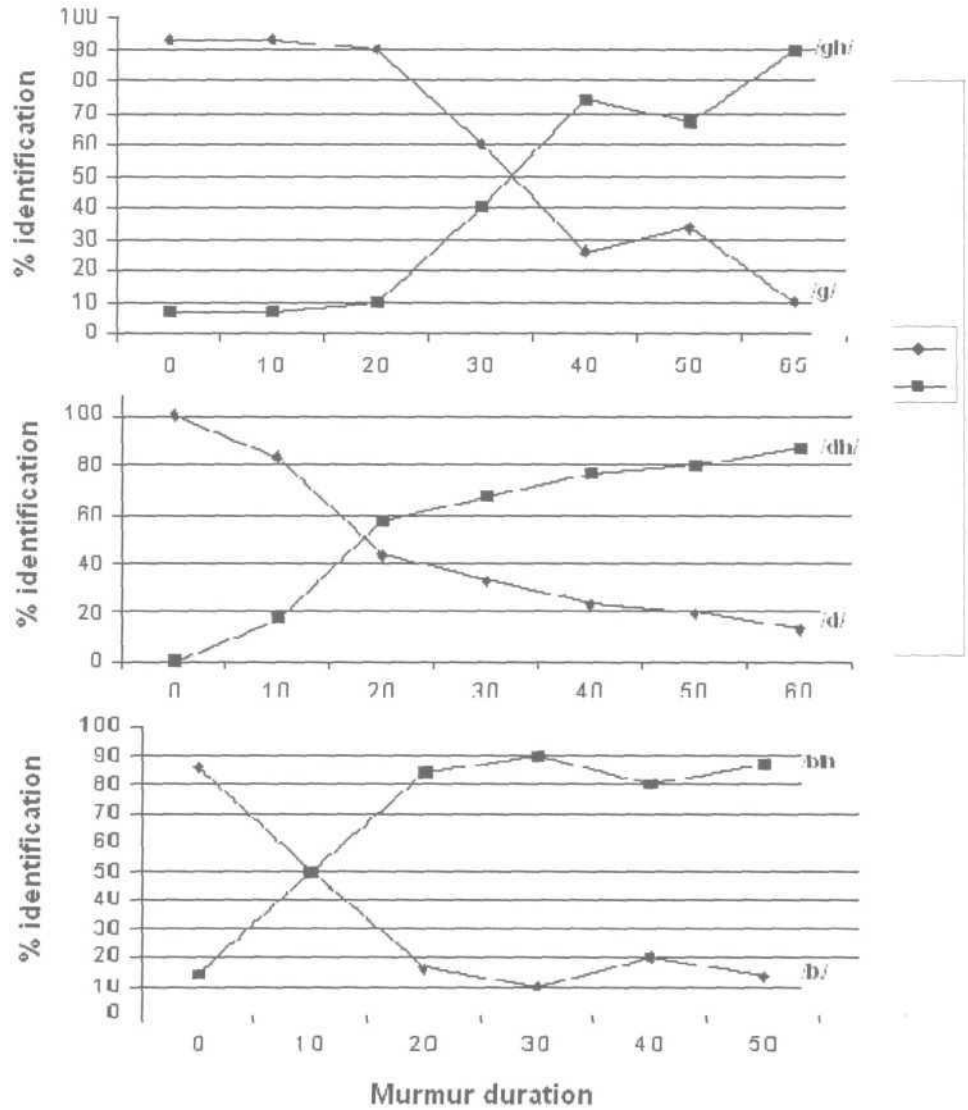

Figure 11: Identification function in Malayalam subjects for all continuums.

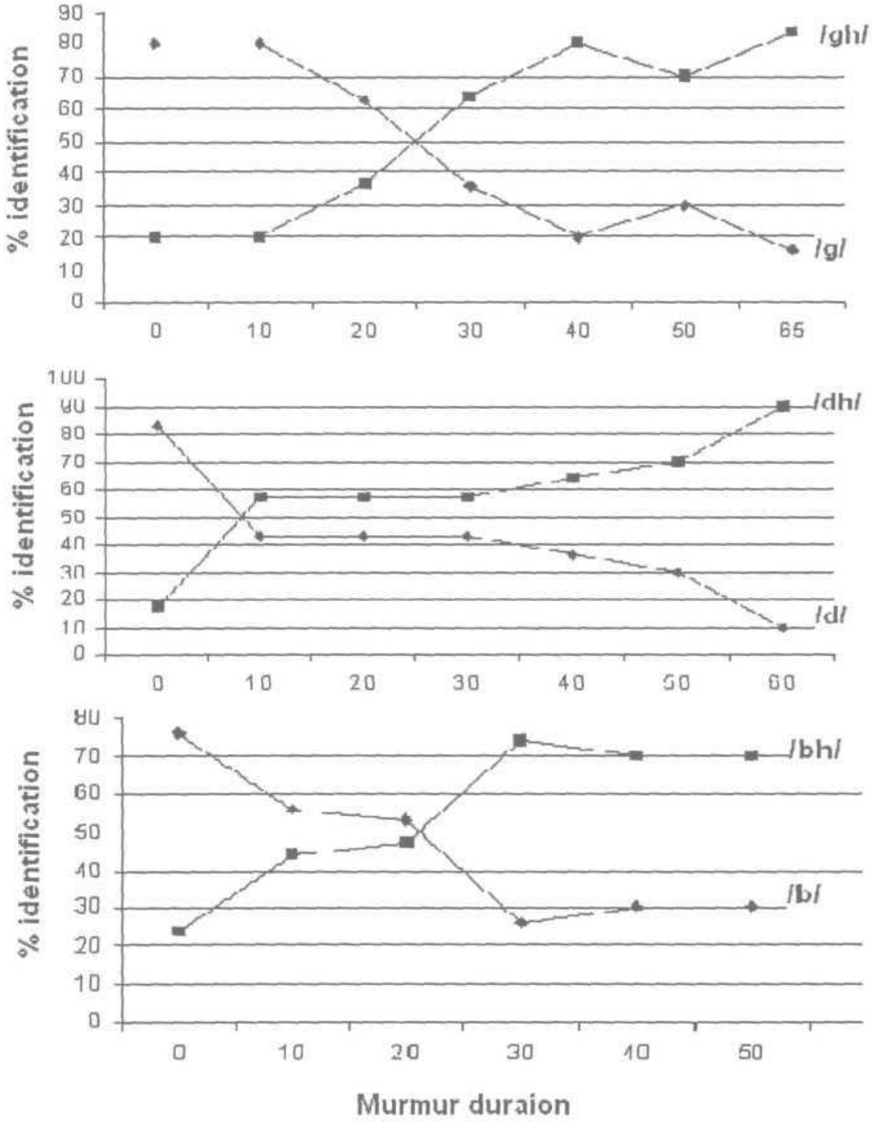

Figure 12: Identification function in Tamil subjects for all continuums.

| b - b^h | continuum: 50% crossover occurred at shorter murmur duration in Tamil when compared to Malayalam. Lower limit occurred at longer murmur duration, and upper limit occurred at shorter murmur duration in Malayalam compared to Tamil. PBW was found to be wider in Tamil than in Malayalam. Table 16 shows all measures for | b - b^h | continuum.

	50% cross over	LL	UL	PBW
Malayalam	36	8	24	16
Tamil	29	5	35	30

Table 16: Average 50% crossover, LL, UL and PBW (in ms) for | b - b | continuum.

| **d - d^h** | **continuum:** 50% crossover occurred at longer murmur duration in Malayalam when compared to Tamil. Lower limit occurred at longer murmur duration, and upper limit occurred at shorter murmur duration in Malayalam compared to Tamil. PBW was found to be wider in Tamil than in Malayalam. Table 17 shows all measures for | d- d^h | continuum.

	50% cross over	LL	UL	PBW
Malayalam	23	16	44	28
Tamil	21	8	58	50

Table 17: Average 50% crossover, LL, UL and PBW (in ms) for | d - d | continuum.

| **g** -g^h | **continuum:** 50% crossover occurred at longer murmur duration in Malayalam when compared to Tamil. Lower limit and upper limit occurred at shorter murmur durations in Tamil compared to Malayalam. However, PBW was found to be wider in Tamil than in Malayalam. Table 18 shows all measures for | g - g^h | continuum.

	50% cross over	LL	UL	PBW
Malayalam	16	30	45	15
Tamil	13	20	42	22

Table 18: Average 50% crossover, LL, UL and PBW (in ms) for | g - g^h | continuum.

Discussion

The results indicate several points of interest. First, the results of experiment I indicated that the percept changed from voiced to unvoiced stops (whenever it happened) in the lag VOT region in both languages. In Tamil, voicing is not phonemic while in Malayalam it is phonemic in the word-initial position. Therefore, it was expected that Tamil speakers would not be able to shift their percept from voiced to unvoiced stops. The results reflect this aspect as Tamil speakers did not shift their percept from voiced |g| to unvoiced |k|. Further, even when there was a shift in the percept from |d| to |t |, Tamil speakers were not able to identify |t | 75% of times. However, Malayalam speakers were also not able to identify /p/ 75% of times. This may be because of the status of voicing in Malayalam. Voicing is phonemic in word-initial position in Malayalam and there is a lax consonant in the word-medial position. Therefore, it is likely that Malayalam speakers are confused with voicing. This is in accordance with earlier studies on cross-language perception (Werker & Tees, 1981, 1984; Polka, 1991 and Pruitt, 1995). However, the question as to how were Tamil speakers, who do not have a distinction between |b-p| or |d -t |, able to identify these? Perhaps, rather than identifying voiced or unvoiced stops, they might have identified the durational differences between these two varieties of speech sounds. In general, a syllable with voiced stop was longer than that of an unvoiced stop in the present study. A comparison of 50% cross over points as obtained by various authors is in Table 19.

Author (year)	Language	Stimuli	50 % cross over
Yeni Komshian et al. (1967)	English	/g/	+35
Simon (1974)	English		+ 15-20
Zlatinet.al. (1975)	English		+25
Flege&Eefting(1986)	English		+36.2
	Spanish		19.9
Williams (1977)	English		+25
	Spanish		-4
lisker & Abramson (1967)	Thai		-20
Sathya(1996)	Telugu	g,d,b	-10 (children) -20 (adults)
Present study	Malayalam Tamil	g,d,b	5.3 6

Table 19: 50% cross over as obtained by various authors.

Williams (1977) speculated that Spanish listeners give greater weight to prevoicing as a cue to voicedness than English listeners and greater weight to the presence of an audible release burst and the lack of low frequency energy immediately following it, as cues to voiceless. The finding of Williams (1977) that the category boundary between /b/ and /p/ occurred along a VOT continuum occurred around -4 ms for Puertoricans who were monolingual Spanish speakers suggest that the phonetic processing of speech may be slowly attuned to the acoustic properties of stops found in a particular language (Aslin & Pisoni, 1980). Flege & Eefting (1986) imply that cross language research suggest that speakers of different languages may learn to perceive stops differently because they are exposed to different kinds of stop consonants. Further, English language environment listeners tend to identify both /b/ and /p/ as the phoneme /b/ and the prevoiced/ voiced contrast is physiologically irrelevant in English. This contrast is perceived categorically in other languages - for example Hindi, Spanish and Thai (Burnham, Earnshaw & Clark, 1991).

In addition to VOT there are three additional acoustic properties that vary in degree across art of the synthetic speech series (Williams, 1980). These variations are restricted to the voicing lag region of the continuum and are as follows:

(a) The presence/ absence of varying duration of aspiration or aperiodic energy in the interval between articulatory release and the onset of voicing. The presence of aspirated formants is an acoustic property that has been demonstrated to provide a positive cue for initial voicelessness to English listeners (Winitz et. al., 1975), (b) The absence of periodic acoustic energy at the level of Fl during periodic excitation of the vocal tract referred to as first formant cut back (liberman, Delattre, & Cooper, 1958). There is also evidence that the presence or absence of periodic energy in the region of Fl provides a perceptual cue for an initial contrast in voicing for English listeners (Delattre, liberman, & Cooper, 1955; liberman et. al., 1958; lisker, 1975), and (c) Differences in the degree and temporal extent of formant transitions under conditions periodic excitation of the vocal tract. There is some evidence that this acoustic variable may also provide a cue for initial voicing in English (Cooper, Delattre, liberman, Borst, & Gerstman, 1952; Stevens & Klatt, 1974; Summerfied & Haggard, 1974).

An examination of spectrograms of *Pol* and /p/ taken from a Malayalam speaker and those provided by Williams (1980) reveals that the stop consonants in Malayalam are entirely different from that of English in all the 3 parameters listed above. Neither aspiration in the lag VOT region, nor Fl cut back is present in Malayalam stop consonants. These differences in the acoustic properties of stops in

English and Malayalam might be reflected in the perception also with a 50% cross over in the lag VOT region for English and short lag VOT region for Malayalam.

While comparing the discrimination data obtained from adult speakers of Thai and English for synthetic bilabial consonants, Aslin & Pisoni (1980) comment that the relative discriminability in the 20 ms of voicing lag is greater than in the - 20 ms region of voicing lead despite that fact that the slopes of the labeling functions for Thai subjects in these regions are very nearly identical. They propose that the smaller incidence of discrimination of VOT (and TOT) differences in the minus region of voicing lead value is probably due to the generally poorer ability of the auditory system to resolve temporal differences in which a lower frequency component precedes a higher frequency component (for un voiced stop and lower frequency component - voicing - precedes a higher frequency component for - burst release - for voiced stop). Aslin & Pisoni (1980) further commenting on infant studies on VOT suggest that the discrimination of the relative order between the onset of first formant and higher formants is more highly discriminable at certain regions along the VOT stimulus continuum corresponding roughly to the location of the threshold for resolving these differences psycho-physically. In the case of temporal order processing, this falls roughly near the region surrounding +/ - 20 ms, a value corresponding to the threshold for temporal order processing (Hirsh, 1959). Further commenting on Pisoni's (1977) experiment on TOT (tone onset time), Aslin & Pisoni say that two distinct regions of high discriminability are present in the discrimination functions. Evidence of discrimination of VOT contrasts that straddle the -20 and + 20 ms regions of the stimulus continuum probably results from general sensory constraints on the mammalian auditory system to resolve small differences in temporal order and not from phonetic categorization.

Second, PBWs, in Tamil was wide. This indicates that Tamil speakers were not decisive about the tokens. This might reflect the fact that voicing is not phonemic in Tamil.

Third, results of experiment II indicate that the percept shifted from unaspirated to aspirated stops at shorter aspiration duration in Malayalam than in Tamil, for all three places of articulation. However, the shift at earlier aspiration duration in Malayalam compared to Tamil. In Tamil, aspiration is not phonemic while in Malayalam it is phonemic in the word-initial position. Therefore, it was expected that Tamil speakers would not be able to shift their percept from unaspirated to aspirated stops. The results reflect this aspect as even though Tamil speakers were able to shift their percept from unaspirated to aspirated stops, this happened at longer aspiration durations. Tamil speakers were not able to identify unaspirated |k| 75% of the time, hence PBW could not be calculated. Further, multiple crossovers in the identification function of Tamil speakers on | k - k^h | continuum revealed that they were confused and could not make a distinction between aspirated and unaspirated stop. Also, in general, PBWs were wider in Tamil, indicating that the confusion persisted for a long duration.

Fourth, results of experiment iv indicate that the percept shift from voiced to voiced murmur occurred at longer murmur duration in Malayalam than in Tamil. This could be attributed to the status of voicing in Malayalam. Malayalam murmured become deaspirated or devoiced as they are borrowed from Sanskrit. Since voicing is phonemic in Malayalam, the subjects might have had greater difficulty perceiving murmur at shorter murmur durations. Murmur is not phonemic in Tamil, while in Malayalam it is phonemic in the word-initial position. Therefore, it was expected that Tamil speakers would not be able to shift their percept from voiced to voiced murmur. However, results indicate that Tamil speakers were able to shift their percept, but with greater difficulty. This is evident from the wider PBWs obtained for Tamil subjects.

To summarize, the results indicated that the non-native speakers also could identify voicing, aspiration and murmur. This could be possible as the phoneme is not present in their language and hence it belongs to an entirely different category thus making the identification easy. However, the wider phoneme boundary widths indicate that the non-native speakers had more confusions than the native speakers. Further, aspiration and murmur are being eliminated unless the speaker is a professional voice user. Thus, the speakers of native and non-native languages might have performed similarly in some place of articulation. The present study has contributed information to the filed of cross-language perception. Both Malayalam and Tamil belong to the Dravidian group of languages. Thus, not much difference may be seen between languages. Future studies could be performed towards identification in more complex designs (ABX) and keeping longer inter-stimulus interval. Under such circumstances it may be apparent that the non-native speakers performance is poorer compared to a native speaker.

Chapter V

Summary and Conclusions

The present study investigated the perception of voicing, aspiration and murmur by Malayalam (native) and Tamil (non-native) speakers. Specifically a discrimination task was used. As voicing, aspiration and murmur are not phonemic in Tamil it was expected that Tamil speakers would perform poorer on tasks involving these parameters. In order to meet the objectives of the study, three experiments were carried out.

Experiment I was on perception of voicing in native and non-native speakers. Ten normal adults, in the age-range of 18-25 years, with mother tongue Malayalam **and** ten with mother tongue Tamil participated in this study. Three voiced stop consonants, i.e. velar /g/, dental /d/ and bilabial /b/, and their counterpart unvoiced cognates were selected. The study used nonsense syllables, which will discard the probability of high response by native speaker. Three syllables with these voiced stops in the initial position and vowel /a/ in the final position were used. Syllables as uttered five times by a native Malayalam adult female speaker were recorded using **a** microphone, kept at **a** distance of 10 cm from the speaker's mouth, in the Speech Science Laboratory. These were recorded on a data acquisition system and were then digitized using a 12 Bit A/D converter, with a sampling rate of 12,000 Hz and stored on to the computer memory. The digitized waveform was displayed on the screen of the computer using the program DISPLAY on the SSL Pro2V2. (Voice and Speech Systems, Bangalore). The original VOT was measured using waveform display of SSL, for each of the stop consonants. A continuum of voiced-unvoiced (-90 to +30 ms) was generated using Waveform Editor of SSL Pro2V2. The original lead VOT was truncated in steps of three pitch periods, from a continuum between lead VOT to 0 ms, i.e. till the burst was reached. When VOT was '0', synthetic tokens with lag VOT were generated by inserting silence in steps of 10 ms between the burst and the following vowel, till 30 ms was reached. The tokens thus generated for each cognate were iterated thrice, randomized and recorded onto a CD with an inter-stimulus interval of three seconds. A total of 126 (14x3x3) tokens formed the material.

Experiment II was on perception of aspiration by native and non-native speakers. Three syllables with aspirated stop consonants, i.e. velar /k^h/, dental /$t.^h$/ and bilabial /p^h/, in the initial position and vowel /a/ in the final position were selected. Native speaker's productions of these syllables (thrice) were audio recorded and digitized and stored onto the computer memory. Using the Waveform Display, aspiration duration was measured for each syllable. The average aspiration duration was calculated for each syllable. The syllables with this average aspiration duration, were then presented to ten native speakers for perceptual analysis. If there was 100% identification of aspiration by native speakers, then that value of aspiration duration was considered for the study. A continuum of aspirated-unaspirated cognate (mean aspiration duration to '0' aspiration) was generated by truncating aspiration noise in steps of 10 ms till '0' aspiration duration was reached. The tokens for each cognate were iterated thrice, randomized and recorded onto a CD with an inter-stimulus interval of three seconds. A total of 78 (velar - 11*3 = 33, dental - 7*3 = 21, bilabial - 8*3 = 24) tokens formed the material.

Experiment iv was on perception of murmured stops by native and non-native speakers. Three syllables with voiced murmured stop consonants, i.e. velar /g^h/, dental /d^h/ and bilabial /b^h/, in the initial position and vowel /a/ in the final position were selected. As in experiment II, the mean murmur duration was calculated for each syllable. Syllables as uttered by a native Malayalam speaker (iterated thrice) were recorded and stored onto the computer memory and a continuum of voiced murmured-voiced unaspirated cognate (mean murmur duration to '0' aspiration) was generated, as in experiment II. The tokens for each cognate were iterated thrice, randomized and recorded onto a CD with an inter-stimulus interval of three seconds. A total of 60 tokens (velar - 7*3 = 21, dental - 7*3 =21, bilabial - 6*3 = 18) formed the material.

The material was audio-presented through headphones at comfortable listening levels. A binary forced-choice response was used. Subjects were instructed to carefully listen to each token and identify it as either belonging to one or the other category (for example: /k/ or /g/) and record their response on the sheet provided. The percent response for each token was calculated. Four measurements - 50% cross over, lower and upper limit of phoneme boundary width, and phoneme boundary width were calculated for each phoneme in two languages.

Results of experiment I indicated no significant difference between languages (native and non-native speakers) in their response to the VOT continuum {$F(1, 13) = 0.559$, $p>0.05$}. Also, there was no significant difference between the three places of articulation in Malayalam {$F(2, 14) = 2.20$, $p>0.05$} and Tamil {$F(2, 12) = 2.213$, $p>0.05$}. Independent t test did not show any significant difference between places of articulation across languages. 50% crossover was observed in the positive region of VOT in both the languages. Malayalam speakers shifted their percept from voiced to unvoiced in all places of articulation. However, Tamil speakers did not shift their percept from voiced to unvoiced in velar place of articulation. Also, the shift was not above 60% for d-t continuum in Tamil speakers.

Results of experiment II indicated no significant difference between languages on aspiration duration {$F (1, 18) = 0.450$, $p > 0.05$}. A significant difference was observed between the places of articulation, within the Malayalam group {$F (2, 18) = 3.102$, $0.05 < p < 0.1$}. Bonferroni's multiple comparison showed that a significant difference between bilabial and velar places of articulation ($p < 0.01$). However, there was no significant difference between the bilabial and dental, and dental and velar places of articulation within the Tamil language group {$F (2, 18) = 1.144$, $p > 0.05$}. 50% cross over occurred earlier in Malayalam compared to Tamil. Also, the phoneme boundary width was shorter in Malayalam compared to Tamil.

Results of experiment iv showed a significant difference between languages {$F (1, 18) = 3.592$, $p < 0.1$} for 50% crossover. A significant difference was observed between the places of articulation, within the Malayalam group {$F (2, 18) = 19.749$, $p < 0.001$}. Bonferroni's multiple comparison showed that all three places of articulation were significantly different from each other. Bilabial and dental places of articulation were significantly different at $p < 0.05$, bilabial and velar places of articulation were significantly different at $p < 0.001$, and dental and velar places of articulation differed significantly at $p < 0.01$. A significant difference was also observed between the places of articulation, within the Tamil group {$F (2, 18) =$

10.478, $p<0.001$}. Bonferroni's multiple comparison showed that significant difference existed between bilabial and velar places of articulation at $p< 0.001$, and dental and velar places of articulation at $p< 0.05$, whereas no significant difference was obtained between bilabial and dental places of articulation. 50% crossover for Malayalam speakers occurred at slightly longer murmur duration when compared to Tamil speakers. Phoneme boundary width was wider in Tamil than in Malayalam.

The results indicate several points of interest. *First, the results of experiment 1 indicated that the percept changed from voiced to unvoiced stops (whenever it happened) in the lag VOT region in both languages.* In Tamil, voicing is not phonemic while in Malayalam it is phonemic in the word-initial position. Therefore, it was expected that Tamil speakers would not be able to shift their percept from voiced to unvoiced stops. The results reflect this aspect as Tamil speakers did not shift their percept from voiced |g| to unvoiced |k|. Further, even when there was a shift in the percept from |d| to |t |, Tamil speakers were not able to identify |t | 75% of times. However, Malayalam speakers were also not able to identify /p/ 75% of times. This may be because of the status of voicing in Malayalam. Voicing is phonemic in word-initial position in Malayalam and there is a lax consonant in the word-medial position. Therefore, it is likely that Malayalam speakers are confused with voicing. This is in accordance with earlier studies on cross-language perception (Werker & Tees, 1981, 1984; Flege, 1989; Polka, 1991 and Pruitt, 1995). However, the question as to how were Tamil speakers, who do not have a distinction between |b-p| or |d -t |, able to identify these? Perhaps, rather than identifying voiced or unvoiced stops, they might have identified the durational differences between these two varieties of speech sounds. In general, a syllable with voiced stop was longer than that of an unvoiced stop in the present study.

An examination of spectrograms of /b/ and /p/ taken from a Malayalam speaker and those provided by Williams (1980) reveals that the stop consonants in Malayalam are entirely different from that of English in that those in Malayalam don't have aspiration, and Fl cut back. Neither aspiration in the lag VOT region, nor Fl cut back is present in Malayalam stop consonants. These differences in the acoustic properties of stops in English and Malayalam might be reflected in the perception also with a 50% cross over in the lag VOT region for English and short lag VOT region for Malayalam.

While comparing the discrimination data obtained from adult speakers of Thai and English for synthetic bilabial consonants, Aslin & Pisoni (1980) comment that the relative discriminability in the 20 ms of voicing lag is greater than in the - 20 ms region of voicing lead despite that fact that the slopes of the labeling functions for Thai subjects in these regions are very nearly identical. They propose that the smaller incidence of discrimination of VOT (and TOT) differences in the minus region of voicing lead value is probably due to the generally poorer ability of the auditory system to resolve temporal differences in which a lower frequency component precedes a higher frequency component (for un voiced stop and lower frequency component - voicing - precedes a higher frequency component for - burst release - for voiced stop). Aslin & Pisoni (1980) further commenting on infant studies on VOT suggest that the discrimination of the relative order between the onset of first formant and higher formants is more highly discriminable at certain regions along the VOT stimulus continuum corresponding roughly to the location of the threshold for resolving these differences psycho-physically. In the case of temporal order processing, this falls roughly near the region surrounding +/ - 20 ms, a value corresponding to the threshold for temporal order processing (Hirsh, 1959). Further commenting on Pisoni's (1977) experiment on TOT (tone onset time), Aslin & Pisoni say that two distinct regions of high discriminability are present in the discrimination functions. Evidence of discrimination of VOT contrasts that straddle the -20 and + 20 ms regions of the stimulus continuum probably results from general sensory constraints on the mammalian auditory system to resolve small differences in temporal order and not from phonetic categorization.

Second, PBWs for VOT was wider in Tamil compared to Malayalam. This indicates that Tamil speakers were not decisive about the tokens. This might reflect the fact that voicing is not phonemic in Tamil.

Third, *results of experiment II indicate that the percept shifted from unaspirated to aspirated stops at shorter aspiration duration in Malay alam than in* ***Tamil,*** for all three places of articulation. However, the shift at earlier aspiration duration in Malayalam compared to Tamil. In Tamil, aspiration is not phonemic while in Malayalam it is phonemic in the word-initial position. Therefore, it was expected that Tamil speakers would not be able to shift their percept from unaspirated to aspirated stops. The results reflect this aspect as even though Tamil speakers were able to shift their percept from unaspirated to aspirated stops, this happened at longer aspiration durations. Tamil speakers were not able to identify unaspirated |k| 75% of the time, hence PBW could not be calculated. Further, multiple crossovers in the and Tamil belong to the Dravidian group of languages. Thus, not much difference may be seen between languages. Future studies could be performed towards identification in more complex designs (ABX) and keeping longer inter-stimulus interval. Under such circumstances it may be apparent that the non-native speakers performance is poorer compared to a native speaker.

References

Abramson, A., and lisker, L., (1965). Voice onset time in stop consonants: Acoustic analysis and synthesis. Fifth international Congress Acoustics, liege, Belgium (1965) cited by Zlatin, N. A, and Koenigsknecht, R. A. Development of voicing contrast: Perception of stop consonants. *Journal of Speech and Hearing Research,* 18, 541-553,1975.

Agarwal, D., & Savithri, S. R. (2005). Perception of Malayalam laterals by native and non-native speakers, *Proceedings of ISC A workshop on plasticity in speech perception* (PSP2005); London; UK; 207-210.

Aslin, R. N., & Pisoni, D. B. (1980). Some developmental processes in speech perception. In G.H. Yeni-Komshian, J.F. Kavanagh, and C.A. Ferguson (eds.), Child Phonolgy, Vol. 2 : Perception. New York : Academic Press.

Aslin, R. N., & Pisoni, D. B., Hennessy, B.L. & Perey, A.J. (1981). Discrimination of voice onset time by human infants : New findings and implications for the effect of early experience. Child Development, 52, 1135-1145.

Best, C. T. (1990). Adult perception of non-native contrasts differing in assimilation to native phonological categories, *Journal of the Acoustical Society of America,* 88, SI 77.

Best, C. T. (1993). Emergence of language specific constraints in perception of non-native speech perception. A window on early phonological development. In (Eds.) B. de Boysson- Bardies, S. de Schonen, P. Jusczyk, P. MacNeilage, and J. Morton *Developmental Neurocognition. Speech and face processing in the First year of life.* Dordreeht: Kluwer.

Best, C. T. (1994a). Learning to perceive the sound pattern of English, In (Eds.) C. Rovee- Collier and I. lipsitt. *Advances in infancy Research.* NJ: Albex, Hillsdale.

Best, C. T. (1995). A direct realistic view of cross-language speech perception, In W. Strange. Speech perception and linguistic experience. *Issues in cross-language research.* York: Baltimore.

Best, C. T., McRoberts, G. W., & Sithole, N. M. (1988). Examination of perceptual reorganization for non-native speech contrasts: Zulu

Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds "L" and "R", *Neuropsychologia*, 9, 317-323.

Guion, S., Flege, J., Yamada, R. & Priutt, J. (2000). An investigation of current models of second language speech perception : The case of Japanese adults perception of English consonants. J. Acoust. Soc. Amer, 107 (5), 2711-2724.

Hamsberger, D. J. (2000). A cross-language study of the identification of non-native nasal consonants varying in place of articulation, Journal *of the Acoustical Society of America*, 108, 764-783.

Henly, E., & Sheldon, A. (1986). Duration and context effect on the perception of |r| and |1| : A comparison of Cantonese and Japanese speakers, Lang. Learning, 36, 505-521.

Himanshu Khanna (1996). *Cross-language differences in vowel perception,* in M. Jayaram & S. R. Savithri (Eds.). Research at AIISH: Dissertation Abstracts, Vol. 3, 187-188.

Hirsh, I. J. (1959). Auditory perception of temporal order. *Journal of the Acoustical Society of America*, 31 (6), 759-767.

House & Fairbanks (1953). Cited in Vowels in consonantial contexts are perceived more linguistically than the isolated vowels: Evidence from an individual scaling study by Rakerd, B. *Status report on speech research* SR-76, 1984.

Ingram, J. C. L., & Park, S. G. (1998). Language, context and speaking effects in the identification and discrimination of English |r(and |1| by Japanese and Korean listeners, J. Acoust. Soc. Amer., 103, 1161-1174.

Jenkins, J. J. (1979). Four points to remember: A tetrahedral model of memory experiments. In L. S. Cermak & F. I.M. Craik (Eds.). *Levels of processing in human memory.* Hillsdale, NJ: Erlbaum.

George, J. (2003). *A cross-language study of consonant perception,* Unpublishd Master's dissertation submitted in part fulfillment of M. Sc. (Sp & Hg) to the University of Mysore.

Keating, P. A., Mikos, M. J. & Ganong, W. F. (1981). A cross-language study of range of voice onset time in the perception of initial stop voicing, J. Acoust. Soc. Amer., 70 (5), 1261-1271.

Kuhl, P. K. (1992). Early linguistic experience and phonetic perception: implications for theories of developmental speech perception, *Journal of phonetics,* 21,125-139.

Laskey, R. E., Syrdal-Laskey, A. & Klein, R. E. (1975). VOT discrimination by four-to six-and-a half-month old infants from Spanish environments, *Journal of Experimental Child Psychology,* 20, 215-225.

Lado, R. (1957). Linguistics across cultures (University of Michigan Press, Ann Arbor, MI).

Lambacher, S., Martin, W., Nelson, B. & Berman, J. (1997). Perception of English voiceless fricatives by native speakers of Japanese. In new sounds 97 : Procedings of the Third International Symposium on the Acquisition of Second Language Speech, edited by J. Leather and A. James (University of Klagenfurt Press, Klagenfurt, Austria), pp. 186-195

Liberman, A. M., Delattre, P. C, & Cooper, F. S. (1958). Cited by M. A. Zlatin in Voicng contrast: Perceptual and productive voice onset time characteristics of adults. *Journal of the Acoustical Society of America,* 56,981-994.

Lisker, L (1975). Is it VOT or a first formant transition detector? *Journal of the Acoustical Society of America,* 57 (6), 1547-1551.

Lisker, L., & Abramson, A. S. (1964). A cross language study of voicing in initial stops: Acoustic measurements. *Word*, 20, 384-422.

Lisker, L., & Abramson, A. (1967). Cited by M. A. Zlatin in Voicng contrast: Perceptual and productive voice onset time characteristics of adults. *Journal of the Acoustical Society of America,* 56, 981-994.

Lisker, L., & Abramson, A. S. (1970). *The voicing dimension: some experiments in comparative phonetics,* in Proceedings of the Sixth International Congress on phonetic science. Prague: Academia.

Lisker, L., & Abramson, A. S. (1972). The voicing dimension : Some experiments in comparative phonetics. In proceedings of the Sixth International Congress on phonetic sciences, Prague, 1967 (pp 563-567) Prague: Academia

Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English *It/* and /I/: A first study, Journal *of the Acoustical Society of America,* 89, 874-886.

MacKain, K. S., Best, C. T. & Strange, W. (1981). Categorical perception of English */r/* and /!/ by Japanese bilinguals. *Applied Psycholinguistics, 2*, 369-390.

Maharani. (2006). *Perception of Tamil Laterals and Trills by native and non-native speakers.* Unpublishd Master's dissertation submitted in part fulfillment of M. Sc. (Sp & Hg) to the University of Mysore.

Miyawaki, K., Strange, W., Verbrugge, R. R., liberman, A. H., Jenkins, J., & Fujimura, O. (1975). An effect of linguistic experience: The discrimination of /r/ and /I/ by native speakers of Japanese and English, *Perception & Psychophysics,* 18,331-365.

Mochizuki, M. (1981). The identification of [r] and [1] by native speakers of Japanese and English, *Perception & Psychophysics,* 18, 331-340.

Morosan, D. E., & Jamieson, D. G. (1989). Evaluation of a technique for training new speech contrasts. : Generalization across voices, but not word position or task, J. Speech Hear. Res. 32, 501-511.

Pisoni, D. B. (1977). Identification and discrimination of the relative onset of two component tones: Implication for the perception of voicing in stops. *Journal of the Acoustical Society of America,* 61, 1352-1361.

Pisoni, D. B., Aslin, R. N., Perey, A. J., & Hennessy, B. L. (1982). Some effects of laboratory training on identification and discrimination of voicing contrasts in stop consonants, *Journal of Experimental Psychology, Human Perception and Performance*, 8,297-314.

Polka, L. (1991). Cross-language speech perception in adults: Phonemic, phonetic and acoustic contribution. *Journal of the Acoustical Society of America*, 89, 2961-2977.

Polka, L. (1992). Characterizing the influence of native language experience on adult speech perception. *Perception & Psychophysics*, 52, 37-52.

Polka, JL, Colantonio, C, & Sundara, M. (2001). A cross-language comparison of (df - \d\ perception : Evidence for a new developmental pattern, *Journal of the Acoustical Society of America,* 100,577-592.

Pruitt, J. S. (1993). Training native English speakers to identify Hindi dental and retroflex consonants. *Journal of the Acoustical Society of America.*, (4) 2392 (A).

Pruitt, J. S. (1995). The perception of Hindi dental and retroflex stop consonants by native speakers of Japanese and American English, doctoral dissertation, University of South Florida, Tampa.

Rakerd, B. (1984). Vowels in consonantial contexts are perceived more linguistically than the isolated vowels: Evidence from an individual scaling study. *Status report on speech research,* SR-76, 87-114.

Ritchie, W. (1968). On the explanation of phonic interference. Language learning, 18, 183-197.

Rochet, B. L. (1991). " Perception of the high vowel continuum: A cross-language study," in Proceedings of the Twelfth International Congress of Phonetic Sciences (Univ. of Province, Aix-en-Provence, France), Vol. 4, pp. 94-97.

Sathya. K. (1996). Personal communication.

Savithri, S. R., & Sridevi, N. (1991). Cross-language study of speech perception. *Journal of the Indian Speech and Hearing Association,* 1, 14-20.

Savithri, S. R., Sreedevi, N., and Santhosh, M. (2001). Voicing Contrast in Indian Languages: Some Acoustic Measurements. Report of the ARF Project, AIISH, India.

Sekiyama, K., & Tohkura, Y. (1993). Inter-language differences in the influence of visual cues on speech perception. Journal of Phonetics, 21,427-444.

Sheldon, A., & Strange, W. (1982). The acquisition of/r/ and *iv* by Japanese learners of English: Evidence that speech production can precede speech perception. *Applied Psycholinguistics,* 3, 243-261.

Simon, C. (1974). Some aspects of development of speech production and perception in children. *Speech communication seminar, 7-15, Stockholm.* London: John Wiley and Sons.

Singh, S., & Black, J. W. (1966). Study of twenty-six intervocalic consonants as taken and recognized by four language groups : *Journal of the Acoustical Society of America.,* (39), 371-387.

Snow, C. E., & Hoefnagel-Hohle, M. (1978). The critical period for language acquisition : Evidence from second language learning. Child Development, 49, 1114-1128.

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

Strange, W., & Dittmann, S. (1984). Effects of discrimination training on the perception of /r-1/ by Japanese adults learning English. *Perception & Psychophysics,* 3 6, 131 -145.

Strange, W., & Jenkins, J. J. (1978). "The role of linguistic experience in the perception of speech," In R. D. Walk and H. L. Pick, Jr. (Eds.). *Perception and experience,* New York: Plenum.

Streeter, L. A., & Landauer, T.K. (1976). Effects of learning English as a second language on the acquisition of a new phoneme contrast, *Journal of the Acoustical Society of America.,* 109(5), 2190-2201.

Sridivya (1997). Cross-language perception of stops in Tamil and Malayalam, LJnpublishd Master's dissertation submitted in part fulfillment of M. Sc. (Sp & Hg) to the University of Mysore.

Stevens, K. N., & Klatt, D. (1974;). Role of formant transitions in the voiced-voiceless distinction for stops. *Journal of the Acoustical Society of America,* 55, 643-659..

Summerfied, A. Q., & Haggard, M. P. (1974). Perceptual processing of multiple cues and contexts: Effects of following vowel upon stop consonant voicing. *Journal of Phonetics, 2,* 279-295.

Sussman, J. E., & Carney, A. E. (1989). Effects of transition length on the perception of stop consonants by children and adults. *Journal of Speech and Hearing Research,* 82, 151-160.

Takagi (1993). Perception of American English */r/* and *iv* by Adult Japanese learners of English: A unified view, unpublished University of California, Irvine, Ph.D. Dissertation.

Takane, Y., Young, F. & DeLeruw, J. (1976). Nonmetric individual differences in multidimensional scaling. An alternating least square method with optimal scaling features. *Psychometrica,* 42, 57-67.

Takata, Y., & Nabelek, A. K. (1990). English consonant recognition in noise and in reverberation by Japanese and American listeners. *Journal of the Acoustical Society of America,* 88, 663-666.

Terbeek, D. (1977). A cross-language multi-dimensional scaling study of vowel perception, Working Papers in Phonetics 37, UCLA, Los Angeles, CA.

Trehub, S. E. (1976). The discrimination of foreign speech contrasts by infants and adults, *Child Development, Al, A66-A72.*

Underbakke, M., Polka, L., Gottfried, T. L., & Strange, W. (1988). Trading relations in the perception of/r/-iv by Japanese learners of English. *Journal of the Acoustical Society of America,* 84, 90-100.

Verbugge, R. R., & Rakerd, B. (1985). linguistic and acoustic correlates of the perceptual structure found in an individual difference scaling study of vowels, *Journal of the Acoustical Society of America, 11,* 296-301.

Vinegard, M. D. (1970). A direct magnitude scaling method to investigate categorical vs continous mode of speech perception. *Status report on speech research,* SR 21-2235, 1970, 147-156.

Vinay, R. (1990). Cross-language study of velar and bilabial stop perception, Unpublishd Master's dissertation submitted in part fulfillment of M. Sc. (Sp & Hg) to the University of Mysore.

Werker , J. F., & Logan, J. S. (1985). Cross-language evidence for three factors in speech perception. *Perception and Psychophysics,* 37, 35-44.

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life, Papers in speech communication : Speech perception, 733-747.

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,* 1865-1878.

Werker, J. F., Gilbert, J. H. V., Humphrey, K., & Tees, R. C. (1981). Developmental aspects of cross-language speech perception, *Child Development,* 52, 339 - 355.

Williams, L. (1977). The perception of stop consonant voicing by Spanish-English bilinguals. *Perception and Psychophysics,* 21, 289-297.

Williams, L. (1979). The modification of speech perception and production in second-language learning. *Perception & Psychophysics,* 26: 95-104.

Williams, L. (1980). Phonetic variation as a function of second language learning. In G. H. Yeni-Komshian, J. F. Kavanagh & C. A. Ferguson (Eds.). *Child Phonology, Vol. 2: Perception.* New York: Academic Press.

Winitz et. al.. (1975). Cited by Williams, L. (1980). Phonetic variation as a function of second language learning. In G. H. Yeni-Komshian, J. F. Kavanagh & C. A. Ferguson (Eds.). *Child Phonology, Vol. 2: Perception.* New York: Academic Press.

Yamada, R. A., & Tohkura, Y. (1992). Perception of American English *hi* and /I/ by native speakers of Japanese. In Y. Tohura, E. Vatikiotis-Bateson, & Y. Sagisaka (Eds.). *Speech perception, production and linguistic structure.* Tokyo: IOS press.

Yeni-Komshian, G., Preston, M, & Cullen, J. (1967). A brief report on a study of imitation of synthetic speech stimuli by children. Haskins Laboratories: Status Report on Speech Research, SR-11, 48-49.

Yoshida, K., & Hirasaka, F. (1983). The lexicon in speech perception. Sophia linguistica, 11, 105-116.

Zlatin, M. A., & Koenegskenecht, R. (1975). Development of the voicing contrast: Perception of stop consonants. *Journal of Speech and Hearing Research*, 18, 541-553.