8-2008

Production of Bangla stops by native English speakers learning Bangla: An acoustic analysis

Dustin D. Miller

PRODUCTION OF BANGLA STOPS BY NATIVE ENGLISH SPEAKERS LEARNING BANGLA: AN ACOUSTIC ANALYSIS

by

Dustin D. Miller
Bachelor of Science, Liberty University, 2000

A Thesis

Submitted to the Graduate Faculty

of the

University of North Dakota

in partial fulfillment of the requirements

for the degree of

Master of Arts

Grand Forks, North Dakota
August
2008

This thesis, submitted by Dustin D. Miller in partial fulfillment of the requirements for the Degree of Master of Arts from the University of North Dakota, has been read by the Faculty Advisory Committee under whom the work has been done and is hereby approved.

Joan Baart
Chairperson

Steve Parker

Stephen A. Marlett

This thesis meets the standards for appearance, conforms to the style and format requirements of the Graduate School of the University of North Dakota, and is hereby approved.

Joseph N. Benoit
Dean of the Graduate School

Date

PERMISSION

Title	Production of Bangla Stops by Native English Speakers Learning Bangla: An acoustic analysis
Department	Linguistics
Degree	Master of Arts

In presenting this thesis in partial fulfillment of the requirements for a graduate degree from the University of North Dakota, I agree that the library of this University shall make it freely available for inspection. I further agree that permission for extensive copying for scholarly purposes may be granted by the professor who supervised my thesis work or, in his absence, by the chairperson of the department or the dean of the Graduate School. It is understood that any copying or publication or other use of this thesis or part thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of North Dakota in any scholarly use which may be made of any material in my thesis.

Signature _____Dustin D. Miller_____

Date _____

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

–A	Unaspirated
+A	Aspirated
ACT	After Closure Time
CD	Closure Duration
CV	Closure Voicing
Fin.	Final
Init.	Initial
L1	Mother Tongue
L2	Non-Mother Tongue
Med.	Medial
ms	Milliseconds
NOT	Noise Offset Time
PVD	Preceding Vowel Duration
SLA	Second Language Acquisition
–V	Voiceless
+V	Voiced
VOT	Voice Onset Time
VVOT	Vowel Voicing Onset Time

ACKNOWLEDGMENTS

It is with a sense of pleasure and satisfaction that I complete this study after many long hours given to it over a period of several years. I would especially like to thank my adviser, Joan Baart, for the time and energy he invested in sharing his knowledge and experience of acoustic phonetics, phonology, and Indo-Aryan languages with me. He patiently answered my questions and gave valuable advice, pointing me in the direction of useful resources and research ideas, and sprinkling our dialogues with a good sense of humor that always helped in retaining my sanity. Thanks is also due to another of my committee members, Steve Parker, who helped with various aspects of research design, the steep learning curve of statistics, and helpful tips for writing clearly and concisely as the manuscript itself was taking shape. I also would like to thank my third committee member, Steve Marlett, for his detailed and thorough comments in the final stages of the project, helping me to see numerous ways in which it could be made more clear, concise, and logical.

I would also like to express gratitude to all of my friends, family, and colleagues who have supported me in various ways throughout this project, whether it was through prayer, encouraging words, locking me in my room, or bringing me a cup of coffee. Between Holly, Anita, and I, we were able to come up with a surprising amount of solutions to the challenges we faced in learning to use Praat, analyzing waveforms, and understanding the intriguing field of acoustic phonetics. Megan, my special friend, it is with a profound sense of gratitude that I recognize your support of me as I made the completion of this project a high priority, even when it meant less time spent with you — you are deeply appreciated.

To Mom & Dad

ABSTRACT

Differences in the phonetic and phonological systems of Bangla and English result in negative transfer in the Bangla stop productions of native English speakers. The phonetic realizations of Voice and Aspiration and their interactions with each other are the key factors in this. A production study was carried out focusing on sixteen of the twenty Bangla stops that are distinguished by a four-way voice/aspiration contrast at four different places of articulation, providing a contrastive acoustic analysis of the pronunciation of L1 and L2 adult speakers. Data containing these stops in an intervocalic environment in word-initial, word-medial, and word-final positions was elicited by digital recording from twelve native Bangla speakers and twelve native English speakers. The data from the L1 speakers was analyzed to investigate production characteristics related to the following acoustic variables: vowel voicing onset time, closure duration, closure voicing, preceding vowel duration, and duration of aspiration noise. The data from the L2 speakers was then analyzed using the same variables.

The primary acoustic correlates of Voice and Aspiration in Bangla were found to be closure voicing and vowel voicing onset time, respectively, and the interaction of these two variables made a clear distinction between the four stop classes of Bangla: voiceless unaspirated, voiceless aspirated, voiced unaspirated, and voiced aspirated. Evidence was found supporting the work of various researchers who have suggested that a [breathy voice] feature is not necessary for a phonological description of the Indo-Aryan languages. The stop productions of the native English speakers indicated a conceptual awareness of the four stop classes, but it was also clear that they lacked a native-like control of the Voice and Aspiration features and their specific interactions with each other. The degree to which the L2 productions of the four stop classes

were different from those of the L1 was directly correlated to each class's similarity to English phonological patterns, providing evidence of certain predictable aspects of L1 transfer. In order to fully apply the results of this study in a pronunciation acquisition context, perceptual studies will need to be done to identify the salience of these acoustic variables for both L1 and L2 speakers. Perceptual studies involving L1 speakers may also give a greater understanding to the ongoing discussion on the best phonological description of the four-way stop systems of the Indo-Aryan languages.

CHAPTER 1

INTRODUCTION

1.1 Background

Bangla (also known as Bengali*) is the mother tongue of most of the people living in Bangladesh. This country was a part of India until 1947, when Pakistan and India were divided into separate nations. At that time Bangladesh was known as East Pakistan and present-day Pakistan was known as West Pakistan, with a large part of India separating them geographically. In 1971 East Pakistan gained its independence and became Bangladesh, 'the land of Bangla,' one of the few countries named after the language of its people. Bengalis (those who share the language and culture of Bengal, which includes Bangladesh and the Indian state of West Bengal) are proud of their language, and are especially proud that in 1999 UNESCO designated February 21 as International Mother Language Day. On this date in 1952 several Bengalis died in an effort to preserve and defend their mother tongue. This day has been commemorated annually in Bangladesh since that time, and was made a national holiday when they gained their independence from Pakistan in 1971 (Islam 2003; Rafique 2003).

As a result of increasing globalization, there continues to be a variety of foreigners coming to live and work in Bangladesh. While some Bengalis in Bangladesh speak English well, it is useful for expatriates to have a functional knowledge of Bangla. For this reason, many invest at least some time in language learning, which has the result of bringing their native language

*ISO 639-3: ben.

into contact with Bangla. The focus of this study is on one aspect of this, that of native English speakers learning Bangla.

Because the sound systems of these two languages are quite different, many of these learners find the pronunciation of some Bangla consonants difficult. While this difficulty relates specifically to the field of second language acquisition, the fields of phonetics and phonology can also be looked to for guidance in overcoming these challenges. This study examines the pronunciation of sixteen of the twenty Bangla stop consonants, first considering some of the relevant literature, and then providing an acoustic description and comparison of these consonants as produced by both native Bangla speakers and native English speakers. Thus, it serves as a foundational study for further research focused on developing practical suggestions for native English speakers on how they can improve their Bangla pronunciation.

1.2 Second Language Pronunciation Acquisition

Because second language acquisition (SLA) involves the mind, learning, and social interaction, it is not surprising that theories originating in other related scientific disciplines have had a significant impact on its history. For instance, the ideas of behaviorism (the branch of psychology focusing on empirically observable stimuli and behavior) can be seen in the work of Leonard Bloomfield (see Bloomfield 1933). Taking a quite different and, to some degree, opposing approach, Noam Chomsky developed his ideas about language using a cognitive framework, thus moving the focus from the observable products of language to its source and systems of generation (see Chomsky 1968). While psychological theories may seem somewhat removed from the practical issues facing language learners and teachers, it is important to note that the theories of cognition and learning that are in vogue at any given time will likely influence the contemporary SLA theories. An awareness of this is essential both for benefiting

from research conducted in other fields, and avoiding inaccurate assumptions generated by an inadequate theory.

1.2.1 Theories and Approaches to Segment Transfer

When considering the various SLA theories, there are numerous concepts that are especially relevant to the issue of cross-linguistic influence. Because similar terms are used in various ways throughout the SLA literature, the following clarification of the terms used in this study is given. The term 'native language' (L1) is used in its usual sense to refer to the mother tongue or first language of a speaker. For ease of reference and in accordance with one of its common uses, the term 'second language' (L2) is used to refer to any non-native language that a speaker interacts with. 'Transfer' is used to refer to the influence that one language has on another in a person's use of those languages.

Transfer is a relevant and important issue in SLA, exemplified by the following types of questions, among many others: How does the word order of an L1 affect the ability to learn an L2 with a different word order? What about differences in intonation patterns, the use of tone, or completely new sounds? How do language learners deal with the differences and similarities between their L1 and an L2 they are attempting to learn? These are just a few of the questions that are important to address when developing a theoretical and practical approach to SLA. There is a large body of literature dealing in some way with the concept of transfer and using the term in a variety of ways; thus, a further clarification of how it is used here would be helpful. The following working definition put forward by Odlin (1989, 27) is sufficient for the present purpose: "Transfer is the influence resulting from similarities and differences between the target language and any other language that has been previously (and perhaps imperfectly) acquired." In light of this definition, the issue being focused on in this study could be stated as 'the influence resulting from similarities and differences between the sound systems of Bangla and

English.' Because of the ambiguity in terminology usage pointed out by Gass and Selinker (1994, 56), 'transfer' is used to refer to the 'process' of cross-linguistic influence, whereas the terms 'positive transfer' and 'negative transfer' are used to refer to two types of 'products' that can occur in that process. Thus, 'positive transfer' refers to a correct linguistic production and 'negative transfer' refers to an incorrect linguistic production, but both are the result of the 'transfer process.' The particular area of transfer this study is concerned with is called 'segment transfer,' which refers to the phonetic and phonological influence that occurs between two languages, having a direct and important impact on L2 pronunciation learning. Many of the ideas explored in relation to transfer in general also have implications at the segment level, so some of these are touched on as well.

Segment transfer has been an area of focus in linguistic research since the early 1950s, gradually increasing in complexity and thoroughness of its treatment as various theories have been formulated and then tested (Hancin-Bhatt 1994, 245). Some of the early studies focused on the phonemic systems of the two languages in question, their unit of analysis being the phoneme. Phoneme charts based on the same framework of articulatory features were presented and compared, and it was hypothesized that the closeness or similarity of the phonemes of one language with the other could be used to predict the type of segment transfer phenomena that would occur (Weinreich 1974; Lado 1957; Stockwell and Bowen 1965). This approach is known as 'contrastive analysis' and was influenced by the structural linguistics popular at that time, illustrated by Lado's comment that "the clear understanding of phonemic differences is the contribution of modern structural linguistics" (Lado 1957, 10). It is commonly thought that behaviorism also had a major influence on this approach, and that may seem to be the case for some proponents of the contrastive analysis hypothesis because of their treatment of language as a set of learned habits (Gass and Selinker 1994, 59–61). However, Odlin (2003, 439) thinks that

this is simply a common misconception, at least in the case of Weinreich (1974) and the majority of researchers Weinreich cites. While contrastive analysis has some predictive value, it also allows for a certain amount of ambiguity in instances where there are two or more L1 phonemes that are somewhat similar to a particular L2 phoneme, and thus could both be potential candidates for substitution (Hancin-Bhatt 1994, 245). For instance, for a native speaker of English encountering Bangla's voiced aspirated bilabial stop $b^{\text{ɦ}}$, contrastive analysis might predict that this phoneme would tend to be replaced by English b or the aspirated allophone of English *p*, since each of these is different from the target phoneme by only one articulatory feature on the typical consonant chart. By the early 1970s this approach had lost its appeal for several reasons (De Bot, Lowie, and Verspoor 2005, 34). First, its predictive value, and thus its usefulness in developing language-learning programs, did not live up to its expectation. Second, and more importantly, the theories of behaviorism and structuralism on which it was actually or supposedly based were considered inadequate. A third, and somewhat related reason, was the growing popularity of Chomsky's proposal that L2 acquisition was similar to L1 acquisition. If this is the case, and if the implication is true that the L2 can be learned without giving direct attention to its structural forms, then the pedagogical importance of contrasting the L1 and L2 is greatly diminished.

In an attempt to provide a better way to understand issues of transfer, Corder (1973) presented an approach called 'error analysis.' This shifts the focus from the segments themselves to the segmental errors L2 speakers produce, using the evidence of negative transfer to try to understand more clearly the characteristics of the two languages in question. This is done by identifying the errors and then analyzing them. Researchers using this approach found that some errors could be traced back to the L1, but that there were also some that could not. They introduced an additional distinction between 'negative transfer' (i.e., errors) and 'positive

transfer,' which refers to elements in the L1 that are similar to the L2 and thus make learning easier (De Bot, Lowie, and Verspoor 2005, 34–35).

The term 'interlanguage' was introduced by Selinker (1972) as a theoretical construct describing the distinct linguistic system that results when someone attempts to learn and use an L2. He asserts that this is an observable entity in itself, which when compared to both the L1 and L2 systems can provide valuable information about the nature of language learning and psycholinguistics. Thus, this approach comes out of a cognitive rather than a behaviorist view of learning, and fits quite well with error analysis, since any errors made by an L2 speaker are considered to be a part of the interlanguage. Although these concepts seem fairly straightforward, the reality is that the analysis of language learning errors can be quite complex, and the fact that the speaker's intention is not always clear can result in uncertainty regarding the type of error that was made. Yet this approach can be useful in obtaining a greater understanding of SLA in general (Crystal 1997).

Following up on his 1953 work (Weinreich 1974), Weinreich (1957) addressed the ambiguity inherent in contrastive analysis by shifting its focus from the phoneme itself to the phonological features of the phoneme. While this certainly deals with some of the weaknesses of the earlier approach, there is still some room for ambiguity in cases where a phoneme of one language differs from two phonemes of another language by only one phonological feature. An example of this would be English t [–distributed, +anterior], which differs from both Bangla dental t̪ [+distributed, +anterior] and Bangla retroflex ʈ [–distributed, –anterior].

A next step came with the exploration of the ideas of feature prominence and related hierarchical structures, which allowed for a differentiation between how two different languages treat the same phonological feature. The degree of prominence assigned to a feature is directly related to its 'functional load,' which was understood to include the number of phonemic

distinctions that it made. The assumption was that there is the potential for phonological interference to the degree that two languages differ in the amount of prominence they give to a particular feature. In his examination of the sound system of Russian, Halle (1959, 34) presented a theory of phonology influenced by the generative grammar of that time, making the comment that "the hierarchy of features seems to provide an explanation for the intuition that not all features are equally central to a given phonological system." Ritchie (1968), as well as Michaels (1973), went on from this proposal to look at the ways in which different languages vary in the way that they rank features in hierarchies of prominence. A suggestion made by Ritchie (1968, 191) as well as Trubetzkoy (1969, 52) was that these hierarchies of prominence not only influence a person's production, but also their perception.

As with any area of SLA, the study of transfer has its share of difficulties. Odlin (1989, 25–43) details some of these, organizing them into the categories of "definition, comparison, prediction, and generalization." Problems of definition are evidenced by common misconceptions about what is meant by transfer, a result of the various ways in which the term has been used. Problems of comparison and prediction refer to the already-mentioned limitations that characterize contrastive analysis of two languages and the resulting predictions. However, a key characteristic of Odlin's approach is that he is not as quick as others to abandon the use of contrastive analysis and the making of predictions. Rather, he feels that these are important tools whose usefulness will increase through a continual integration with developing theories of SLA and transfer. Problems related to generalization deal with the question of which findings in a particular instance of transfer can be applied to virtually all instances of transfer. As he notes, progress in this area will most probably be closely linked to advances in answering the question of what things about language itself can be generalized.

In recent years other fields have been looked to for adaptable theories that would be more adequate in dealing with the complexity of SLA. One example of this is Larsen-Freeman's (1997) discussion of chaos/complexity science and its potential relevance to language-learning theory. In her own words, she was seeking to "call attention to the similarities among complex nonlinear systems occurring in nature and language and language acquisition" (142). In their more recent work on SLA, De Bot, Lowie, and Verspoor (2005, 14) conduct their discussion in the framework of Dynamic Systems Theory, which comes from the field of biology and "seeks to clarify systems that seem to be chaotic and self-organising." In this theory, a dynamic system is one in which every variable is interacting with the others, resulting in the whole system being changed over time. The strength of these types of approaches to SLA is that they more adequately recognize the complexity involved in issues of transfer, thus avoiding some of the apparent oversimplifications of earlier theories, such as the anticipated comprehensive predictive ability of contrastive analysis.

As is clear from this brief and by no means exhaustive overview, there has been considerable theoretical development in the area of second language pronunciation acquisition over the last sixty years. It seems that this is simply a reflection of the growing awareness among researchers that language acquisition is an incredibly complex process, with a large number of variables exerting an influence. While this may seem daunting from a scientific perspective, it is also an exciting challenge for contemporary linguists as they build on the work of those before them. We now turn to look briefly at some examples of comparisons of two languages' sound systems as well as some more specific research similar to the present study.

1.2.2 Brief Survey of Similar Studies

A comparison of the sound systems of Bangla and English was done in 1961 by Hai and Ball. Recognizing that these two languages are both in the Indo-European family, they put this in

perspective with the picturesque assertion that it is "only on the fringes of Linguistic theory that the main language of East Pakistan and English brush against each other: but the touch is as light as thistledown and neither bears the impress of the other's influence" (Hai and Ball 1961, 1). One of the things they noted is the important difference between how the two languages use aspiration. Using the common plosives p, t, and k as examples, they observed that Bangla uses aspiration phonemically whereas English does not. A second thing they noted was that Bangla has four palatal phonemes to English's corresponding two, and that these are more like plosives than affricates. Thirdly, they discovered important differences in intonation patterns and the use of stress, differences which play a key role in pronunciation (19–58).

A more recent work by Ghosh compares and contrasts the sound systems of Bangla and French. Her main purpose in doing this contrastive analysis was to describe the linguistic differences between the two languages, although she also points out the potentially ensuing problems for language learning and possible solutions (Ghosh 2003, vii). Using an articulatory phonetic classification, she compares both the vowels and the consonants of the two languages, but also includes a section on the suprasegmental features of word and sentence stress. She makes several interesting observations about the Bangla and French consonant inventories. She points out that the feature of aspiration is one of the primary reasons for the striking contrast in the number of stops in these languages. The phonetic system of aspiration, which the twenty Bangla stops use with maximal differentiation, is entirely absent in the six French stops. Because of this, it is predicted that native French speakers will find it difficult to learn to recognize and produce the differences between the unaspirates and their counterparts. As a way of dealing with this potential difficulty, she goes on to suggest that it could be helpful to explain the articulatory characteristics of phonemic aspiration, and use lists of minimal pairs both for listening comprehension and pronunciation exercises. In this useful descriptive summary of the

similarities and differences of the Bangla and French sound systems, a foundation and springboard is provided for a range of studies focused on developing effective and efficient language-learning resources.

Experimental investigation of differences in actual speech production is also useful in understanding language transfer. One example of this is a study that focuses on the pronunciation of English word-final *t* and *d* by Spanish and Mandarin speakers, and whether this contrast can be successfully learned (Flege, Munro, and Skelton 1992). The authors had native English speakers listen to the productions of five groups of people: native English speakers, and Spanish and Mandarin speakers at both beginning and advanced levels of experience in using English. The task of these listeners was to identify whether they thought they were hearing a *t* or a *d*. The results of the study showed that the productions of the L2 speakers were identified incorrectly significantly more often than those of the L1 speakers, and that the degree of experience in English that L2 speakers had did not make a significant difference in this. Also, acoustic measurements of preceding vowel duration, closure voicing duration, stop closure duration (when possible to measure), and F1 (first formant) offset frequency of the L1 and L2 tokens were compared to see if any differences there could explain the disparity in accuracy of perception. It was found that important acoustic correlates that cue the perception of *t* vs. *d* in English were present in the L2 productions as well, but typically with smaller differences between the contrasting values.

An example of a cross-linguistic study that involves the productions of L1 speakers is Cho and Ladefoged's 1999 comparison of the voice onset time (VOT) values of stops in eighteen different languages. Their aim was to examine the differences and similarities related to place of articulation, and thus to gain a better understanding of which characteristics of VOT might be universal, and which seem to be more language-specific or context-specific. They found that

most of the variation in VOT between the languages and within each language could be explained by "universally applicable phonetic rules," but they recognized that there was some variation that could only be dealt with in the given language's grammar. Besides its contribution to the literature dealing with VOT, this work by Cho and Ladefoged is also a good example of how a study can build upon previous research, in that the data they used came out of previous acoustic analyses of each of these different languages.

The present study contains similar elements to those found in the works mentioned above. The difference in the use of phonemic contrast of aspiration between Bangla and English noted by Hai and Ball is examined, especially those aspects of it that could potentially cause pronunciation problems for a language learner. A contrastive analysis similar to Ghosh's is done, but only for eight of the ten pairs of Bangla consonants distinguished by aspiration; however, a prediction of potential transfer issues and their solutions has not been attempted. Similar to Flege, Munro, and Skelton's study of Spanish and Mandarin speakers of English, speech samples of both L1 and L2 speakers of Bangla are analyzed acoustically to determine L1 pronunciation tendencies and any significant L2 departures from those; however, no perceptual data is examined. Lastly, as in Cho and Ladefoged's work, VOT-like post-release vowel onset data plays an important role, but the comparisons and contrasts made involve L1 and L2 speakers of only one language. Thus, the present study lays a foundation for the development of practical L2 pronunciation acquisition strategies for native speakers of English learning Bangla.

1.3 Phonetics and Phonology

Although the fields of phonetics and phonology are vast, it is useful to look at the big picture they offer to identify particular areas that are helpful. There are three branches of phonetics: articulatory phonetics (focusing on the production of speech sounds), acoustic

phonetics (focusing on the movement of sound waves through the air), and auditory phonetics (focusing on the perception of speech sounds by a listener) (Roach 2000, 204–205). This study examines some of the speech sounds of Bangla from the viewpoint of acoustic phonetics, but related research in the other two areas of phonetics is crucial if practical solutions to the pronunciation problems of language learners are desired.

Phonology seeks to understand and describe the sound systems of languages, including the variety of contextually determined phonetic realizations of phonemes. From a language-learning perspective, deeply ingrained phonological habits can be a hindrance in acquiring the L2 phonology required for successful communication. Thus, an understanding of both the L1 and L2 phonologies is crucial when a more natural pronunciation is sought, since a comparison of these can point out potential problem areas. However, it is also important to note that the L1 phonology can sometimes be helpful in acquiring the L2 phonology, especially in cases where there are similar sounds or patterns. Bada's 2001 study of Japanese learners of English is an example of how a detailed look at phonology can provide ideas for improving L2 pronunciation acquisition.

Major (2002) provides an interesting perspective on the interaction between the L1 and L2 phonologies during language learning. He proposes the Ontogeny Phylogeny Model to describe the process of language acquisition (ontogeny) as well as the process by which languages as a whole change in various ways (phylogeny). This model is based on the results of many studies that have sought to discover what takes place during the learning of an L2 phonology; however, he purposefully abstains from framing it in a contemporary phonological theory. One of his key ideas is that the language learner has an interlanguage phonology, which is comprised of the sum of the effects of the L1, the L2, and certain universal tendencies in learning linguistic sound systems. Thus, at the time of initial exposure to the L2, the 'ideal' language learner's production

in that language would be completely influenced by the L1, but after achieving native-like proficiency, it would be completely influenced by the L2. The universals mentioned above show themselves in the acquisition process, and are manifested by the production of sounds that are not a part of either the L1 or L2 phonology, but can be recognized as an attempt at producing the L2 sound. With this model, Major also shows how stage of language learning, speech style, markedness of sounds, and similarity of L1 and L2 sounds each have their particular influence on what the ratio of the above-mentioned effects will be on the interlanguage phonology at any given point in time.

1.3.1 English and Bangla Consonants

Determining the number of distinct sounds in English is not an easy task, due to the multiple dialects that exist, and different ways of interpreting the same linguistic data. According to one textbook on the sound system of English, there are three nasals, six plosives, two affricates, nine fricatives, three central approximants, and one lateral approximant, making a total of twenty-four consonants (see Table 1 below). There are also about eleven vowel sounds, along with around eight diphthongs. There is some variation in degree of aspiration for the voiceless plosives, but this is dependent on context and is not used for phonemic contrast (Roach 2000).

Bangla also has a number of dialects that vary somewhat in their consonant inventories. However, most dialects would contain the consonants present in Standard Colloquial Bangla (the focus of this study): three nasals, sixteen plosives, four affricates, three fricatives, two taps, two central approximants, and one lateral approximant (see Table 2 below). There are seven basic vowel sounds with seven correlating nasalized vowels, and at least twenty diphthongs (see Chatterji 1921; Bhattacharya 2001; and Ferguson and Chowdhury 1960). One of the biggest differences seen when comparing the consonant inventories of these two languages is their respective number of plosives and affricates. Bangla uses five places of articulation and all the

possible combinations of the features [voice] and [spread glottis] for these, whereas English only makes use of the feature [voice] and four places of articulation. Other differences to note are the lower number of both fricatives and basic vowel sounds in Bangla, as well as its greater number of diphthongs.

A consonant inventory of English is shown in Table 1 below. It is based on a chart in Hammond's (1999, 2) phonology of English and a chart in one of Ladefoged's works (2005, 118), and uses the standard IPA (International Phonetic Alphabet) symbols. As Hammond notes, this chart is an "oversimplification" in some ways, and is "intended to be representative of the most standard dialect of American English" (1999, 2). However, it is a sufficient representation for the purposes of this study. The consonant inventory of Standard Colloquial Bangla is shown in Table 2. It has been adapted from the charts of Kostić and Das (1972, 168–169) and Masica (1991, 107).

Table 1. Consonants of English

	Labial	**Labiodental**	**Dental**	**Alveolar**	**Palatal**	**Velar**	**Glottal**
Stops	p b			t d	t͡ʃ d͡ʒ	k g	
Nasals	m			n		ŋ	
Fricatives		f v	θ ð	s z	ʃ ʒ		h
Central Approximants	(w)			ɹ	j	(w)	
Lateral Approximants				l			

Note: Where there are two symbols in one box, the second symbol is voiced. Also, the stops include both plosives and affricates.

One SLA-related question that arises from a consideration of the corresponding stop inventories included in Tables 1 and 2 is what happens when these two languages interact with each other; for instance, when a native English speaker attempts to learn Bangla, or vice-versa. One way of hypothesizing what might happen is to do a comparison and contrast of the phonetics and phonologies of the two languages in order to highlight gaps or conflicts that need to be overcome in order to gain a more native-like proficiency in the target language. The next sections do this from the perspective of a native English speaker attempting to learn Bangla.

Table 2. Consonants of Standard Colloquial Bangla

	Labial	**Dental**	**Alveolar**	**Retroflex***	**Palatal**	**Velar**	**Glottal**
Unaspirated Stops	p b	t̪ d̪		ʈ ɖ	t͡ʃ d͡ʒ	k g	
Aspirated Stops	p^h $b^ɦ$	$t̪^h$ $d̪^ɦ$		$ʈ^h$ $ɖ^ɦ$	$t͡ʃ^h$ $d͡ʒ^ɦ$	k^h $g^ɦ$	
Nasals	m	n				ŋ	
Fricatives			s		ʃ		h
Taps			ɾ	ɽ			
Central Approximants	(w)				j	(w)	
Lateral Approximants			l				

Note: Where there are two symbols in one box, the second symbol is voiced. Also, the stops include both plosives and affricates, but only the results of the plosives are reported in this study. *Kostić and Das (1972) refer to these as "alveolar retroflexives" and Chatterji's (1921) description of them is comparable.

In her contrastive study of the sound systems of Bangla and French, Ghosh (2003, 3) defines contrastive analysis as "a method of systematically comparing two linguistic systems in

order to find out their similarities and differences." While it is true, as noted earlier, that contrastive analysis in its strongest form has been shown to be inadequate to account for the variety and complexity of pronunciation challenges faced by a language learner, the comparison and contrast of the L1 and L2 still seems like a good starting point for making some initial hypotheses about L2 learning. Furthermore, if done well and continually adjusted to account for new information, it can serve as a useful reference point for continued SLA studies related to the languages in focus. Since a full contrastive analysis of Bangla and English is beyond the scope of this study, a brief contrastive survey is done of the parts of their phonetic and phonological systems that are most relevant to the consonants being dealt with.

One of the first things evident in comparing the above two tables is the large difference in the number of plosives (not counting affricates) — English has six and Bangla has sixteen. However, they do share both the labial and velar places of articulation. Instead of using only an alveolar articulation of stops as English does, Bangla makes contrastive use of a dental and retroflex articulation. Another important difference is that Bangla uses the feature of aspiration extensively as a phonemic differentiator, whereas English does not. This combination of having an additional place of articulation and a heavy reliance on phonemic aspiration is what gives Bangla the potential to have so many more stops. As we can see in Table 2, it fully uses this potential.

Throughout this study the terms 'voiced' and 'voiceless' are used in their phonological sense to describe classes of stops, regardless of their actual phonetic realization. Following the order of Table 1 and Table 2, the focus first falls on the labials. As with the other voiceless stops in English, p has both an unaspirated and aspirated allophone, whereas Bangla has two phonemes p and p^h distinguished by aspiration. In the current study this phonemic distinction was primarily realized as an average (across all positions in the word) of about fifty milliseconds

difference in post-release vowel onset time. Two additional dialectal and allophonic varieties of /p^h/ are the fricative [f] and one that is more like the affricate [p͡f]. In regards to this variability, Radice (1994, 28) notes that [f] is quite common in Bangladesh and [p͡f] is more common in West Bengal (a state of India bordering Bangladesh on the west), going on to suggest to language learners that [p͡f] may be the "most 'standard' pronunciation to aim for." The voiced plosives of English are realized phonetically as "sometimes fully voiced, sometimes partly voiced and sometimes voiceless" (Roach 2000, 33). In Bangla, however, the eight voiced plosives always have vocal fold vibration during their closure, making them truly voiced in the phonetic sense (Mikuteit and Reetz 2007, 248). This is a natural outcome of the different means by which the voice distinction is made in each of the languages, which will be described in a later section.

For the coronal plosives (which includes the English alveolar stops, and the Bangla dental and retroflex stops), most of what was mentioned above for the labials is also true here. However, it should be noted that the phonetic realization of the English plosives can be affected by stress, which plays a more significant role in English pronunciation than in Bangla. Additionally, in some dialects such as Standard American English, medial *t* and *d* are regularly realized as taps, as in the words *fatty* and *caddy*. When these plosives are in word-final position such as in the words *bat* and *bad*, there is very little difference between each pair in regards to presence or absence of low-amplitude periodicity or other acoustic cues; the main, if not only, salient indicator of the feature [voice] is the length of the preceding vowel — those preceding the voiceless stops are much shorter (Roach 2000, 34–35). When considering the Bangla coronal plosives, one of the most important differences to note is the place and manner of articulation. The two corresponding English plosives are produced with the tongue tip or blade in the alveolar area, whereas Bangla's four dental plosives are produced with it in the dental area and the four retroflex plosives are produced with it in a slightly curled-back or 'retroflex' position, often in

the general alveolar area (Kostić and Das 1972). The same authors also note a difference in the nature of the aspiration of the voiceless aspirated plosives, commenting (in regards to Bangla $\underset{..}{t}^h$) that the aspiration is "more friction-like in the first part of the aspirative period" (Kostić and Das 1972, 79). The combination of the dental/retroflex distinction and the voice/aspiration four-way contrast results in Bangla having eight coronal plosives (not counting affricates), while English has only two.

Lastly, the velars also show the same general pattern in regards to voice and aspiration as the other stops in their respective languages. In their fairly detailed description of the Bangla phonemes, Kostić and Das (1972, 63–64) observe that "Bengali k and English k are approximately the same in terms of articulation and acoustic structure but their phonological behaviour is different." They do not elaborate on what this difference is, but it is presumably related to the difference noted in English unaspirated k occurring only in a limited number of environments. The same authors make a further important point about aspiration in Bangla: "All Bengali aspirated plosives may vary in the degree of aspiration. When they are in an unstressed and final position, the aspiration may be distinctly reduced, but they still differ from their unaspirated opposites" (Kostić and Das, 92).

A variety of consonant clusters are possible in Bangla, with position in the word heavily influencing their frequency of occurrence. According to Ferguson and Chowdhury (1960, 45), "medial consonant clusters are frequent, initial consonant clusters are rare, and final clusters are almost nonexistent" in Bangla. The observation they make that is most relevant to this discussion is that "the distribution of aspirated stops is fully parallel to that of unaspirated stops, whether singly or in clusters, e.g. *t-*, *th-*; *st-*, *sth-*." This is important because it means that Bangla's phonemic contrast of aspiration is present in two environments corresponding to two distinct

environments in English, one in which a stop's aspirated allophone appears, and the other in which its unaspirated allophone appears.

As can be seen from this brief look at the stop inventories of English and Bangla, there are important differences in their respective acoustic characteristics, but there are also some potentially useful similarities. Based on these observations, the following hypotheses about segment transfer, both positive and negative, can be made in regards to a native English speaker using Bangla as an L2:

1. Aspiration is contrastive in Bangla, but contextually conditioned in English. Thus, it is likely that there will be a tendency for the voiceless unaspirated stops to be perceived as voiced unaspirated stops and produced as voiceless aspirated stops when in word-initial and word-final positions and not following *s* or *ʃ* (i.e., not resulting in a cluster providing an environment similar to that in which the English unaspirated allophone is used); however, when they are in word-medial position or follow an *s* or *ʃ*, positive transfer may be evident. Likewise, there should be positive transfer for the voiceless aspirated stops in word-initial and word-final positions, but there may be a little difficulty with them in word-medial position or when they follow an *s* or *ʃ*.
2. Voiced stops are consistently realized with a significant amount of glottal buzz during the closure phase in Bangla, but not in English. Because both the English voiced stops and the Bangla voiceless unaspirated stops *(p, t̪ , ʈ , k)* normally have their glottal buzz beginning right after the burst release, the Bangla stops may be perceived as English *b, d, g,* which in turn may result in production difficulties.
3. Aspiration is used to contrast the four pairs of voiceless plosives in Bangla, a contrast not found in English. However, the existence of this contrast can normally be understood to some degree by native speakers of English, resulting in their ability to produce these phonemes with at least some differentiation. When this is the case, there may be a

tendency towards a more English-like realization of aspiration for each of the phonemes in casual speech, resulting in a post-release vowel voicing onset for the unaspirates that is longer and one for the aspirates that is shorter than that of a native Bangla speaker. In the careful speech of an L2 user, however, there may be an exaggeration of aspiration (and thus post-release vowel voicing onset) for the aspirates, but it is unlikely that this speech style would result in the unaspirates having a much shorter post-release vowel voicing onset, since the resulting production would seem like a voiced sound to the L2 user.

4. As the result of the contrasts of voice and aspiration at the dental and retroflex places of articulation, Bangla has eight coronal plosives whereas English has only two. Thus, it is likely that the greatest concentration of negative phonetic transfer will occur with these eight phonemes.
5. The Bangla voiced aspirated stops do not have intuitively recognizable corollary phonetic realizations in English, as do the other three stop classes. Thus, there is likely to be more interspeaker variability in their production, especially in the earlier stages of SLA.

1.3.2 English and Bangla Phonological Processes

There are also important differences in the phonological processes of English and Bangla. At the segmental level, aspiration of the English stops is contextually determined. For instance, *t* in the word *top* is aspirated, but when it follows *s* in the word *stop* it is unaspirated (Hammond 1999, 3). Thus in this allophonic variation a phonetic distinction similar to that between the corresponding Bangla stops can be seen. Context also plays a role in Bangla, evidenced by aspirated stops in word-final (hereafter 'final') position losing much if not all of the aspiration usually present in their post-release phase, a process sometimes referred to as 'de-aspiration.' This results in a phonetic realization more similar to their unaspirated

counterparts, an effect that sometimes occurs in word-medial (hereafter 'medial') position as well. However, it is clear from the presence of minimal pairs in word-initial (hereafter 'initial') position that the aspirated stops in Bangla retain their distinctiveness in a consonant cluster such as *st̪ʰ* (see Ferguson and Chowdhury 1960, 45–46). This is important to note as a contrast to the allophonic use of aspiration in English which results in stops being unaspirated when following *s*, a contrast that predicts an interesting interaction of both positive and negative transfer for the native speaker of English learning the pronunciation of the Bangla voiceless stops. Kostić and Das (1972) also detail some other articulatory and acoustic characteristics of the aspirated stops that could possibly serve as perceptual cues in differentiating them from the unaspirated stops in contexts where aspiration is minimal or absent.

Both English and Bangla share the common trait of using the feature [voice] to differentiate between phonemes. However, as noted above, the way this feature is realized phonetically in each of the languages creates the potential for a certain amount of perceptual confusion, most notably between the voiced stops of English and the voiceless unaspirated stops of Bangla. Differences in suprasegmental features in each of these languages also exert their influence on the [voice] feature. In final position English voiced stops usually have little voicing, and retain their phonemic contrast primarily by the acoustic cue of preceding vowel duration. The corresponding stops in Bangla also have a tendency to have less voicing than their initial and medial counterparts (cf. Chatterji 1921, 4).

Several other characteristics of these two phonological systems should be noted. While in Bangla the contrastive use of aspiration plays a crucial part, in English the contrast between stops and fricatives is more important than in Bangla (see Pandit 1964, 203). Thus, in English there is allophonic variation using aspiration, and in Bangla the fricativization of stops is somewhat common, sometimes causing them to be fully realized as fricatives. English has more

possible types of syllable structure than Bangla, at least when considering words that are considered native Bangla, as opposed to borrowings from Sanskrit, Arabic, English, or other languages. Rules for word stress are more complex in English, whereas Bangla uses relatively weaker stress in a simpler system that Hayes and Lahiri (1991, 55) summarize with the following "inviolable" rule: "Stress the initial syllable of a word." They go on to note that while the two languages display a great deal of difference in the phonetic aspects of stress, the intonational systems they use are fairly similar (Hayes and Lahiri 1991, 92–93).

Based on these phonological characteristics of each language, the following additional hypotheses about segment transfer can be made in regards to a native English speaker using Bangla as an L2:

1. It is probable that, at least in the early stages of SLA, each of the Bangla stops will be perceived as some variant of an English stop, accompanied by a tendency to use the L1 phonological rules related to it. For instance, if Bangla $b^{\text{ɦ}}$ in a given context is considered to be most like English p^h by one L2 user and most like English b by another, they may have different phonologically motivated productions of the same phoneme as the result of the different ways they are conceptualizing it. This acknowledges the role that phonetic transfer plays in phonological transfer.
2. In regards to the word-initial position, L2 speakers will find it challenging to produce and perceive the voiceless unaspirated stops, tending to aspirate them in production and perceive them as voiced stops.
3. In regards to the word-medial position, L2 speakers will find it challenging to produce the voiceless aspirated stops with sufficient aspiration/frication to distinguish them from their productions of the voiceless unaspirated stops. The one exception to this would be the labials, since Bangla p^h would likely be treated as English f.

4. In regards to the word-final position, L2 speakers will find it challenging to perceive the differences between stops at the same place of articulation because of the partial neutralization of both aspiration and voicing. Their production may also seem affected if they overgeneralize by always seeking to pronounce a stop in the same 'correct' way, not realizing the influence of context and position in the word.
5. When voiceless unaspirated stops directly follow s or $ʃ$, L2 speakers will be able to benefit from positive transfer. However, when voiceless aspirated stops are in that same phonetic environment, L2 speakers will find it challenging (even more so than for the medial position mentioned above) to produce a sufficient amount of aspiration/frication.
6. For some L2 speakers who produce alveolar taps for medial English t and d, there may be a tendency to replace the Bangla coronal plosives with a tap. This would probably depend on the degree to which they are relying on their L1 phonology, and may be offset by the tendency to overgeneralize the pronunciation of a segment across each of the positions in the word.
7. Because the voiced aspirated stops are new for the L2 speaker, and because they are different from both the English voiced and voiceless stops by only one feature, L2 speakers will likely show some interspeaker variability as to which English phoneme they will associate these with, which will also have implications in the area of phonological transfer. This will be especially true in relation to Bangla $pʰ$ and $bʱ$ since these may also be associated with English f and v, respectively.
8. For the Bangla palatals and $pʰ$, the free variation (between stop and fricative) present in the productions of native speakers will likely be confusing for L2 speakers. They will probably choose one phonetic expression for each of these phonemes, which will probably result in some interspeaker variability but less intraspeaker variability than for native speakers. The greatest challenge here would likely be in the area of perception rather than production.

9. L2 speakers may have a tendency to try to use their L1 stress patterns while speaking in Bangla. Because of how different these are from the L2 patterns, this could have a big impact on the stop productions of L2 speakers, as already noted for the coronal plosives.

While the current study will not be able to speak to each of the specific hypotheses mentioned up to this point, they provide the rationale for the general hypothesis that the sixteen Bangla plosives being analyzed (see Table 2) will be a challenge for native English speakers to learn.

1.3.3 Acoustic Variables Investigated

A comparison of the voice-onset time (VOT) of Bangla stop consonants and their English counterparts is interesting to consider. Lisker and Abramson (1964), in their comparative study of the VOT of utterance-initial stops in eleven different languages, found three common patterns in the timing of voice onset. Some stops had glottal buzz beginning before the burst ("voicing lead"), others had it beginning soon after the burst ("short lag"), and still others had it beginning a while after the burst ("long lag"). The above-mentioned study and others have included investigations of the VOT of languages such as Indian English, Hindi, Nepali, and Marathi, the last three of which have nearly identical stop inventories to that of Bangla (see Davis 1995; Davis and Beckman 1983: Kagaya and Hirose 1975; and Poon and Mateer 1985). It can be inferred from these studies that Bangla stops possess all three of the above-mentioned VOT-related categories, but that VOT by itself would not be a sufficiently distinguishing feature for their four-way phonemic contrast. From these studies, it can also be predicted that the feature [voice] in Bangla is primarily distinguished by the presence vs. absence of lead voicing (i.e., negative VOT). These characteristics stand in contrast to the VOT of English stops, for which only two of the above-mentioned categories are evident, with [voice] primarily distinguished by the contrast between short lag time and long lag time, although a short amount of lead voicing is

a possible production as well for the voiced stops. The voiced stops of Bangla, however, usually have an earlier onset of voicing, most often realized by a significant amount of vocal fold vibration during the closure phase. Figure 1 gives an example of this lead voicing in the production of an initial Bangla *g*, immediately preceded by silence.

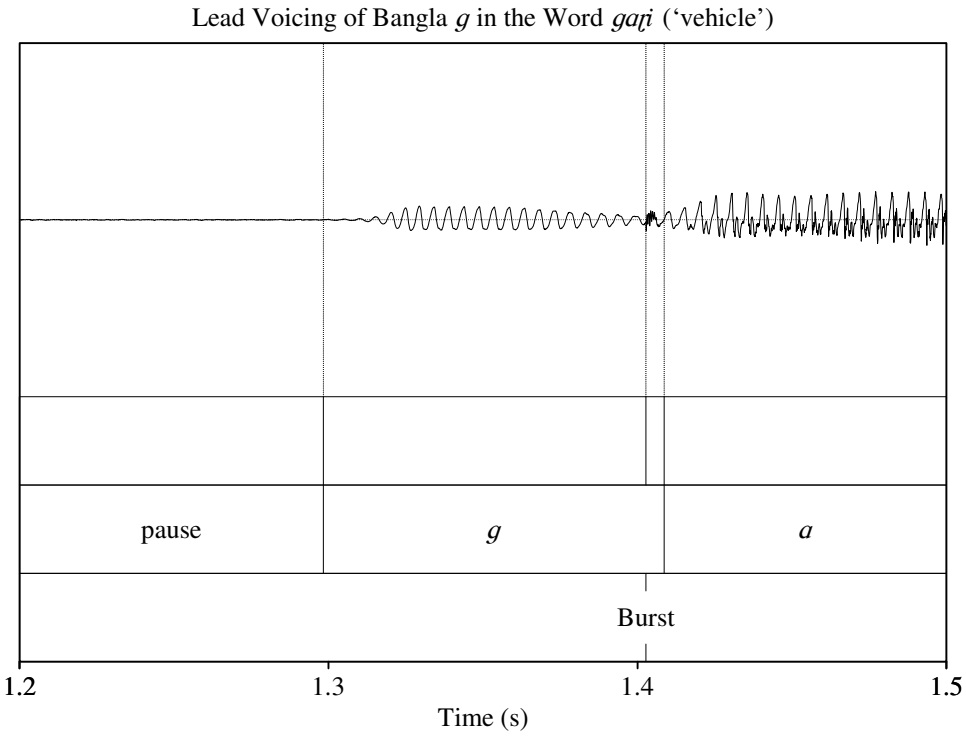


Figure 1. Example of Lead Voicing of Bangla *g* in Word-Initial Position

Note: All waveform and spectrogram images in this paper were produced with Praat (Boersma and Weenink 2008).

As touched on above, Bangla is similar to other Indo-Aryan languages like Hindi in possessing the following characteristics:

1. Phonemic voicing is associated with lead voicing — that is, voiced stops display low-amplitude periodicity during their closure phase.
2. The three-way contrast usually associated with VOT (lead time, short lag time, long lag time) is not sufficient to account for the four-way stop contrasts found in these languages.

The traditional explanation for the type of four-way stop contrast found in the Indo-Aryan languages is that it is simply the result of the various possible combinations of the voice and aspiration features. However, in the early seventies Ladefoged (1971, 1973, 1975) suggested that the combination of a three-way feature based on the timing of phonetic voicing (lead time, short lag time, long lag time) and a two-way phonation feature (normal/modal vs. breathy/murmur) provides a better way of explaining this contrast. See Figure 2 for a set of matrices (adapted from Ingemann and Yadav 1978, fig. 1–2) illustrating these two approaches.

	Unaspirated	**Aspirated**
Voiceless	*p*, etc.	p^h, etc.
Voiced	*b*, etc.	$b^{ɦ}$, etc.

(a)

	Voice Onset Time		
	Lead Time	**Short Lag**	**Long Lag**
Normal Voice	*b*, etc.	*p*, etc.	p^h, etc.
Murmur	$b^{ɦ}$, etc.		

(b)

Figure 2. Matrices of (a) Traditional and (b) Ladefoged's Approach to the Four-Way Stop Contrast of Indo-Aryan Languages

Since then, other researchers have used various means to suggest that this introduction of a [breathy voice] feature is not necessary or helpful and that the voiced aspirates can be adequately described as [+voice, +spread glottis] stops (see Ingemann and Yadav 1978; Yadav 1984; Dixit 1987; also see Dutta 2007, 5–12 for a recent summary of this discussion). As early as 1977, Catford suggested that VOT could be useful for describing the voiced aspirates as well if it was redefined as "delayed onset of *normal* voicing" (Catford 1977, 113). Davis (1994) took this approach by introducing the term "noise-offset time" (NOT), which was the duration from the burst release to the onset of F2 (second formant) in the following vowel. She showed that with this acoustic variable it is possible to distinguish between all four classes of the Hindi initial

velar stops she was investigating, claiming NOT to be a more useful measure than VOT in that context. Using a similar measure, Mikuteit and Reetz (2007) showed the same basic results for Bangla medial stops at all five places of articulation. While they found that their lag time measurement alone was not sufficient to distinguish between all four stop classes at each place of articulation, they showed that it could clearly distinguish the aspiration contrast in both the voiced and voiceless groups, thus showing that a [breathy voice] feature is not necessary to differentiate the voiced aspirates from their unaspirated counterparts.

In their study of Bangla stops and affricates, Mikuteit and Reetz (2007) called their post-release vowel onset measure 'after closure time' (ACT), which is a useful amendment of the traditional VOT concept. For instance, the VOT and ACT durations of voiceless consonants would probably be almost identical. The only difference is that ACT is not determined in relation to the low-amplitude periodicity during the closure and/or release of the voiced stops; rather, it is determined as the duration from the beginning of the burst to the onset of regular pulsing associated with the following vowel. Both NOT and ACT represent the type of measure that is useful for analyzing a four-category stop system. In this study I introduce a related but different post-release measure called 'vowel voicing onset time' (VVOT). It is similar to the other two in that it can be generally described as the duration from the burst release to the onset of the following vowel. Its distinction from NOT is that its endpoint is generally closer to the onset of the vowel's first rather than second formant, and its distinction from ACT is the result of a different interpretation of vowel onset. This difference with ACT is probably most noticeable with the voiced aspirates. A significant part of the post-release 'breathy voiced' portion is considered by Mikuteit and Reetz (2007) to be part of the following vowel. However, in this study it is considered to be part of the stop's voiced aspiration. Thus, especially for the voiced aspirated stops, VVOT is probably most similar to Davis's (1994) NOT. An example of the

VVOT measure is given in Figure 3. (See Section 2.1.3 for a description of how VVOT and the other acoustic variables in this study were measured, and see Appendix 6 for the segmentation criteria used in marking the measurement boundaries.)

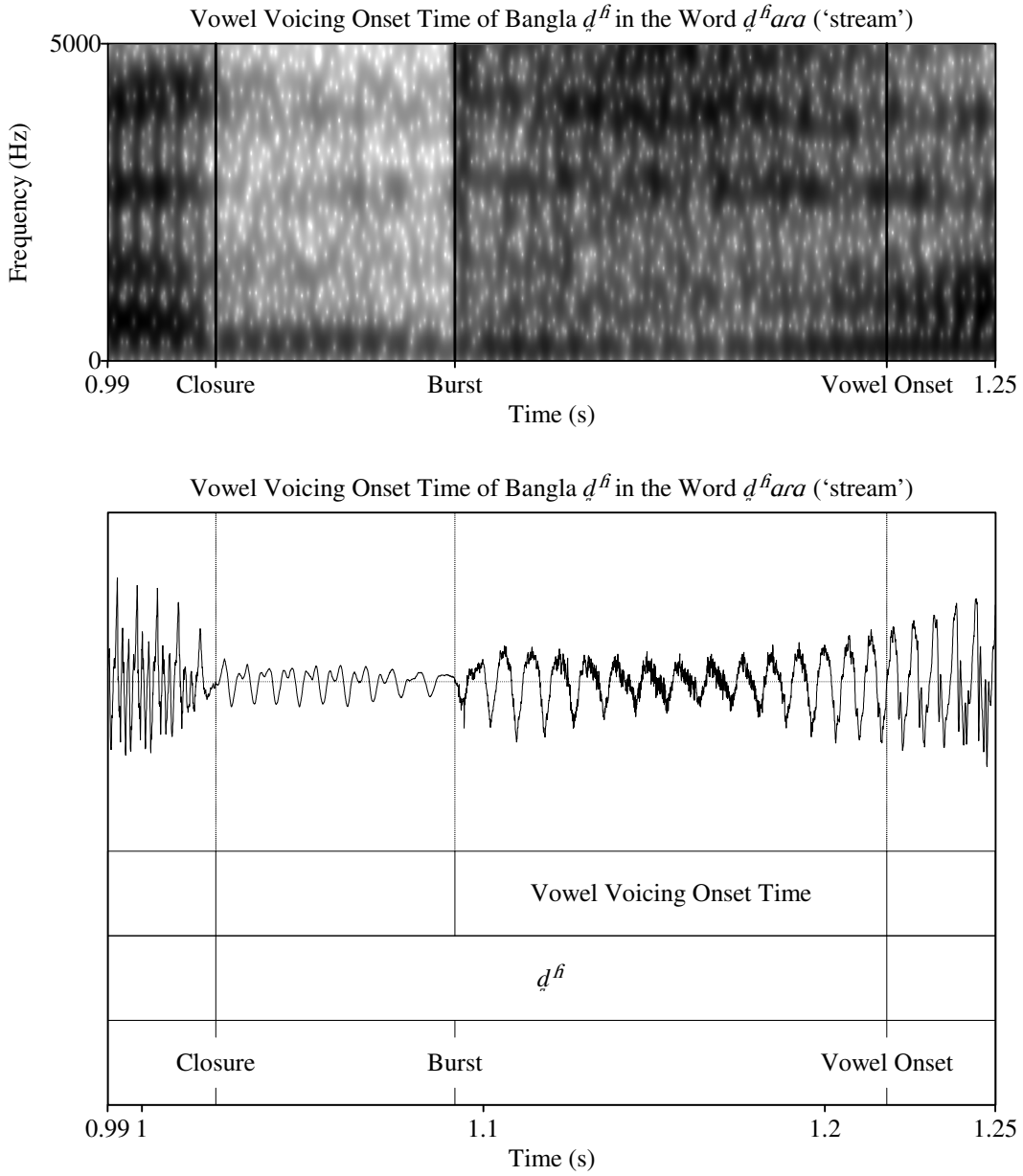


Figure 3. Example of Vowel Voicing Onset Time of Bangla ḍʱ in Word-Initial Position

Another way in which English and Bangla stops can be compared is by their closure duration. From a brief look at a number of studies measuring CD in English (among others: Byrd 1993; Sharf 1962) and in languages such as Bangla with a four-way stop contrast (among others: Mikuteit and Reetz 2007; Dutta Majumder, Dutta, and Ganguli 1973), it appears likely that Bangla stops exhibit a longer CD than the corresponding English stops.

A stop's CD as used in this study refers to the time between the oral closure after a preceding vowel and the sudden release of that closure, which allows the airstream to flow out. As Figure 4 shows, it is characterized acoustically by a relatively flat line on the waveform, but there may also be some low-amplitude periodicity present, especially in the case of voiced consonants. Lisker (1986) includes CD in a list of sixteen acoustic properties that are known to sometimes cue the voice distinction for English *p* and *b* in medial position, as in the words *rapid* and *rabid*. Closure duration can also sometimes serve as a contrastive feature of aspiration, as it does in Hindi (see Dutta 2007), and thus presumably in the other Indo-Aryan languages with similar stop inventories. This is an important acoustic variable to consider in an investigation of Bangla's four-category stop system, whose main contrastive features are voice and aspiration.

In their study of Bangla stops, Mikuteit and Reetz (2007) found that the CD of geminates was more than twice as long as that of singletons, that the CD of voiceless consonants was longer than for voiced, and that place of articulation had some effect on CD. They also found that the unaspirated stops had a longer CD than the aspirated, a similar result to what Dutta (2007) observed for the Hindi stops. When they compared each of the four stop classes by CD, they ranked in descending order as follows: voiceless unaspirated (135 ms), voiceless aspirated (116 ms), voiced unaspirated (106 ms), and voiced aspirated (96 ms) (Mikuteit and Reetz 2007, 258).

As in the four-category stop systems of other Indo-Aryan languages, the Bangla voiced stops are characterized by lead voicing (see Kostić and Das 1972; and Mikuteit and Reetz 2007,

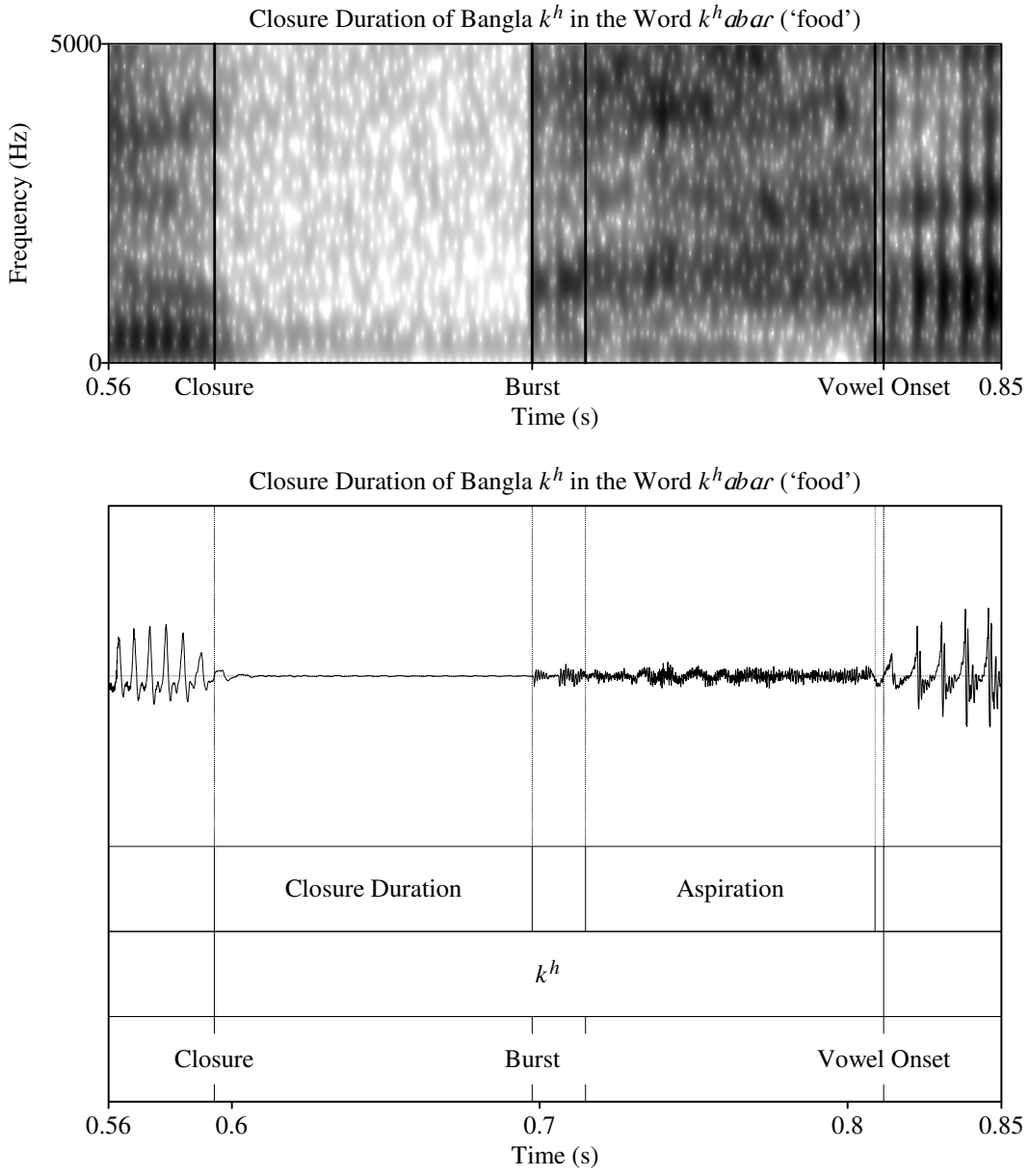


Figure 4. Example of Closure Duration of Bangla k^h in Word-Initial Position

249), evidenced by the consistent presence of low-amplitude periodicity for a large portion of their CD. This, as well as the voicing from the preceding vowel continuing for a short while into the closure phase of the consonant (sometimes referred to as a 'voice tail'), are included in the measure of closure voicing (CV) (see Figure 5 for an example). One of the characteristics of CV

is that it is directly related to the duration of the stop closure, as Dutta (2007) observed in his dissertation on Hindi stops. He asserted that "the epiphenomenon of voicing during closure serves to enhance the contrast between voiceless and voiced stops, while the duration of the closure serves to distinguish between the aspirated and unaspirated stops" (Dutta 2007, 53).

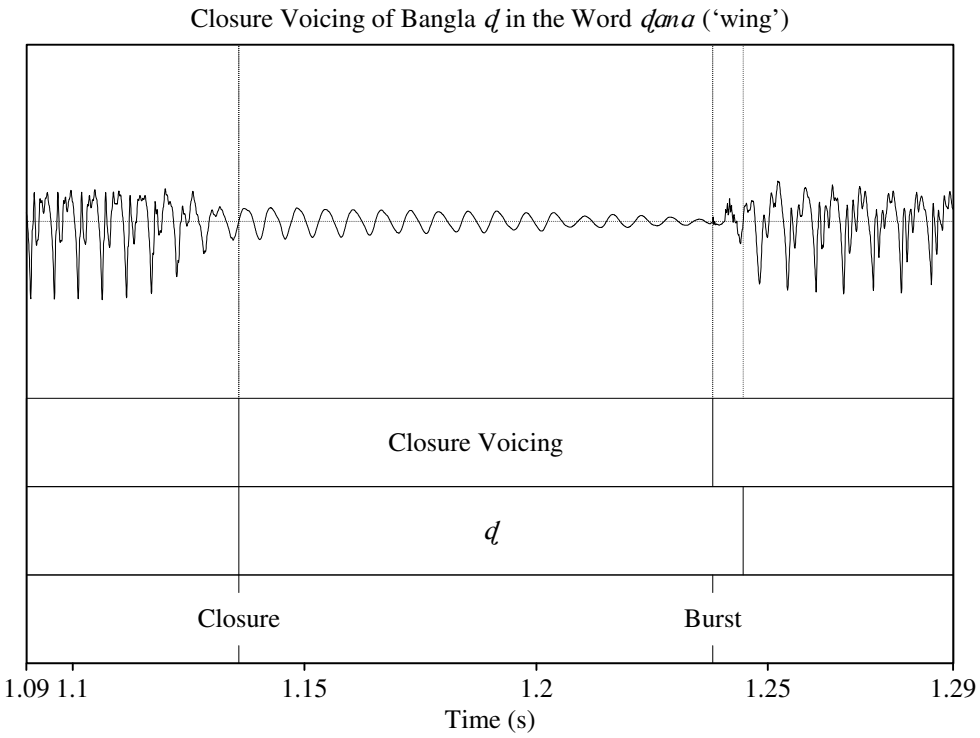


Figure 5. Example of Closure Voicing of Bangla ɖ in Word-Initial Position

It is generally the case that in most languages vowels preceding voiced consonants will normally be longer than those preceding voiceless consonants (see Chen 1970). Mikuteit and Reetz (2007) found this same pattern in their study of Bangla geminate and singleton stops in medial position, even though they did not control for vowel quality. The present study, which used more but not total control for vowel quality, examines the effect of stop voicing on preceding vowel duration (PVD), and also investigates whether there is any effect from the stop's position in the word.

Aspiration is a term commonly used in phonetic and phonological studies of a language; thus, it would be useful to know more about what it actually is. Following a detailed but useful discussion on possible definitions for aspiration, Ladefoged and Maddieson (1996, 66–70) define it as "a period after the release of a stricture and before the start of regular voicing (or the start of another segment, or the completion of an utterance) in which the vocal folds are markedly further apart than they are in modally voiced sounds" (Ladefoged and Maddieson 1996, 70). This definition serves the purposes of this study since it also includes the 'breathy voicing' of the voiced aspirates. More specifically, this study uses the term 'aspiration noise' to refer to the audible whispery noise following the release of a stop, characterized by a considerable amount of noise excitation (with or without non-vowel-related periodicity) in the signal. In Figure 6 a

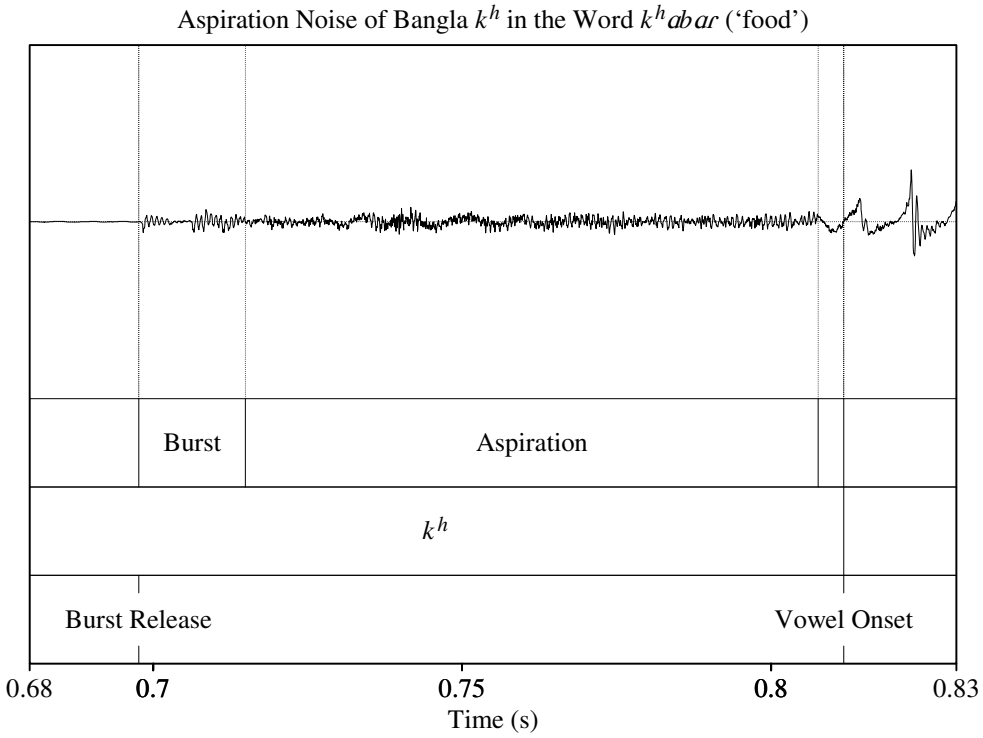


Figure 6. Example of Aspiration Noise of Bangla *kʰ* in Word-Initial Position

portion of the waveform in Figure 4 is given as an illustration of the acoustic variable of aspiration noise. As is already evident from the above discussion, aspiration plays a key role in the Bangla stop system, and the dilemma of how to most accurately explain the voiced aspirates from a phonetic and phonological perspective remains an important issue. Because of the difference in the English and Bangla stop inventories, it is also expected that the phonetic realization of aspiration in different contexts would vary considerably.

Mikuteit and Reetz (2007) note that aspiration is used for the voice distinction in English, but that in Bangla it distinguishes unaspirated from aspirated stops. They also hypothesize that the acoustic realization of the aspiration feature for voiceless stops is not the same for all languages. In addition to this, they introduce the term 'superimposed aspiration' to refer to the breathy quality sometimes observed in the first part of vowels following aspirated stops, especially the voiced aspirates. They were not aware of anyone mentioning the concept before in relation to its occurrence after voiceless aspirates, but they found the phenomenon to be quite evident in their study. Most interestingly, they observed a compensatory relationship between the durations of ACT and superimposed aspiration such that the total length of aspiration noise (including any superimposed aspiration) for both classes of aspirated stops was actually the same — it was just realized differently depending on voice. The assumption they made from this is that "aspiration as a whole is assigned as a single fixed length, irrespective of the consonant's state of voice" (Mikuteit and Reetz 2007, 266). They go on to mention a brief physiological explanation for this based on Benguerel and Bhatia's (1980) and Dixit's (1989) related studies. In this study only the aspiration noise occurring in the post-release phase of the stops will be investigated.

1.4 Research Objectives

The phonetic and phonological systems of Bangla and English are different enough that native English speakers face considerable difficulty in mastering a sufficiently adequate pronunciation of a large number of Bangla stops. In this study a contrastive acoustic analysis of the pronunciation of these stops by L1 Bangla speakers and L1 English speakers is carried out with the aim of providing useful foundational information for the successful acquisition of Bangla stop production by native English speakers.

The focus of the current research is on sixteen of the twenty Bangla stops that are distinguished by a four-way voice/aspiration contrast at four different places of articulation. Data containing these stops in word-initial, word-medial, and word-final positions was elicited by digital recording from twelve native Bangla speakers and twelve native English speakers. The data from the L1 speakers was analyzed to investigate production characteristics related to various acoustic variables, such as VVOT, CD, CV, PVD, and duration of aspiration noise. The data from the L2 speakers was then analyzed using the same variables, in order to discover and describe ways in which their production differed from that of the L1 speakers.

In the remainder of this thesis the results for the L1 stop productions are first detailed, observing general patterns in the data and significant effects of a variety of factors such as position in the word. Then the corollary L2 results are presented, showing variations from the L1 patterns. The results are then discussed in light of related literature, some subsequent characterizations of the four classes of stops are presented, and some resulting questions are raised in relation to Bangla pronunciation acquisition. This is followed by some suggestions for further research.

CHAPTER 2

BANGLA STOP PRODUCTION STUDY RESULTS

2.1 Methodology

2.1.1 Stimuli

For each of the twenty Bangla stop consonants, three real words were chosen for use in a production experiment, containing instances of the stop in word-initial, word-medial, and word-final position, respectively. In word-medial position, Bangla has a contrast between single and geminate stops. In this study, only singletons were used. Real words were used so as to ensure naturalness of the stimuli; because of this, there were some minor limitations in the type of target words available. No real words were found containing $q^{\text{ɦ}}$ in final position; consequently, this segment in that context is not represented. Two of the target words are loans from English *(soda, rod)*, but have been considered to be Bangla words for the purposes of this study, since they are often heard in everyday usage, have phonetic and syllabic characteristics compatible with integration into Bangla's sound system, and have also been included in a standard Bengali-English dictionary (Biswas 2000, 903, 1027). Additionally, there is one instance of a single target word being used for two different stops. The results of the four affricates are not included with those of the sixteen plosives in this study due to their different articulatory and acoustic characteristics. (See Appendix 1 for a list of the fifty-eight target words used.)

In order to eliminate as many variables as possible, an attempt was made to make the environments in which the target stops occurred as similar as possible, while still using real

words occurring in natural sentence contexts. Thus, no target words of more than two syllables were used, and all are nouns apart from one adjective. The words containing target stops in initial position were all disyllabic except one monosyllabic word with a CVC structure. The words with target stops in medial position were all disyllabic, and those with target stops in final position were all monosyllabic with a CVC structure. Because of Bangla's simple word stress assignment (the first syllable almost always carries word stress), the similar word structure in each of these three groups controlled for any variability caused by stress.

In order to elicit data in which the stops occurred in similar phonetic environments (even across word boundaries), the target words were embedded in sentence frames. As much as was feasible, the attempt was made to use the vowel *a* on either side of the target stop, and this was accomplished for forty-six of the fifty-eight target words. Of the remaining twelve instances, four stops were preceded by *ɔ*, four were preceded by *u*, one was preceded by *o*, one was followed by *o*, one was followed by *i*, and one was preceded by *ɔ* and followed by *e*. A frame with the meaning, "There is a/an _____ over there," was used for forty-one of the target words, and a similar frame with the meaning, "There is some _____ over there," was used for another twelve of the words. This was due to the different grammatical requirements for count and non-count nouns. Somewhat similar frames were used for the remaining five target words, the differences being due to word usage and grammatical restrictions. The resulting list of fifty-eight sentences was then checked with a native speaker of Bangla to determine grammatical accuracy, appropriate word usage, and naturalness of expression. While a few of the sentences would not be considered grammatical for written language, and some of them would only make sense in fairly specific contexts, it was affirmed that each of the sentences was a possible production of an L1 speaker. (See Appendix 2 for a list of the fifty-eight sentences.) This list of stimulus sentences was then used for data elicitation in this study.

2.1.2 Subjects

None of the subjects participating in this study were paid; however, for a few of them some minimal transportation costs were reimbursed. Twelve were native speakers of English, who had learned or were learning Bangla as a second language, and twelve were native Bangla speakers. Within each of these groups, six were male and six were female. None were perceived to have any speech disorders or a strong regional or idiolectal L1 accent, but a variety of mild accents (due to geographical origin and/or culture) were represented. Their ages ranged from twenty-two to sixty-one, with only three being over the age of forty-four. Their education level was between ten years of formal schooling and the completion of a Masters degree, and there was a similar distribution in each language group. Some were siblings or spouses of each other, and the researcher himself was one of the subjects.

Even though there is significant regional variation in the speech of Bengali Bangladeshis, the L1 Bangla speakers were each able to comfortably speak in standard colloquial Bangla, which is the form of Bangla taught in schools, and which is also used by educated people for most formal purposes. This is also the form of Bangla taught to expatriates in language schools. All of the L1 speakers reported having at least some knowledge of English, which is normal since it is one of the subjects taught in Bangladeshi schools. A few were also somewhat functional in languages similar to Bangla, such as Hindi and Urdu. All were current residents of Dhaka, Bangladesh. The majority of their first twenty-one years was spent in one or more of the following areas of Bangladesh: Kushtia Town, Faridpur District, Gopalganj District, Dhaka City, Gazipur District, and Narshingdi District.

The slight accent variation among the native English speakers was not considered to be a significant factor for purposes of this study of stops, since the primary cause of this is variability in the pronunciation of vowels. One reported that he thought his mother tongue was a dialect of

German, but that he was bilingual; the researcher also concurred that he possessed a native proficiency in English. About half reported having learned or taken classes in an L2 other than Bangla, but none of these were languages closely related to Bangla. All were current residents of Dhaka, Bangladesh, except for two who were residing in Kushtia, Bangladesh. Seven spent the majority of their first twenty-one years in the United States, two spent those years in England, one in Africa, one in South Africa, and one in both Canada and Indonesia equally. The amount of time that they had lived in Bangladesh ranged from eight months to eighteen years, although only one had lived there for more than seven years. Of this time, the amount of time given to intentional Bangla language learning ranged from six months to three years, which included time spent in formal language learning ranging from five months to one year. Five reported no former training in pronunciation, two reported some training while taking classes for other languages, five reported one to two days training as part of a two-week language and culture acquisition course, one reported the successful completion of an undergraduate-level three-credit-hour articulatory phonetics class, and one reported some training in Bangla pronunciation in the context of Bangla singing classes. (See Appendix 3 for more detailed subject information.)

2.1.3 Procedures

The stimulus sentences were presented as black printed text on white A4-sized paper, using the (non-roman) Bangla script. The free and commonly used (in Bangladesh) SuttonyMJ font was used, at a 20-point size, and without any additional formatting. The order of the fifty-eight stimulus sentences (as given in Appendix 2) was randomized, a similar filler sentence was added at the beginning and end, and the resulting sixty sentences were organized in two columns of thirty sentences each on one page, with a vertical black line separating the columns. The order of this randomized list was then reversed to create a second list, and the use of these two lists was distributed equally among each language group and each gender. The sixty stimuli were

numbered from first to last, as they appeared on the page, with the numerals appropriate to the L1 of the subjects. (See Appendix 4 for a nearly identical example of the two lists of stimulus sentences that were used, numbered with numerals familiar to a native English speaker.)

The researcher, a native speaker of English who was also one of the subjects, performed all of the data elicitation, recording each subject individually, with all forms and verbal communication provided in the L1 of the subjects. The subject was first welcomed and then made to feel as comfortable as possible. Then, the subject was asked to sign a consent form for participation in the study, Bangla and English copies of which can be found in Appendix 5. Following this, the subject information mentioned above was obtained using a short series of verbal questions. Just before the actual elicitation, the subject was given the page containing the stimulus sentences and encouraged to become familiar with it in preparation for reading the list as part of the recording process. An opportunity was also given for any questions related to pronunciation, word meanings, spelling and grammar, which were then answered as clearly and accurately as possible. Any questions related to the specific focus of the study beyond what was included in the consent form were not answered, or were only answered after the data elicitation was complete for that subject.

Once the subject, who was seated at a table, indicated readiness to begin recording, the microphone was set up and adjusted as necessary and specific verbal instructions similar to the following were given:

- You will need to read this list of sentences three times with a normal speaking rate and in a clear manner. You will not be evaluated on your performance; the researcher is simply interested in how you normally/naturally speak.
- In order to get a clear recording, you will need to speak loudly and clearly and make sure you are close to the microphone, and speaking in direct line with it.

- You may pause for any length of time between each sentence, but please do not pause anywhere within a sentence. If you can read it more fluently if you take the time to read it over to yourself first, that is fine.
- The researcher will have the same list of sentences in front of him and will follow along as you are reading them. After you have read the list once, if there are any sentences he would like you to read again, he will say the number of each sentence, after which you will read it again. If you ever feel that you have not read a sentence according to the instructions given by the researcher, you may simply read that sentence again before going on.

Once these instructions were given, subjects were given a chance to ask any questions they had until they were satisfied that they had understood what they were expected to do. At this time they were instructed to read the first 3–5 sentences as a practice run, to give the researcher the chance to adjust recording levels and also to see whether they had understood the task. Further instructions, clarifications and modeling were provided as needed to ensure a good recording signal and data consistent with the aims of the study; these were usually related to speech rate, naturalness, pronunciation, position in relation to the microphone, and the undesired insertion of pauses. Instructions related to pronunciation took the form of reminding native Bangla speakers that the focus of the study was on standard colloquial Bangla, and reminding native English speakers of the pronunciation of a particular Bangla character when it was clear that their mispronunciation was a case of faulty text processing. The need for this type of guidance in pronunciation was relatively rare, and was not considered to have a significant impact on naturalness of speech.

Once the researcher was satisfied with these various details, recording was begun and the subject was requested to read through the list of stimuli one time, and then to re-read any

sentences necessary until it was decided that the data elicited was satisfactory. In some cases it was necessary to explain why a sentence needed to be read again in order to elicit a usable rendering. At this point, the recording was stopped, and the subject was given any necessary feedback regarding their performance of the task. This same process was used for the second, third, and any subsequent recordings as well.

Almost all of the recordings were done in a moderately quiet office in Dhaka, Bangladesh, using a professional, unidirectional, dynamic microphone plugged into the microphone jack of the Creative SoundBlaster sound card of a desktop computer. The only exception to this was some of subject S40's data being recorded in a quiet room in Millersburg, Ohio, in the United States, using the same microphone plugged into a Griffin iMic USB audio adapter plugged into a laptop computer. The microphone was either held by the subject or positioned on a microphone stand, such that it was 1–3 inches from the lips and in direct line with the airflow from the mouth. Professional microphones are generally designed with a low sensitivity or signal level that assumes the use of an amplifier, whereas computer sound cards are normally designed according to the consumer standard that is used for accessories like computer microphones that are usually plugged directly into other electronic devices. This had the advantage of providing a recording setup that enabled a good, clean signal to be obtained regardless of moderate noise in the recording environment. However, it also created the disadvantage of requiring a very close proximity to the microphone and a position in which the airflow was directly aligned with it. In order to avoid any noise resulting from air being blown into the microphone, a foam windscreen was used, and a similar microphone but with a somewhat higher sensitivity level was used when necessary. The two microphones used for the data in this study were a Sekaku PRO-14L with a frequency response of 60–15,000 Hz and sensitivity of –72±3 dB at 1,000 Hz, and a Shure SM58 with a frequency response of 50–15,000

Hz and sensitivity of –54.5 dB at 1,000 Hz. When necessary, the 'Mic Boost' option was selected in the 'Advanced' section of the microphone settings in the 'Recording Control' feature of Microsoft Windows. Some recordings were done with a standard microphone cable using a 1/4" to 1/8" adapter in order to plug it into the sound card, and some were done with a microphone cable especially made for connecting a professional microphone to a sound card or other electronic device using an 1/8" plug, in which case no adapter was necessary. No noticeable difference was found in the respective recordings.

All subjects but one were recorded using SIL's Speech Analyzer software (SIL International 2008) with a 70 Hz highpass filter applied and the signal digitized as 16-bit mono at a 22,050 Hz sampling rate. The exception was subject S24, who was recorded using the Audacity software (Mazzoni, et al. 2008) with the same settings except without the highpass filter. The Cool Edit 2000 software (Syntrillium 2000) was then used to apply a 70 Hz highpass filter to it, thus ensuring its data characteristics were as much like the others' as possible. No difference was noted in the quality of the data recorded by the different software. All of the recordings were saved and retained in the WAV audio format for all archival and analysis purposes. Cool Edit 2000 was used to extract the individual tokens from each set of data and file them according to an initial assessment of recording quality and usability for this study. The Praat software (Boersma and Weenink 2008) was later used for the final assessment and selection of the tokens used for analysis. Praat was also used for segmentation and labeling of the tokens, and for subsequent measurements.

For purposes of this study, twenty-four subjects read these fifty-eight stimuli three times which, along with the fact that one target word possessed two target stops, resulted in an initial total of 4,248 tokens. These were then reviewed and the token with the best recording quality and most natural production in each set of repetitions was chosen for analysis, as subjectively judged

by the author on the basis of factors such as: signal-to-noise ratio; no electronic interference, interference from background noise, 'clipping,' or indications of 'popping' or wind noise in the part of the signal being analyzed; clarity and naturalness of diction; no hesitations or pauses; a natural speech rate; the subject's self-monitoring at the time of data elicitation; and no mispronunciations as the result of dialectal variations from Standard Colloquial Bangla or errors in text processing. Generally, most subjects seemed to be more relaxed and better able to perform the production task successfully in the later recordings; thus, these were given preference when individual tokens were being chosen for analysis. For recordings determined not to be of sufficient quality for the study, the respective subjects were later requested to return for further data elicitation. The above-mentioned selection process resulted in a total of 1,416 tokens. Excluding the 288 affricate tokens left a total of 1,128 plosive tokens, the data set on which the results in this paper are based.

For the VOT-like post-release vowel onset measurement referred to in this study as VVOT, the duration from the beginning of the burst release to the onset of the following vowel was measured. In some word-final tokens there was a glottal stop inserted before the following vowel; these were not assigned VVOT values for purposes of this study. Closure duration was measured from the end of the preceding vowel to the beginning of the burst release. Closure voicing was measured as the total duration of any low-amplitude periodicity in the speech signal that occurred after the end of the preceding vowel, but before the beginning of the burst release (i.e., within the stop closure phase). This included the fading periodicity after the closure of voiceless stops, resulting in small CV values for most of them as well. For PVD, the time from the onset of that vowel to its offset (i.e., the closure of the stop) was measured. Aspiration noise was measured from the beginning of fricative-like energy in the lower frequencies up to the onset of the following vowel, or to the point at which this acoustic feature was no longer evident. In

cases where there was no clear burst release, the burst was considered to be synchronous with the onset of the vowel, and measurements were made accordingly. Throughout the presentation of this study, duration measurements are given in milliseconds.

In order to measure the values of the above-mentioned variables, boundaries were marked at the vowel onset of both the preceding vowel and the following vowel, the closure of the stop, the beginning and/or end of any low-amplitude periodicity during the closure period, the burst release, and the beginning and end of aspiration noise. In all instances, the spectrogram was used to provide general guidance, the waveform was used for more exact placement, and the resulting boundaries were confirmed by impressionistic listening. The segmentation criteria used to ensure consistency in marking these boundaries is given in Appendix 6, and an example of boundary marking is given in Figure 7. (Also see Figure 58 in Appendix 6 for a broader view of the segmented waveform in Figure 7, with an accompanying spectrogram.)

2.2 Results of the L1 Speakers

Five different durational measures from the L1 stop productions were analyzed: vowel voicing onset time (VVOT), closure duration (CD), closure voicing (CV), preceding vowel duration (PVD), and the duration of aspiration noise. The results of this analysis serve the dual purpose of being an acoustic characterization of the four stop classes of Bangla as well as being a standard by which the L2 results presented in the next section can be evaluated. All references to significance in this study are derived from one-way ANOVAs, using an α-level of 0.05. Basic statistical data from these various tests, including obtained f values and degrees of freedom, is given in the tables in Appendix 7. Because of the multiple references to the various stop classes in the following sections, the abbreviations '–V–A,' '–V+A,' '+V–A,' and '+V+A' are often

used for ‘voiceless unaspirated,’ ‘voiceless aspirated,’ ‘voiced unaspirated,’ and ‘voiced aspirated,’ respectively.

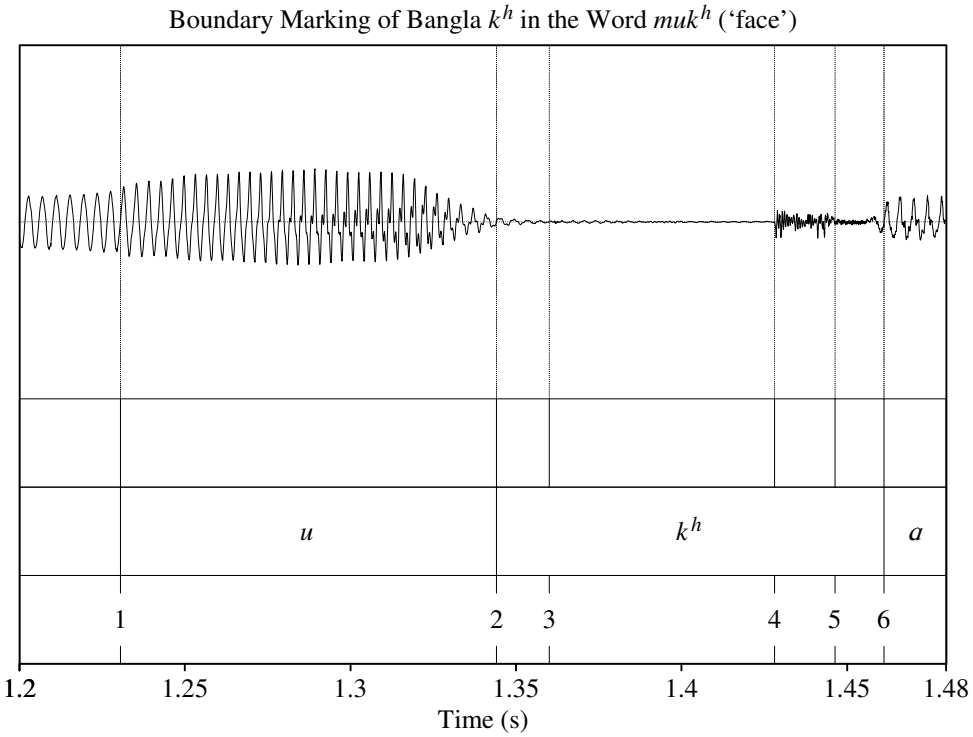


Figure 7. Example of Boundary Marking [Key: 1 = Beginning of preceding vowel. 2 = Closure and beginning of stop. 3 = End of low-amplitude periodicity during closure. 4 = Beginning of burst. 5 = Beginning of aspiration noise. 6 = Vowel onset.]

2.2.1 Vowel Voicing Onset Time for L1 Speakers

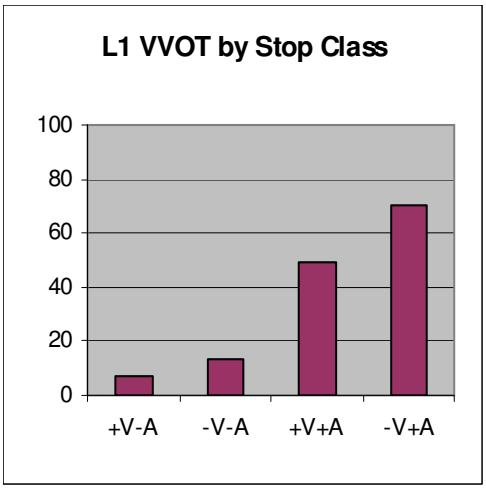
The L1 VVOT means are shown in Figure 8 by Stop Class, the effect of which was significant ($p = 0.000$). There was a significant effect of Aspiration ($p = 0.000$) as well as Voice ($p = 0.003$), with the voiceless stops having greater values on average. As Figure 9 shows, VVOT marks Aspiration more clearly than it does Voice.

The clear effect that Position in the Word had on VVOT, as illustrated in Figure 10, was found to be significant ($p = 0.000$). In initial position the effect of Voice was insignificant

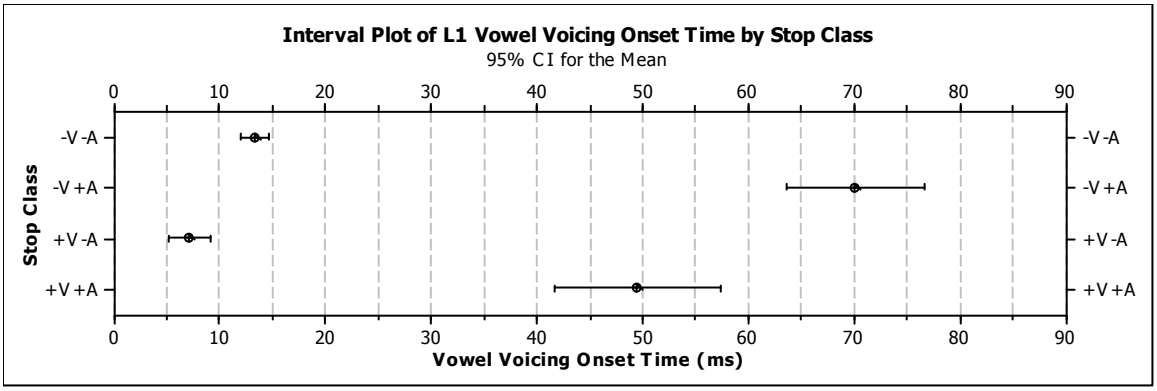
	Unaspirated	Aspirated
Voiceless	13.3 (7.6) [129]	70.1 (31.9) [93]
Voiced	7.1 (11.2) [128]	49.4 (42.8) [115]

Standard deviations in (parentheses) and sample sizes in [brackets].

(a)

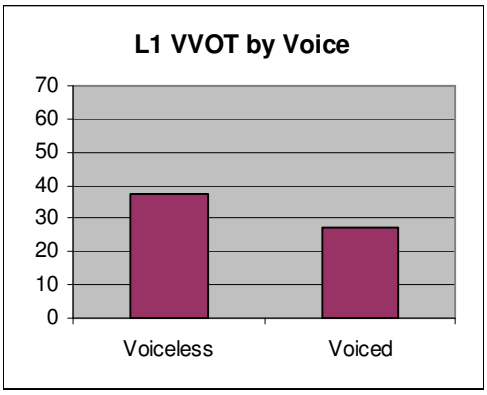


(b)

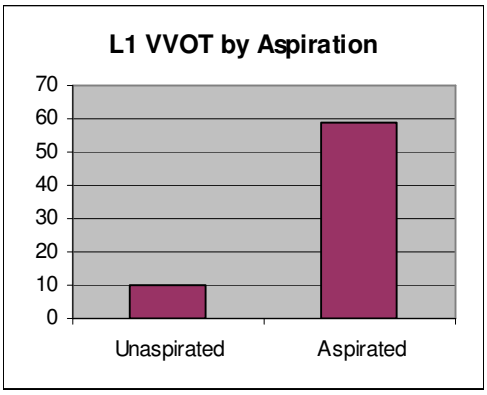


(c)

Figure 8. L1 Vowel Voicing Onset Time Means (ms) by Stop Class in: (a) Voice and Aspiration Matrix; (b) Bar Chart; (c) Interval Plot



(a)



(b)

Figure 9. Effect of (a) Voice and (b) Aspiration on L1 Vowel Voicing Onset Time Means (ms)

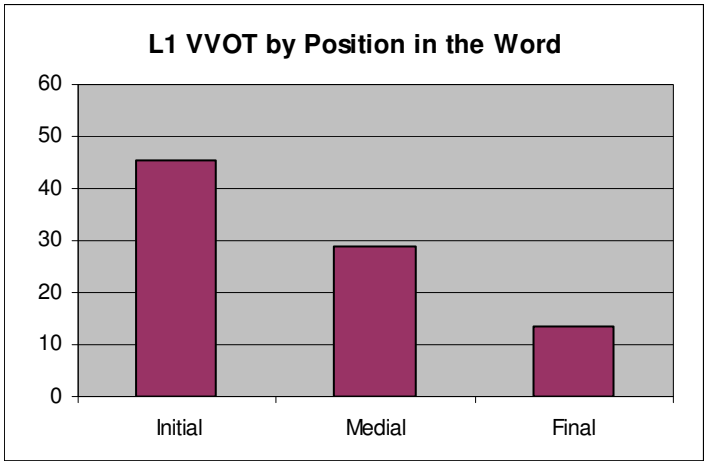


Figure 10. L1 Vowel Voicing Onset Time Means (ms) by Position in the Word

(p = 0.348), but the effect of Stop Class and Aspiration were both significant (p = 0.000). In medial position there was a significant effect of Stop Class (p = 0.000), Voice (p = 0.004), and Aspiration (p = 0.000). In final position, the effect of Stop Class (p = 0.000) and Voice (p = 0.003) were significant, but the effect of Aspiration was insignificant (p = 0.211).

The general stop class pattern in Figure 8(b) is at least somewhat evident at each position in the word as seen in Figure 11, but it is also clear that the positional effect observed in Figure 10 is mainly due to the two aspirated classes. The means of both of these groups are smaller in the medial and final position, but the difference in the +V+A values in the latter instance is noticeably greater, resulting in a mean slightly smaller than that of the two unaspirated classes.

2.2.2 Closure Duration for L1 Speakers

The CD means are shown by Stop Class in Figure 12, the effect of which was significant (p = 0.000). As the interval plot in Figure 12(c) shows, it is only the –V+A and +V–A stop classes that are not clearly distinguished from each other by CD. The effects of Voice and Aspiration on CD, evident in the bar chart in Figure 12(b) and more clearly seen in Figure 13, were both found to be significant (p = 0.000).

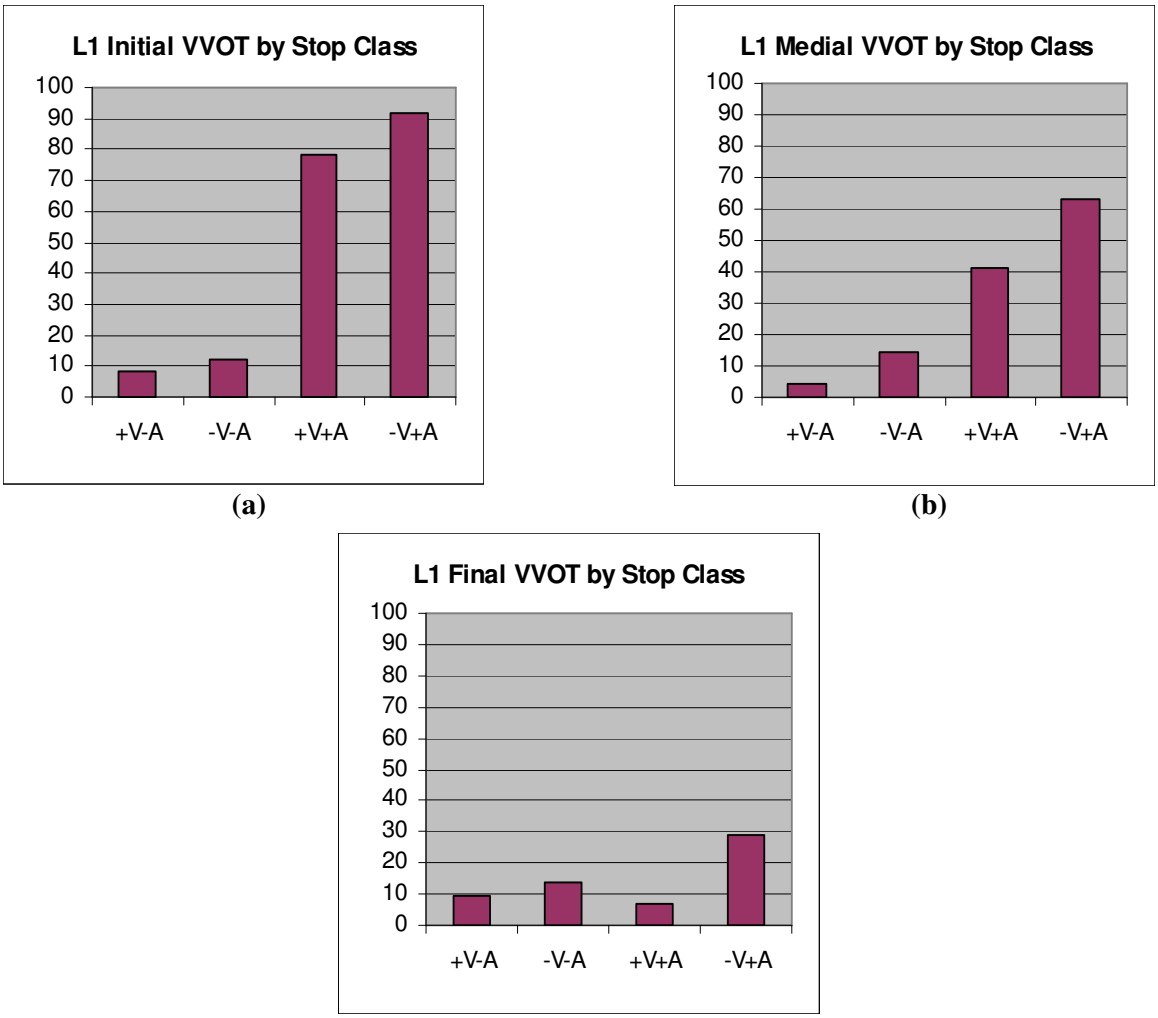


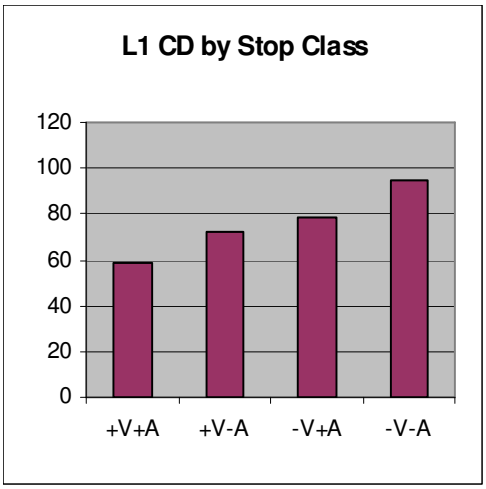
Figure 11. L1 Vowel Voicing Onset Time Means (ms) by Stop Class in: (a) Initial, (b) Medial, and (c) Final Positions in the Word

The general pattern evident is that Voice and Aspiration are each correlated with a smaller CD. The effect of Position in the Word on CD, as illustrated in Figure 14(a), was significant ($p = 0.000$). As shown in Figure 14(b)–(d), the CD means in each position in the word follow the same stop class pattern as pictured in Figure 12(b), except for the –V+A stops in initial position, which are shorter than would be expected. In spite of this, the effect of Voice is also significant ($p = 0.000$) word-initially.

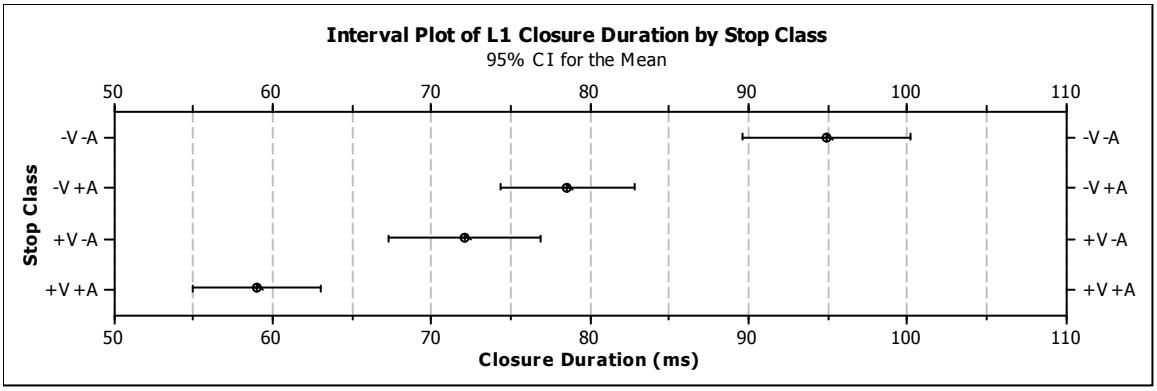
	Unaspirated	Aspirated
Voiceless	94.9 (31.4) [138]	78.5 (23.5) [118]
Voiced	72.0 (26.7) [121]	59.0 (22.6) [123]

Standard deviations in (parentheses) and sample sizes in [brackets].

(a)

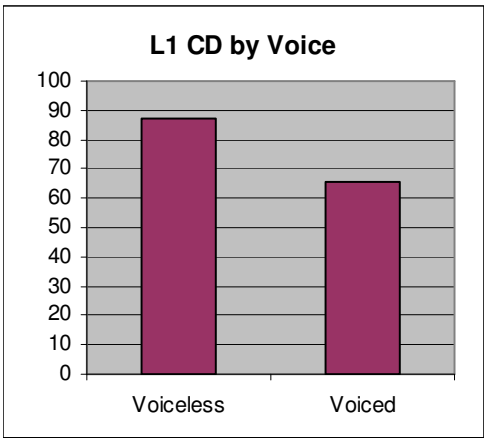


(b)

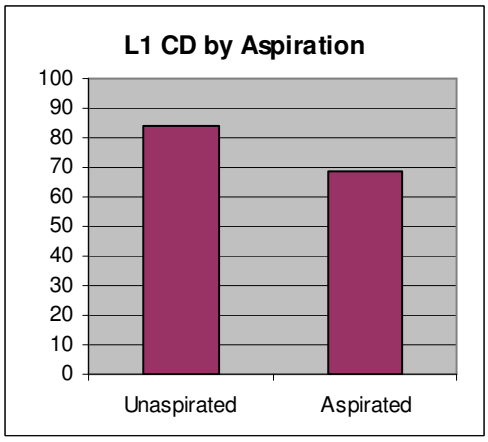


(c)

Figure 12. L1 Closure Duration Means (ms) by Stop Class in: (a) Voice and Aspiration Matrix; (b) Bar Chart; (c) Interval Plot



(a)



(b)

Figure 13. Effect of (a) Voice and (b) Aspiration on L1 Closure Duration Means (ms)

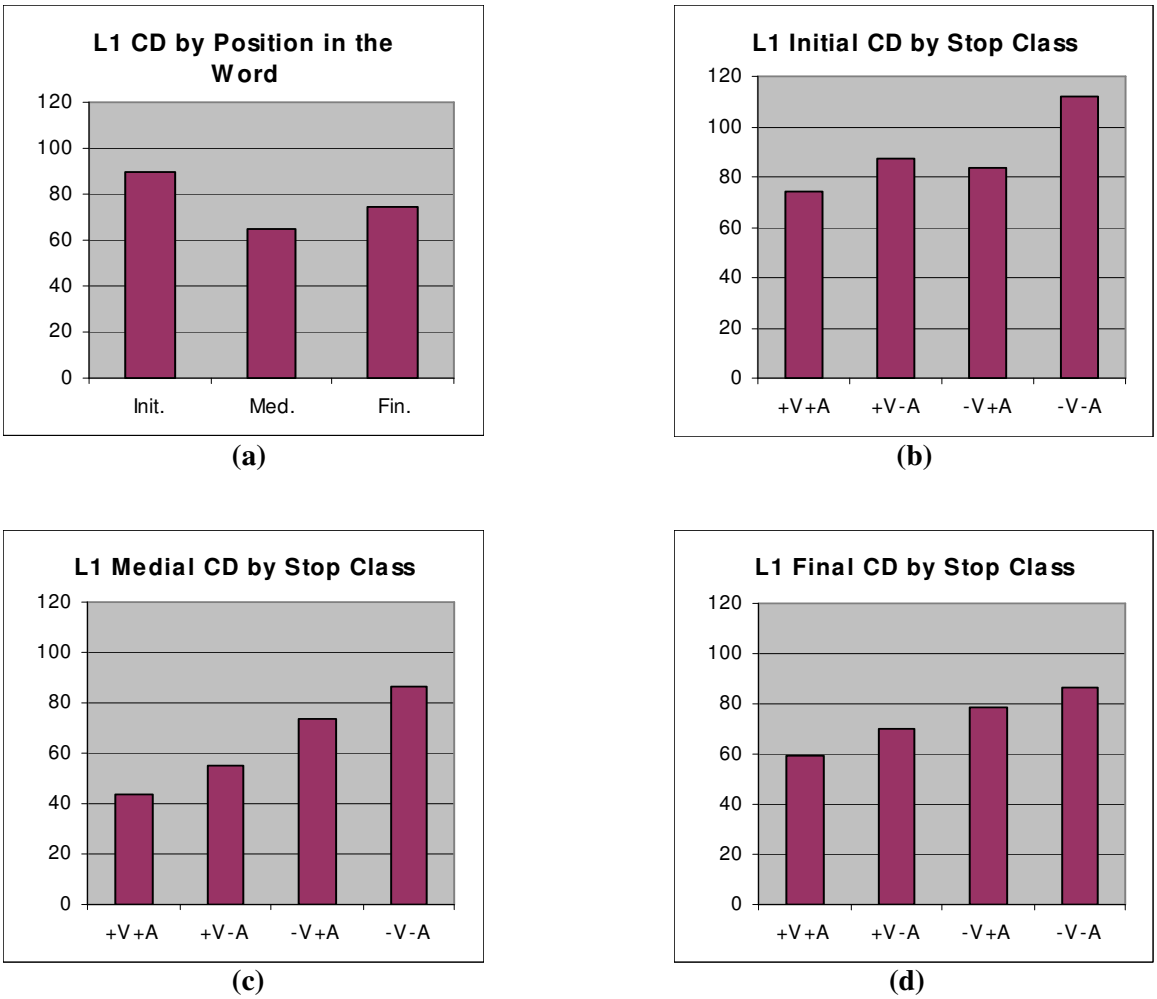


Figure 14. L1 Closure Duration Means (ms) by: (a) Position in the Word; and by Stop Class in: (b) Initial, (c) Medial, and (d) Final Positions in the Word

2.2.3 Closure Voicing for L1 Speakers

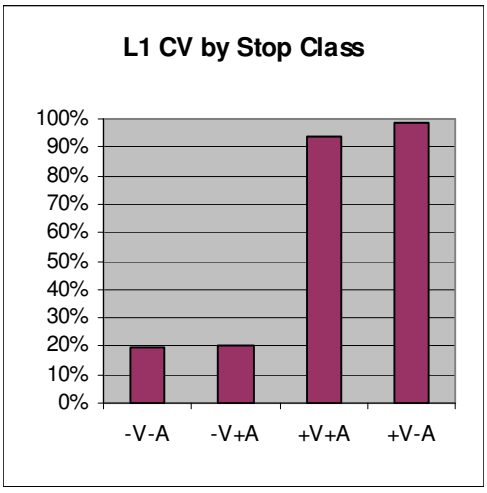
Closure voicing is an important acoustic correlate of Voice. The CV data in this study lead to the same conclusion (at least for singleton stops) as Dutta (2007) arrived at for Hindi stops — namely, that the duration of CV is directly related to CD. Thus, CV is given in relative values, calculated as the mean of each token's CV value divided by its CD value, and expressed as a percentage. The effect of Stop Class on CV was significant ($p = 0.000$), and the effect of

Aspiration was insignificant ($p = 0.591$). The main effect of Voice, as illustrated in Figure 15, was significant ($p = 0.000$).

	Unaspirated	Aspirated
Voiceless	19.4% (10.9) [138]	20.5% (9.9) [115]
Voiced	98.6% (5.4) [121]	93.6% (14.4) [123]

Standard deviations in (parentheses) and sample sizes in [brackets].

(a)



(b)

Figure 15. L1 Closure Voicing Means (as % of Closure Duration) by Stop Class in: (a) Voice and Aspiration Matrix; (b) Bar Chart

For the full set of stops there was a tendency for those in initial position to have the greatest CV values, followed by those in the medial and then the final positions, as illustrated in Figure 16(a). However, this effect was not significant ($p = 0.326$). The general stop class pattern in Figure 15(b) was evident for CV in each position in the word, as Figure 16(b)–(d) shows. There was a slight difference in the relationship between the values of the voiced stops in the final position, with the +V+A stops showing smaller values than expected. The interaction of VVOT and CV in differentiating the four stop classes is illustrated in Figure 17.

2.2.4 Preceding Vowel Duration for L1 Speakers

The PVD means by Stop Class are given in Figure 18. The general pattern in this data is that the vowels preceding voiced stops are longer than those preceding voiceless stops. As evidenced by the interval plot in Figure 18(c), the effect of Voice was significant ($p = 0.000$), but

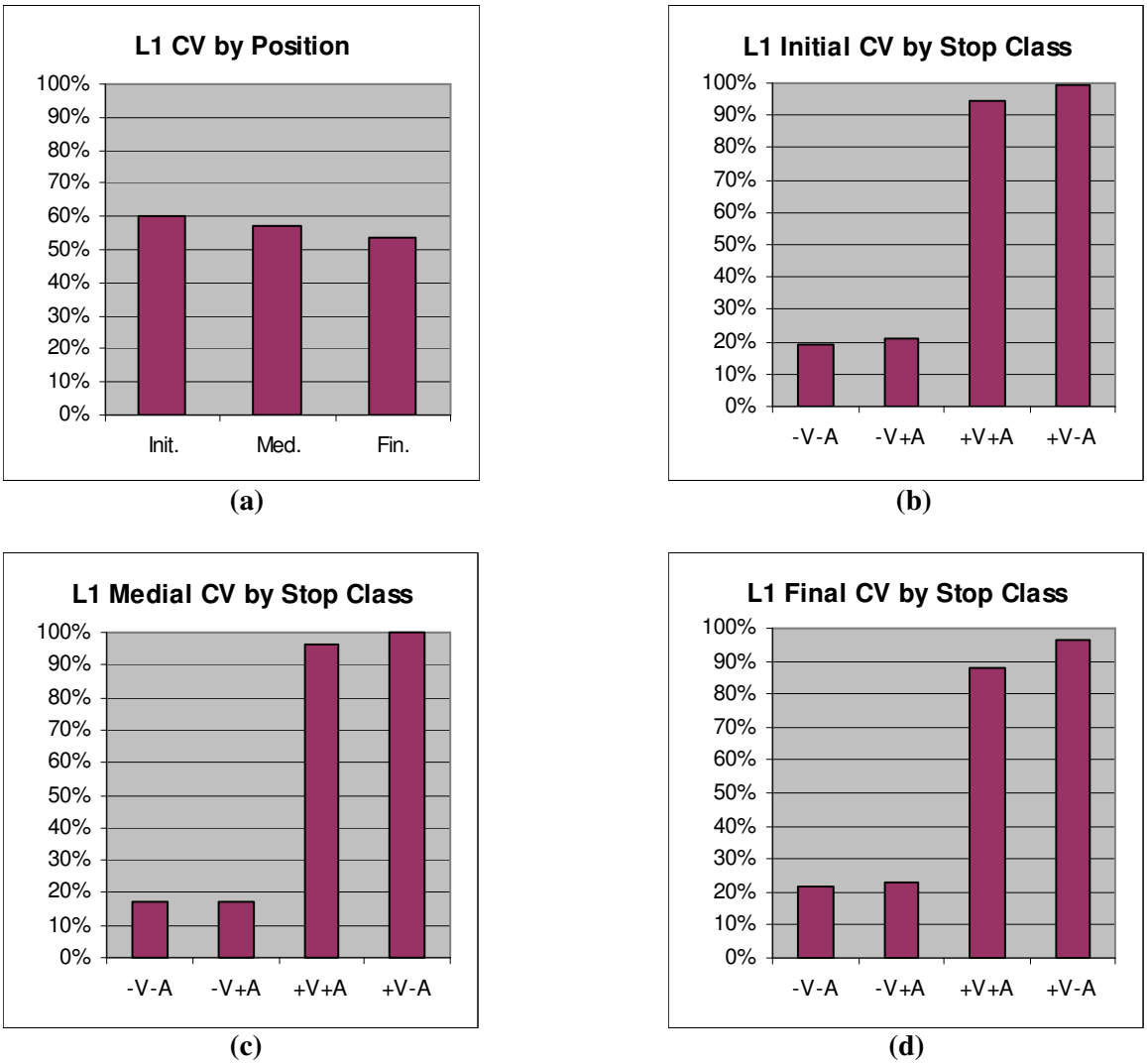


Figure 16. L1 Closure Voicing Means (as % of Closure Duration) by: (a) Position in the Word; and by Stop Class in: (b) Initial, (c) Medial, and (d) Final Positions in the Word

the effect of Aspiration was insignificant (p = 0.390), indicating that PVD is an acoustic correlate of Voice but not of Aspiration. The effect of Position in the Word on PVD, as shown in Figure 19, was significant (p = 0.000). The large differences observed are probably mainly the result of the way in which the stimulus material was constructed. For the initial stops the preceding vowel was an unstressed vowel, for the medial stops it was the stressed vowel of a two-syllable word, and for the final stops it was the stressed vowel of a monosyllabic word. Stressed vowels and

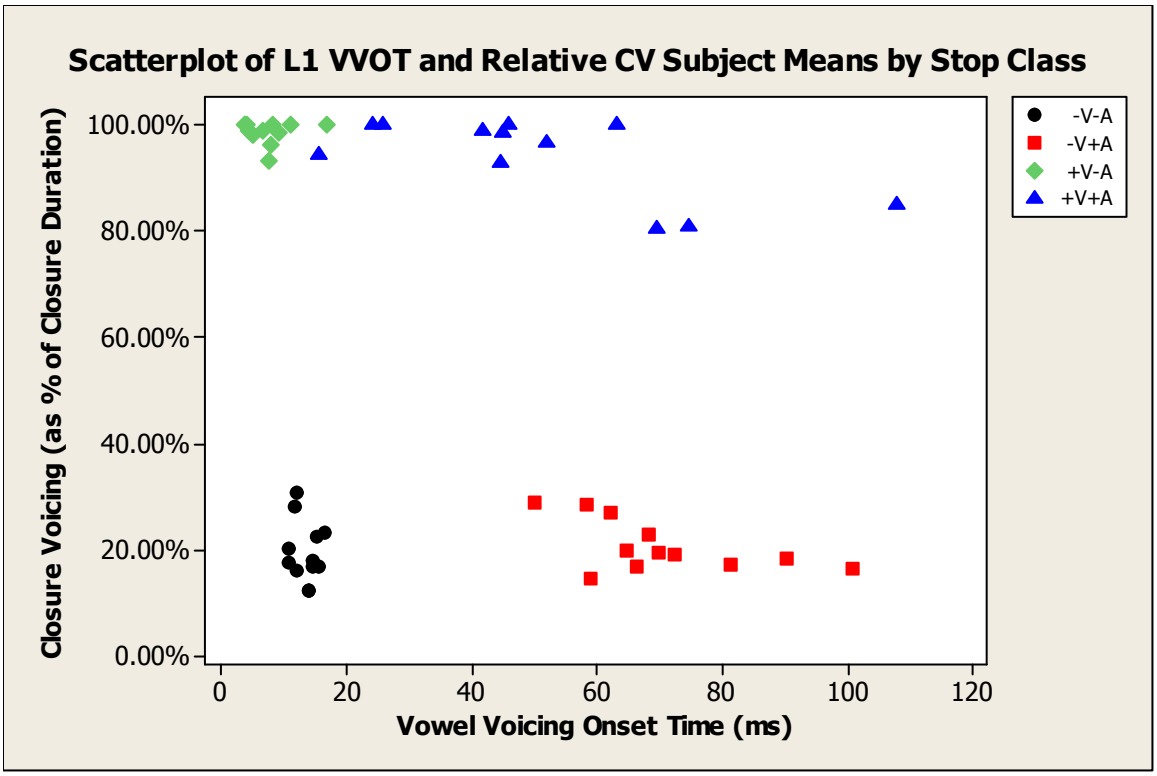


Figure 17. Interaction of L1 Vowel Voicing Onset Time Means (ms) and Closure Voicing Means (as % of Closure Duration) in Differentiating the Four Stop Classes

vowels in shorter words are generally relatively longer, resulting in the observed correlation of Position in the Word and PVD. As Figure 20 shows, the stop class pattern in Figure 18(b) is somewhat evident at each position in the word, the main exception being that the +V–A mean is lower than expected in the medial position. However, the effect of Voice is still significant ($p = 0.012$) in the medial position.

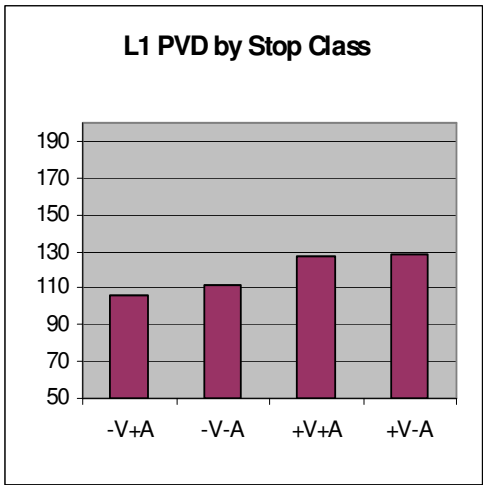
2.2.5 Aspiration Noise for L1 Speakers

There was no trace of aspiration noise in any of the L1 unaspirated stops; thus, the results in this section will deal only with the two aspirated stop classes, where it was present in 57% of the tokens. The overall comparison of the –V+A and +V+A stops is given in Figure 21.

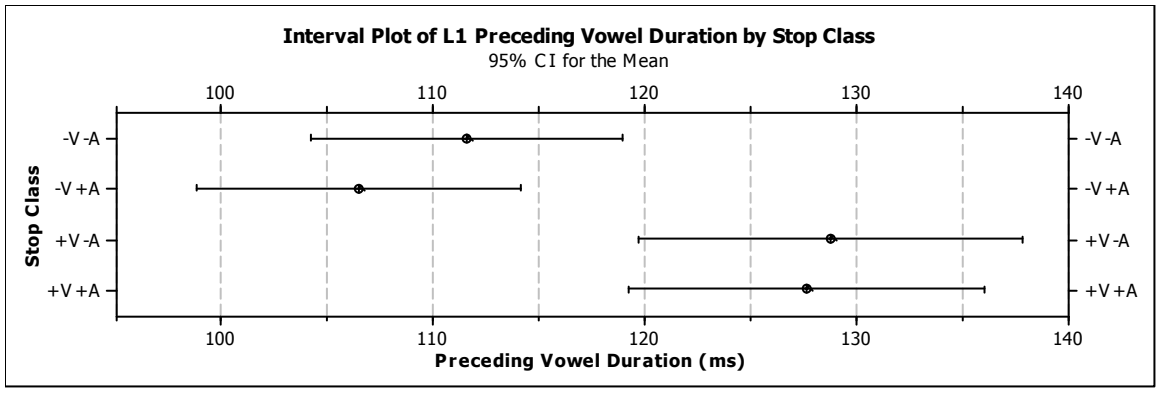
	Unaspirated	Aspirated
Voiceless	111.6 (44.8) [144]	106.4 (46.4) [144]
Voiced	128.8 (55.3) [144]	127.6 (48.8) [132]

Standard deviations in (parentheses) and sample sizes in [brackets].

(a)



(b)



(c)

Figure 18. L1 Preceding Vowel Duration Means (ms) by Stop Class in: (a) Voice and Aspiration Matrix; (b) Bar Chart; (c) Interval Plot

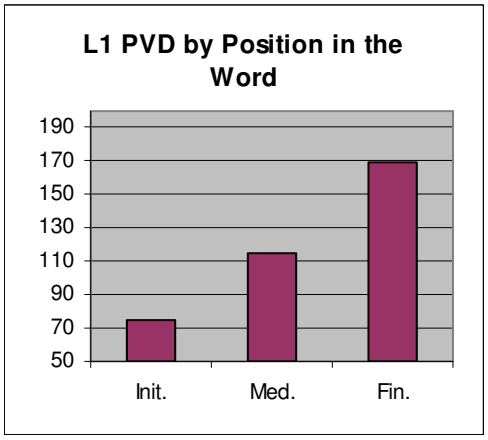


Figure 19. L1 Preceding Vowel Duration Means (ms) by Position in the Word

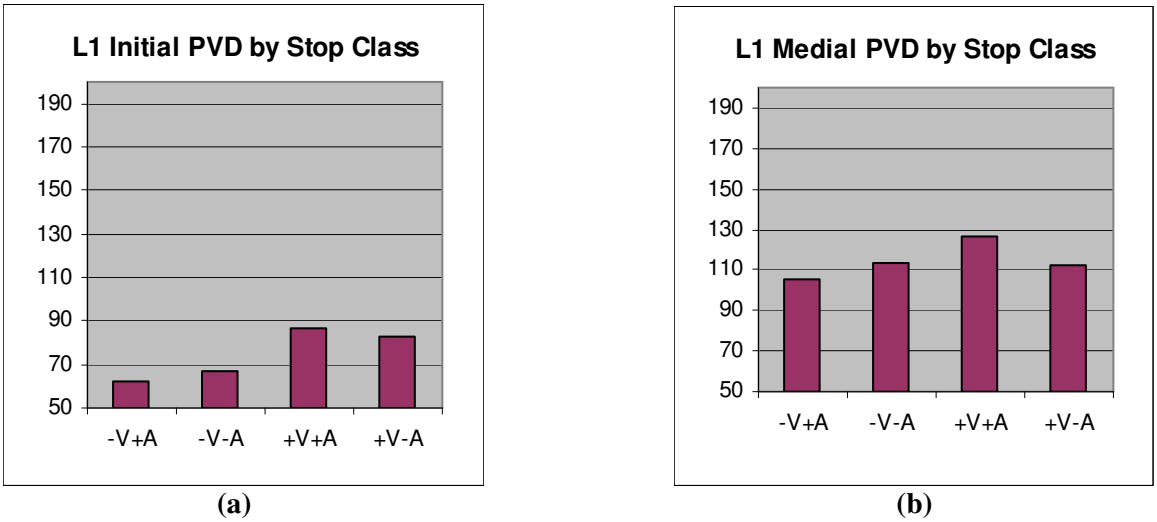


Figure 20. L1 Preceding Vowel Duration Means (ms) by Stop Class in: (a) Initial, (b) Medial, and (c) Final Positions in the Word

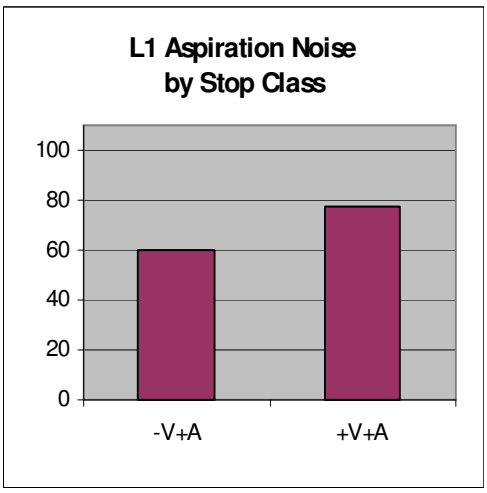
The general pattern is that the voiced aspirates tend to have longer aspiration noise durations, and this difference is significant ($p = 0.000$).

The patterning of aspiration noise duration by Position in the Word can be seen in Figure 22, the effect of which was significant ($p = 0.000$). As Figure 23 shows, the stop class pattern observed in Figure 21(b) remained consistent for each position, although the initial position showed a smaller difference and the final position a greater difference. Additionally, only the

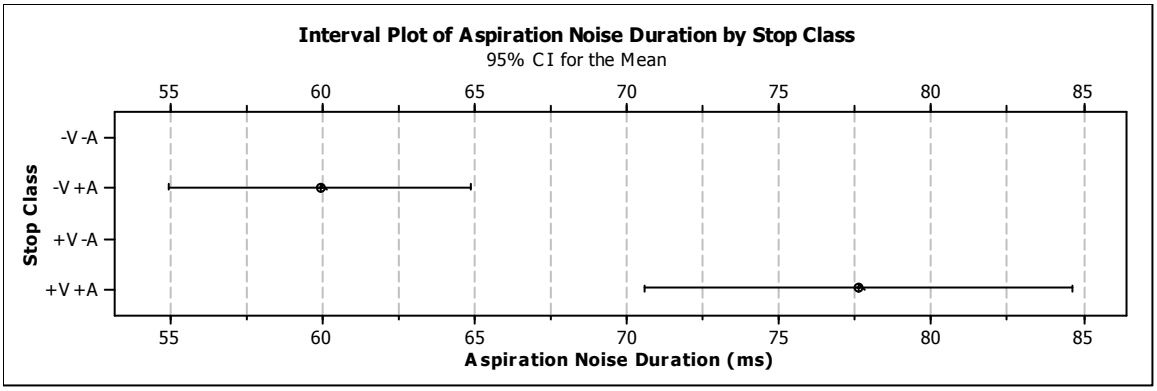
	Unaspirated	Aspirated
Voiceless	N/A	59.9 (24.3) [95]
Voiced	N/A	77.6 (27.4) [61]

Standard deviations in (parentheses) and sample sizes in [brackets].

(a)



(b)



(c)

Figure 21. L1 Aspiration Noise Duration Means (ms) by Stop Class in: (a) Voice and Aspiration Matrix; (b) Bar Chart; (c) Interval Plot

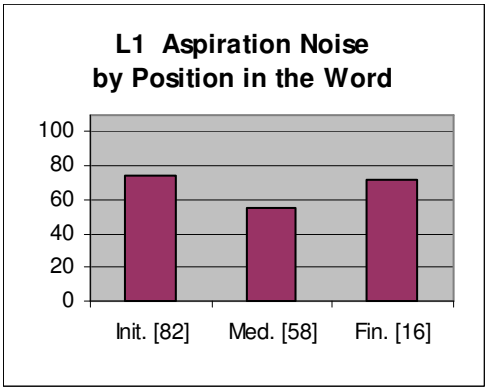


Figure 22. L1 Aspiration Noise Duration Means (ms) by Position in the Word (Sample sizes in [brackets].)

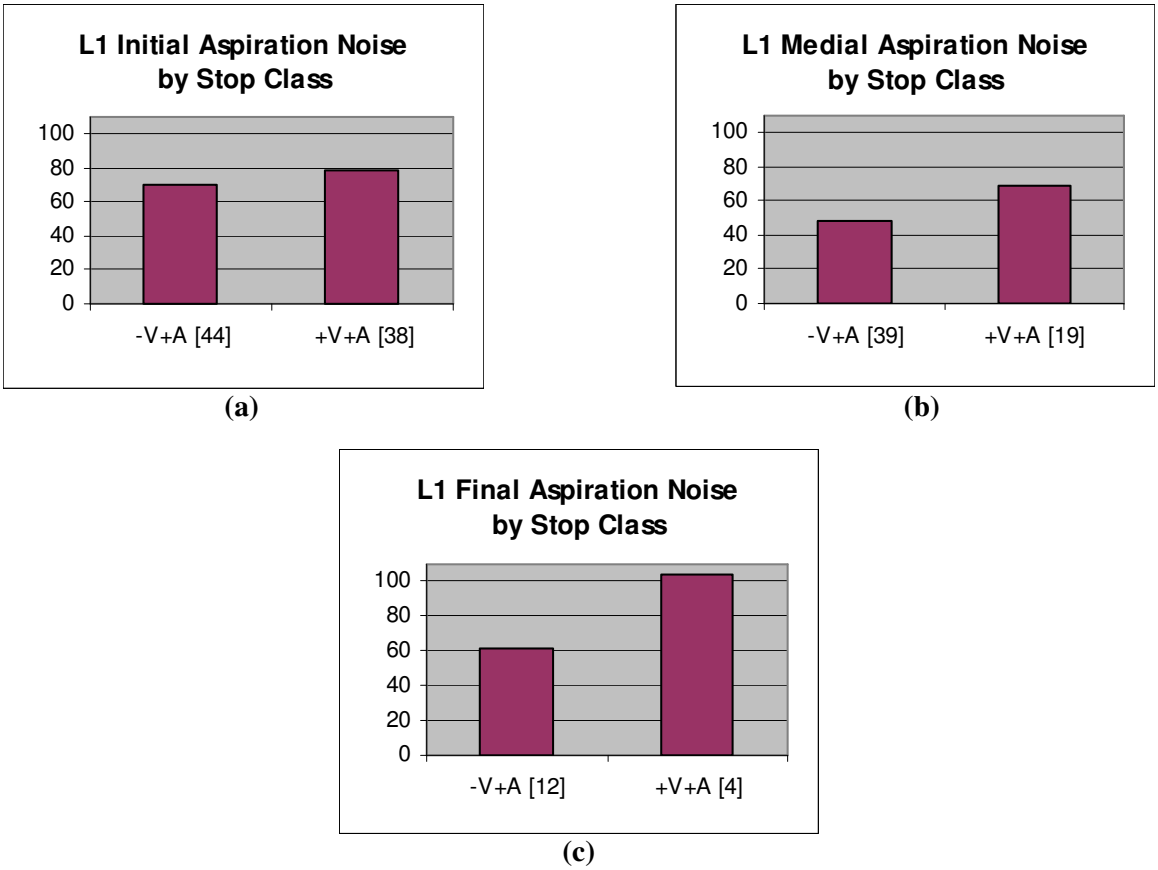


Figure 23. L1 Aspiration Noise Duration Means (ms) by Stop Class in: (a) Initial, (b) Medial, and (c) Final Positions in the Word (Sample sizes in [brackets.])

difference in medial position was significant ($p = 0.000$). The distribution of instances of aspiration noise was 53% for initial, 37% for medial, and 10% for final position.

2.3 Results of the L2 Speakers

One of the main objectives in this study was to investigate the ways in which the pronunciation of the native English speakers was different from that of the native Bangla speakers. Now that the L1 results have been detailed, the L2 results will be examined in light of them, and some of the similarities and contrasts noted.

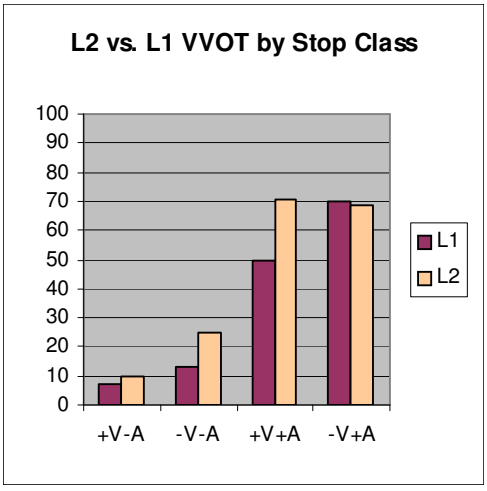
2.3.1 Vowel Voicing Onset Time for L2 Speakers

The VVOT means by Stop Class for the L1 and L2 speaker groups are presented in Figure 24. As is suggested by the bar chart in Figure 24(b), the L2 speakers had a greater average VVOT than the L1 speakers for the whole group of stops, and this difference was significant

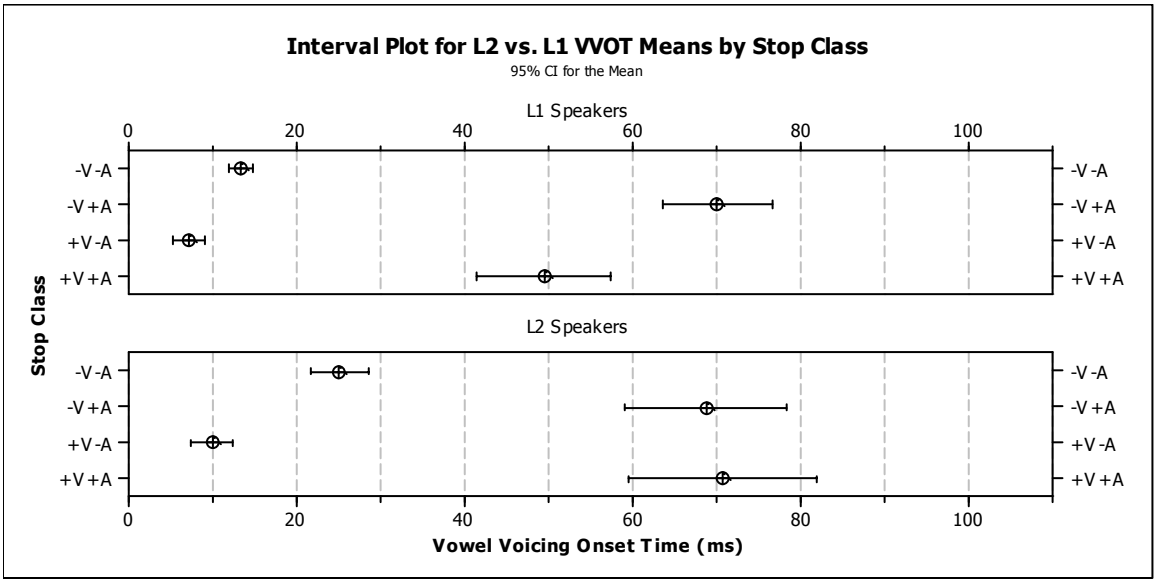
	Unaspirated	Aspirated
Voiceless	25.1 (18.9) [116] {13.3 (7.6) [129]}	68.7 (49.1) [100] {70.1 (31.9) [93]}
Voiced	10.0 (14.0) [120] {7.1 (11.2) [128]}	70.7 (60.8) [114] {49.4 (42.8) [115]}

Standard deviations in (parentheses) and sample sizes in [brackets]. L1 values in {braces}.

(a)



(b)

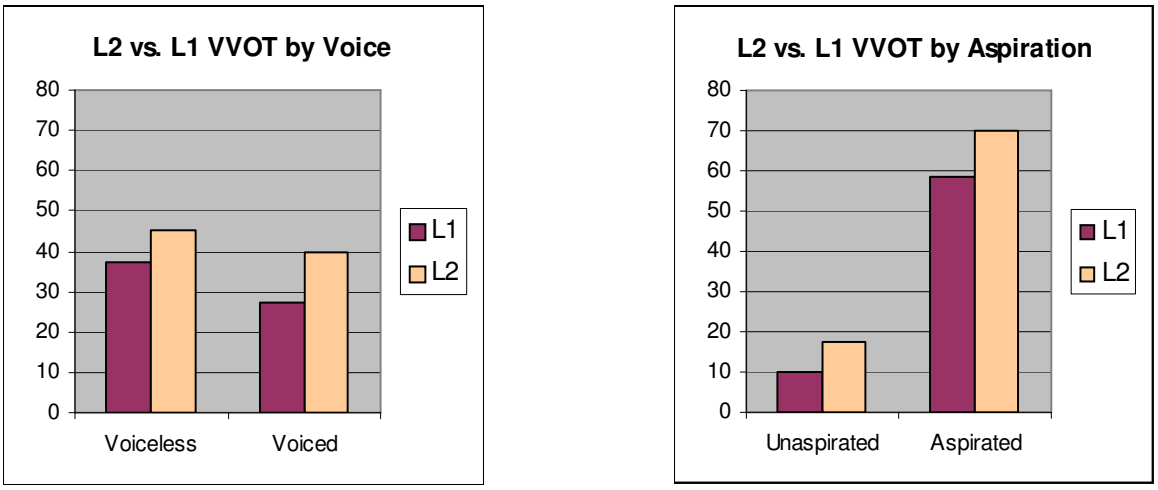


(c)

Figure 24. L2 vs. L1 Vowel Voicing Onset Time Means (ms) by Stop Class in: (a) Voice and Aspiration Matrix; (b) Bar Chart; (c) Interval Plot

(p = 0.000). This same pattern is recognizable for each of the stop classes except –V+A, which shows similar means for both groups.

As with the L1 subjects, Stop Class had a significant effect (p = 0.000) on VVOT, as did Aspiration (p = 0.000), but the effect of Voice was insignificant (p = 0.208). The L2 means also did not exhibit as great of a tendency to be larger for the voiceless stops (cf. Figure 25). As can be seen in Figure 24(c), the L2 means showed a greater deal of variance for each of the stop classes; additionally, the L2 speakers showed the most difference in Voice-related duration with the unaspirated stops, whereas for the L1 speakers this was true for the aspirated stops. As Figure 25 shows, the L2 means follow the same patterns as those of the L1 speakers in regards to Voice and Aspiration.



(a) **(b)**

Figure 25. Effect of (a) Voice and (b) Aspiration on L2 vs. L1 Vowel Voicing Onset Time Means (ms)

The significant effect (p = 0.006) that Position in the Word had on VVOT is illustrated in Figure 26. Particularly interesting to note is the much greater difference in VVOT among the subject groups in the final position. As Figure 27(c) shows, this effect is especially concentrated

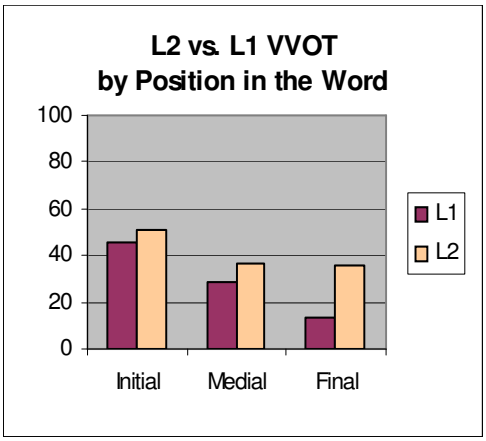
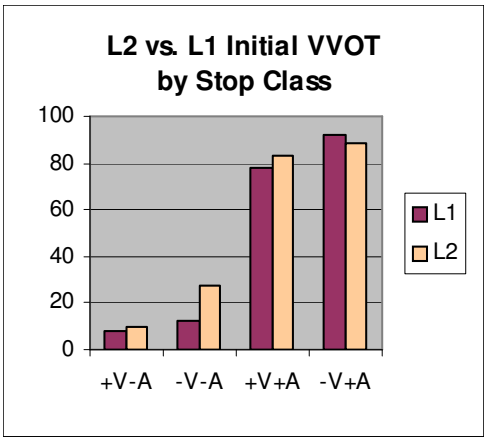
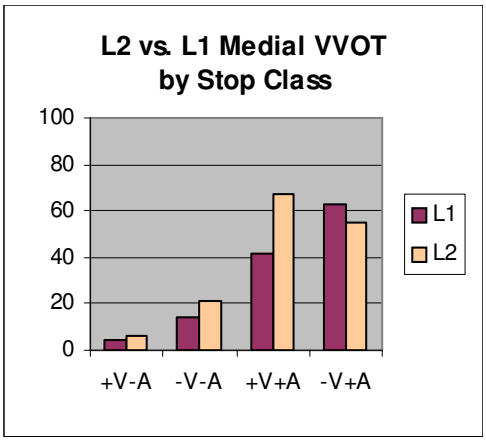


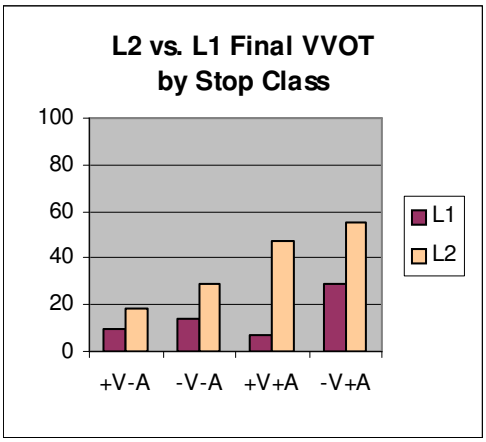
Figure 26. L2 vs. L1 Vowel Voicing Onset Time Means (ms) by Position in the Word



(a)



(b)



(c)

Figure 27. L2 vs. L1 Vowel Voicing Onset Time Means (ms) by Stop Class in (a) Initial, (b) Medial, and (c) Final Positions in the Word

in the +V+A stop class, where the L2 mean is seven times greater than the L1 mean, as opposed to being about two times greater in the other three stop classes. The primary cause of this was found to be three L2 speakers whose individual means were 5–12 times greater than the largest individual L1 speaker mean. The L2 stop class pattern observed in Figure 24(b) is somewhat evident at each position in the word, but there is a fair amount of position-dependent variation among the aspirated stops.

Other items to note in Figure 27 are as follows:

1. Among the –V–A stops, the L2 means are about twice the size of the L1 means in the initial and final positions, but this effect is lessened in the medial position.
2. Among the +V+A stops, both the L1 and L2 means decline from initial to medial to final position, but at a much different rate.
3. Among the –V+A stops, the L2 means tend to be smaller than those of the L1 in initial and medial positions, even though their tendency in the other stop classes is to be larger. However, in final position they are noticeably larger than those of the L1, remaining about the same as the means for medial position.

2.3.2 Closure Duration for L2 Speakers

As shown in Figure 28, the L2 CD means for the +V+A and –V+A stop classes differ most noticeably from those of the L1. As Figure 29 shows, the L2 means were noticeably greater for the initial and medial positions, but in the final position were similar to those of the L1 group. The overall stop class pattern in Figure 28(b) is evident in each position in the word, as seen in Figure 30, except for the +V–A stops in initial position, where the L2 mean is noticeably greater.

2.3.3 Closure Voicing for L2 Speakers

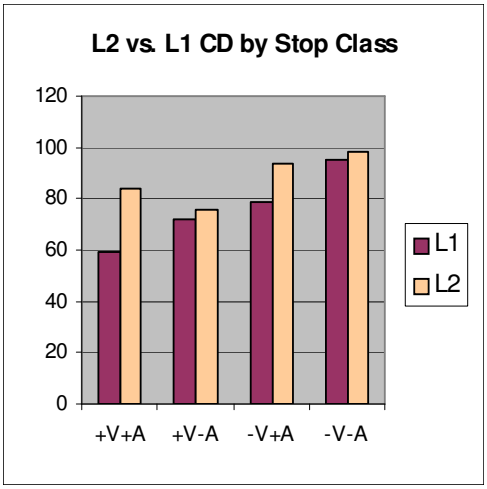
As illustrated in Figure 31, the L2 CV means as well as their patterning over the four stop classes are fairly similar to those of the L1 except for the +V+A class where they are

noticeably smaller. A comparison of L1 and L2 CV by Stop Class in each position in the word is presented in Figure 32, in which it is evident that the L2 CV means are most similar to those of the L1 in the final position, and that in both the initial and medial positions the L2 +V+A means are relatively smaller than those of the L1 speakers.

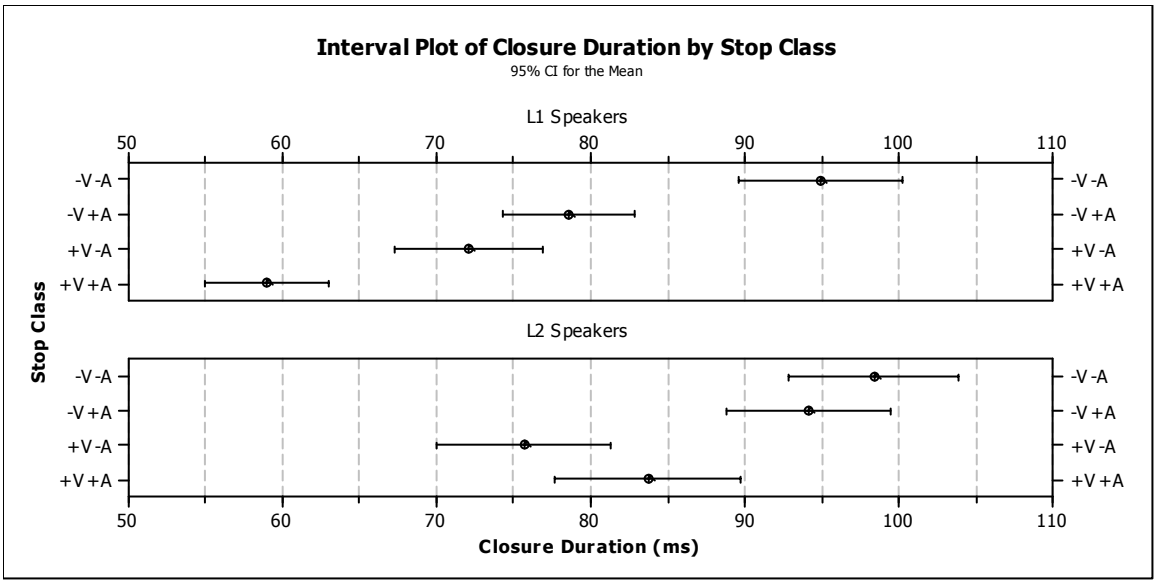
	Unaspirated	Aspirated
Voiceless	98.4 (33.3) [143] {94.9 (31.4) [138]}	94.1 (29.6) [122] {78.5 (23.5) [118]}
Voiced	75.7 (32.7) [133] {72.0 (26.7) [121]}	83.7 (34.2) [126] {59.0 (22.6) [123]}

Standard deviations in (parentheses) and sample sizes in [brackets]. L1 values in {braces}.

(a)



(b)



(c)

Figure 28. L2 vs. L1 Closure Duration Means (ms) by Stop Class in: (a) Voice and Aspiration Matrix; (b) Bar Chart; (c) Interval Plot

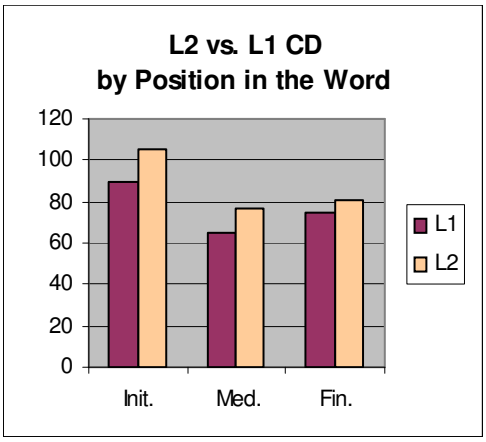


Figure 29. L2 vs. L1 Closure Duration Means (ms) by Position in the Word

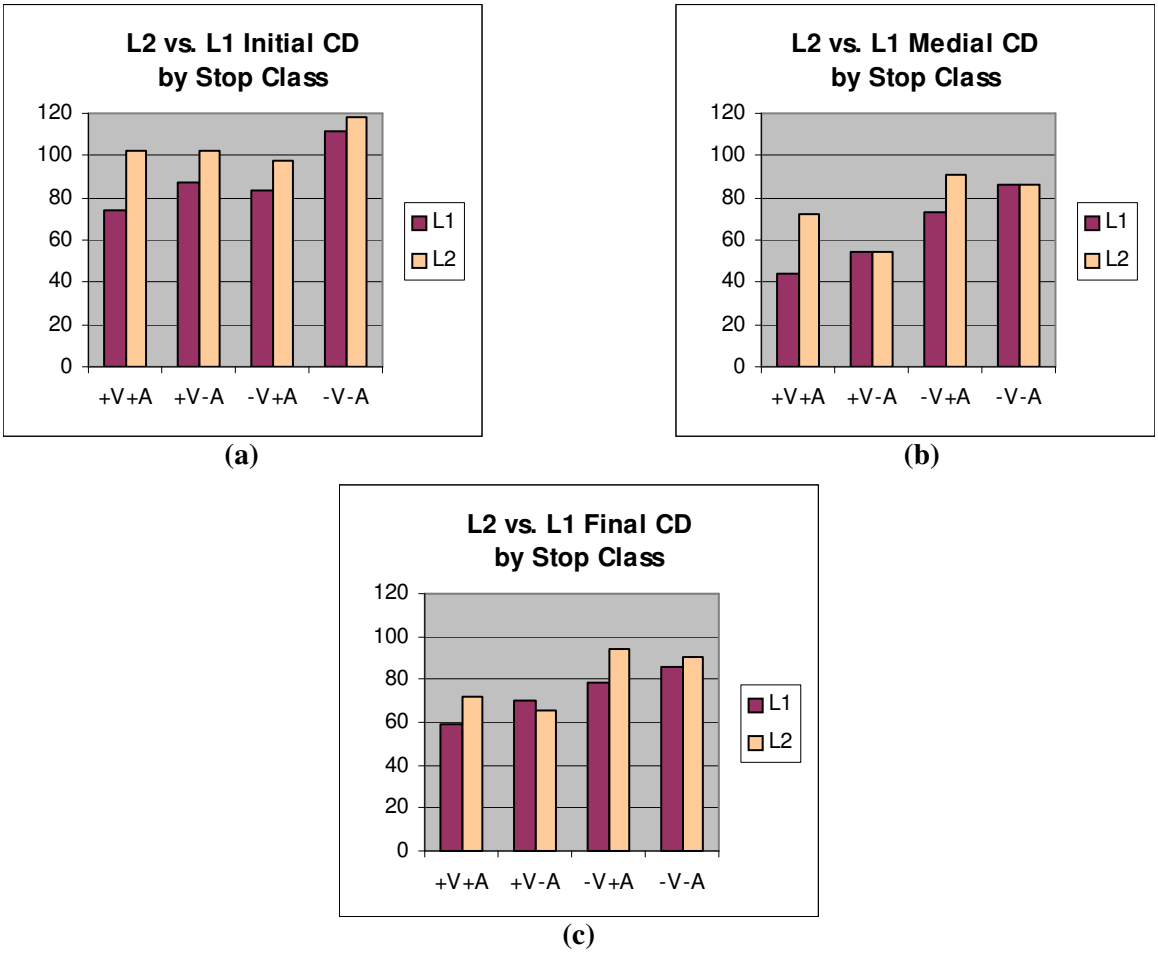
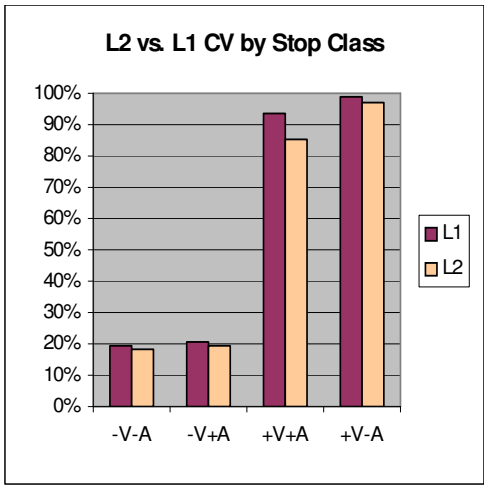


Figure 30. L2 vs. L1 Closure Duration Means (ms) by Stop Class in: (a) Initial, (b) Medial, and (c) Final Positions

	Unaspirated	Aspirated
Voiceless	18.4% (12.4) [137] {19.4% (10.9) [138]}	19.7% (12.6) [121] {20.5% (9.9) [115]}
Voiced	97.0% (8.7) [133] {98.6% (5.4) [121]}	85.2% (22.4) [126] {93.6% (14.4) [123]}

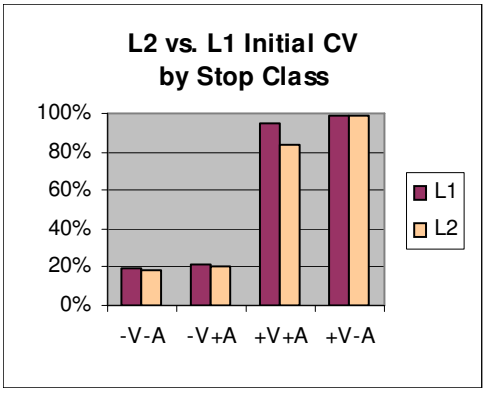
Standard deviations in (parentheses) and sample sizes in [brackets]. L1 values in {braces}.

(a)

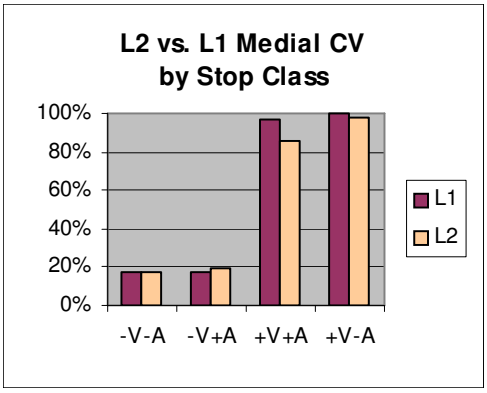


(b)

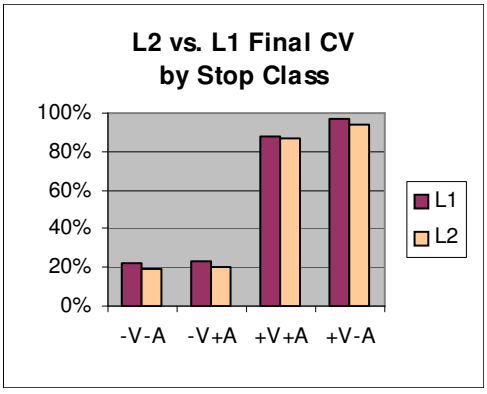
Figure 31. L2 vs. L1 Closure Voicing Means (as % of Closure Duration) by Stop Class in: (a) Voice and Aspiration Matrix; (b) Bar Chart



(a)



(b)



(c)

Figure 32. L2 vs. L1 Closure Voicing Means (as % of Closure Duration) by Stop Class in: (a) Initial, (b) Medial, and (c) Final Positions in the Word

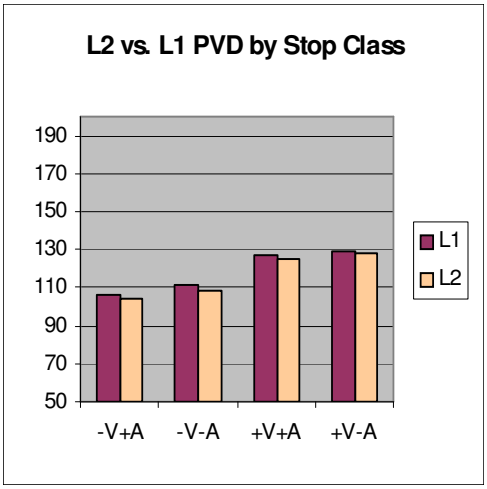
2.3.4 Preceding Vowel Duration for L2 Speakers

The PVD means for L2 vs. L1 by Stop Class are presented in Figure 33. The L2 means follow the same general pattern as those of the L1 speakers, although aside from the +V–A class, they are noticeably smaller. As can be seen in Figure 34, the difference between the L1 and L2 PVD means is not consistent across all positions in the word.

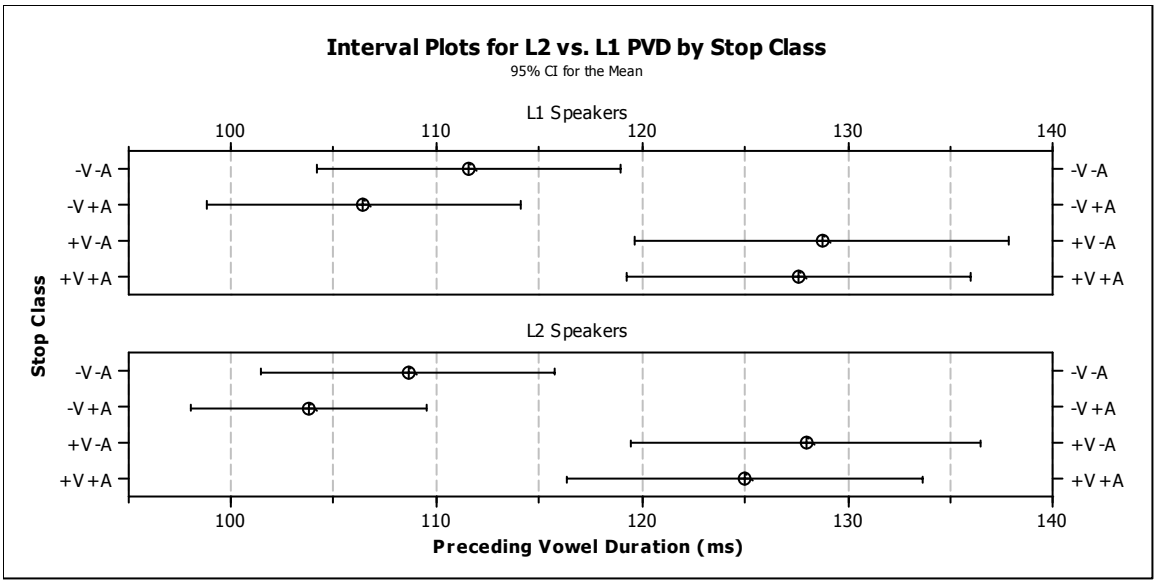
	Unaspirated	Aspirated
Voiceless	108.6 (43.6) [144] {111.6 (44.8) [144]}	103.8 (34.8) [144] {106.4 (46.4) [144]}
Voiced	128.0 (51.7) [144] {128.8 (55.3) [144]}	125.0 (50.4) [132] {127.6 (48.8) [132]}

Standard deviations in (parentheses) and sample sizes in [brackets]. L1 values in {braces}.

(a)



(b)



(c)

Figure 33. L2 vs. L1 Preceding Vowel Duration Means (ms) by Stop Class in: (a) Voice and Aspiration Matrix; (b) Bar Chart; (c) Interval Plot

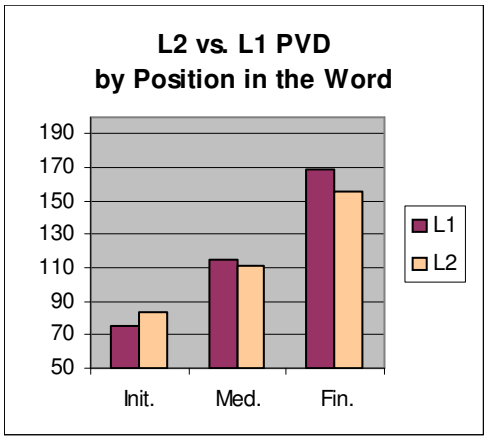


Figure 34. L2 vs. L1 Preceding Vowel Duration Means (ms) by Position in the Word

As seen in Figure 35, the –V+A and –V–A stops have larger means in initial position, the

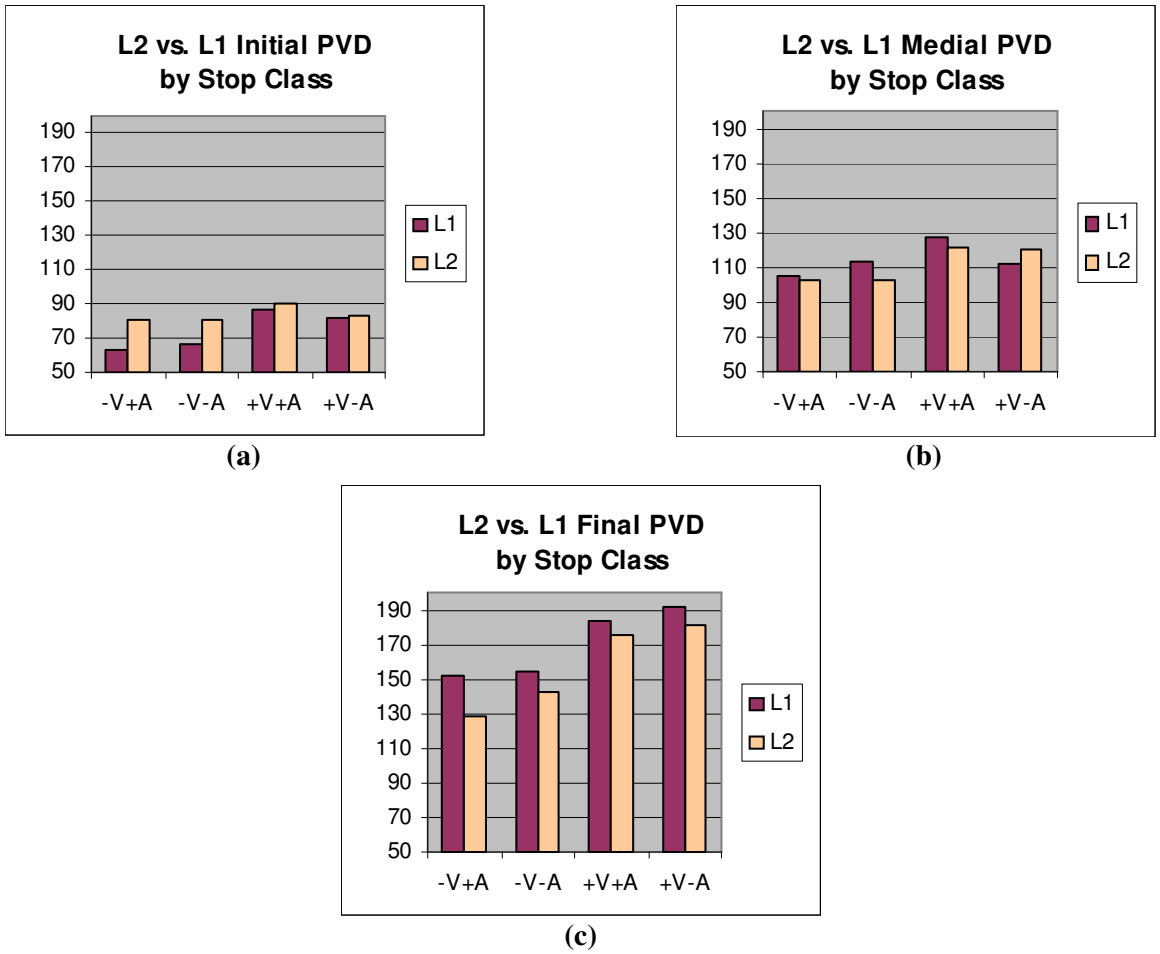


Figure 35. L2 vs. L1 Preceding Vowel Duration Means (ms) by Stop Class in: (a) Initial, (b) Medial, and (c) Final Positions in the Word

–V–A stops have smaller means in the medial position, and the –V–A, +V–A, and –V+A stops have smaller means in final position, with the difference for the latter group being quite large.

2.3.5 Aspiration Noise for L2 Speakers

Unlike the L1 group, a small percentage of the L2 –V–A stops had aspiration noise in their post-release phase; thus, they are included with the two aspirated stop classes in the following results. The percentage of L2 aspirated stops in which this acoustic correlate was present was 71%, as opposed to 57% of the corresponding L1 stops. As illustrated in Figure 36, the main differences in the L2 means were the presence of aspiration noise for the –V–A stop class and the larger means for the +V+A stop class.

Comparisons of the L1 and L2 aspiration noise duration means at each position in the word are shown in Figure 37. Besides the obvious difference of the presence of aspiration noise in the –V–A productions of the L2 speakers, the following were the most important observations made:

1. The L2 +V+A means are larger in both initial and medial positions.
2. The L2 –V+A means are larger in final position.
3. The L2 –V–A means are smaller in medial position than in the initial and final positions.

The distribution of instances of aspiration noise was 43% for initial, 34% for medial, and 23% for final position, as opposed to the L1 distribution of 53%, 37%, and 10%, respectively.

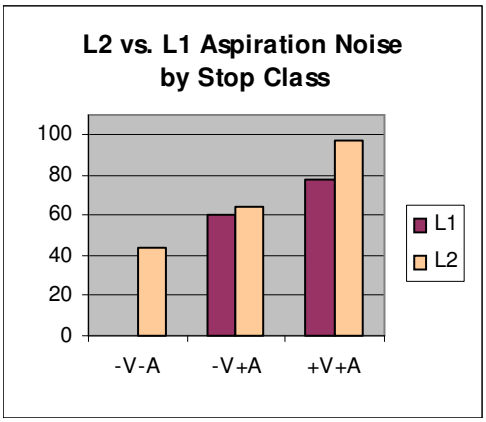
2.4 Gender Effects

There was a significant effect ($p = 0.000$) of Gender on L1 VVOT for the +V+A stop class, as illustrated in Figure 38, consisting of a female mean that was about twice as large as the male mean.

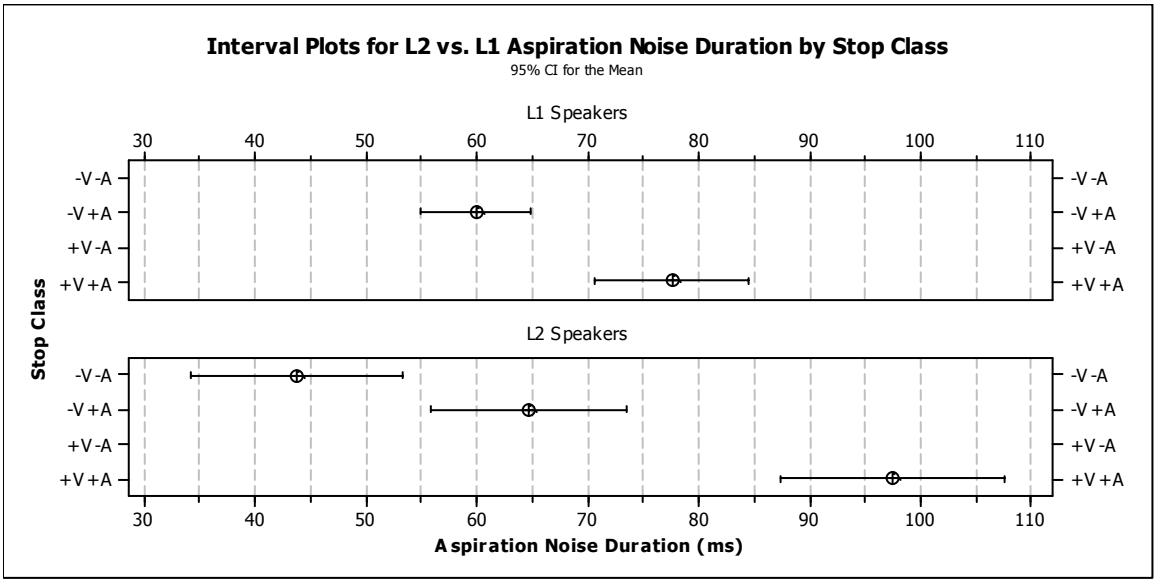
	Unaspirated	Aspirated
Voiceless	43.8 (22.7) [24] {N/A}	64.7 (44.3) [99] {59.9 (24.3) [95]}
Voiced	N/A {N/A}	97.5 (43.1) [73] {77.6 (27.4) [61]}

Standard deviations in (parentheses) and sample sizes in [brackets]. L1 values in {braces}.

(a)



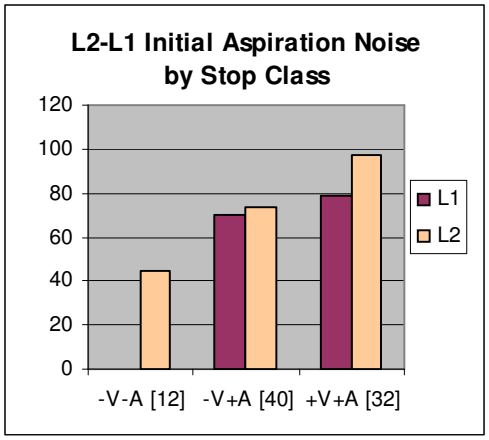
(b)



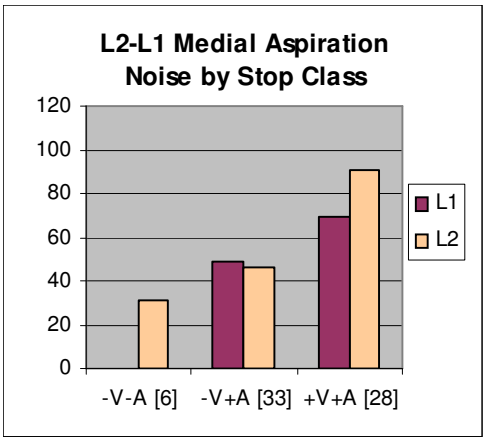
(c)

Figure 36. L2 vs. L1 Aspiration Noise Duration Means (ms) by Stop Class in: (a) Voice and Aspiration Matrix; (b) Bar Chart; (c) Interval Plot

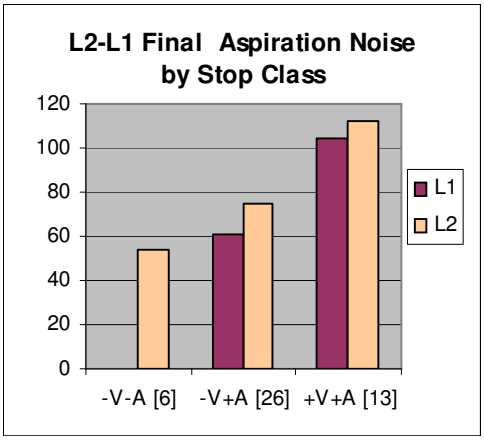
As shown in Figure 39, the L2 effect of Gender on VVOT was similar to that of the L1 (cf. Figure 38) in that the L2 female means were larger than those of the males for the aspirated stops. The L2 gender difference for the +V+A stops was smaller but still significant ($p = 0.018$), and for the –V+A stops the difference was larger and near the border of significance ($p = 0.062$). Another related observation is that each of the four cross-linguistic gender groups exhibited



(a)



(b)



(c)

Figure 37. L2 vs. L1 Aspiration Noise Duration Means (ms) by Stop Class in: (a) Initial, (b) Medial, and (c) Final Positions (L2 sample sizes in [brackets]. L1 sample sizes are as follows: Initial –V+A = 44, and +V+A = 38; Medial –V+A = 39, and +V+A = 19; and Final –V+A = 12, and +V+A = 4.)

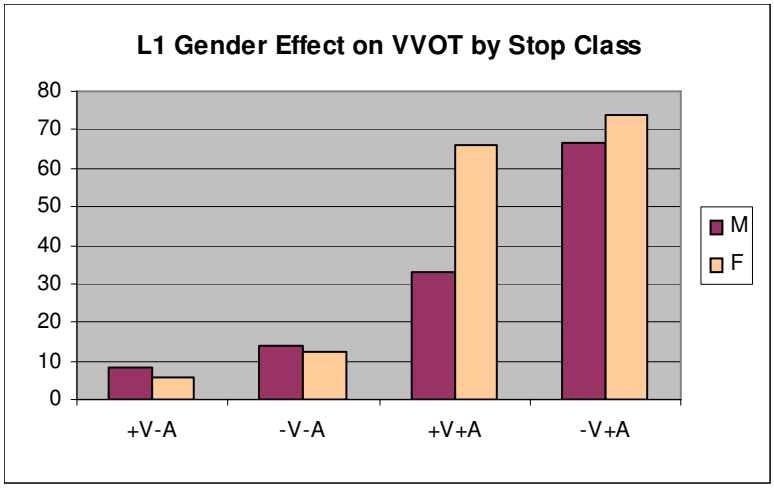


Figure 38. L1 Gender Effect on Vowel Voicing Onset Time Means (ms) by Stop Class [Key: M = Male. F = Female.]

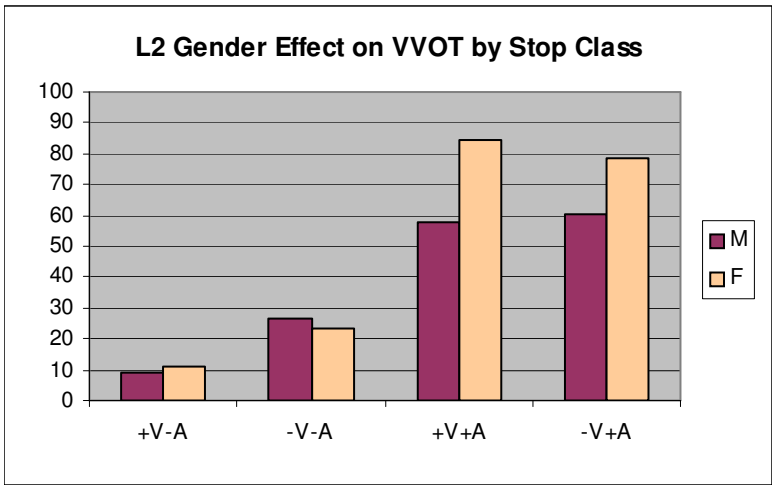


Figure 39. L2 Gender Effect on Vowel Voicing Onset Time Means (ms) by Stop Class [Key: M = Male. F = Female.]

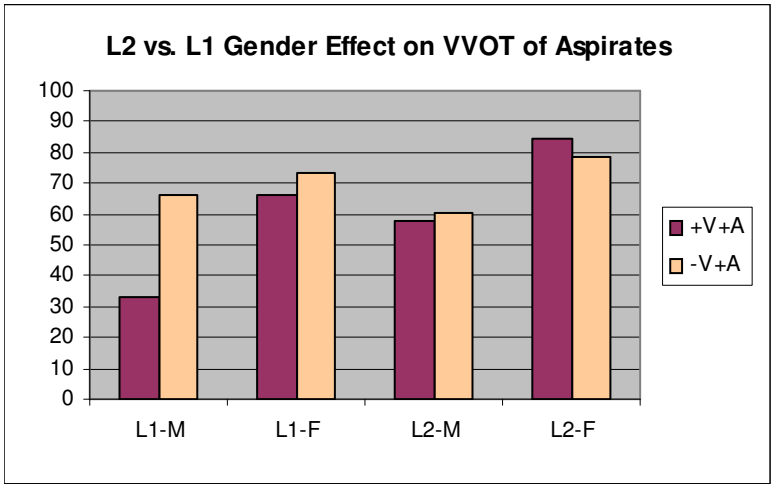


Figure 40. L2 vs. L1 Gender Effect on Vowel Voicing Onset Time Means (ms) for the Aspirated Stop Classes [Key: M = Male. F = Female.]

fairly similar means for both classes of aspirated stops except for the L1 males, whose +V+A means were only about half as large as their –V+A means (see Figure 40).

There was an effect of Gender on L1 CD in that the female means were consistently larger than those of the males for each of the stop classes, as shown in Figure 41. This effect was

significant ($p \leq 0.036$) for all of the stop classes except +V+A where the effect was insignificant ($p = 0.108$). For the L2 group, there was a similar effect of the female means being consistently larger for each stop class, as shown in Figure 42. However, this effect was near the borderline of significance for the –V–A ($p = 0.050$), –V+A ($p = 0.042$), and +V–A ($p = 0.076$) stops, and was insignificant ($p = 0.483$) for the +V+A stops.

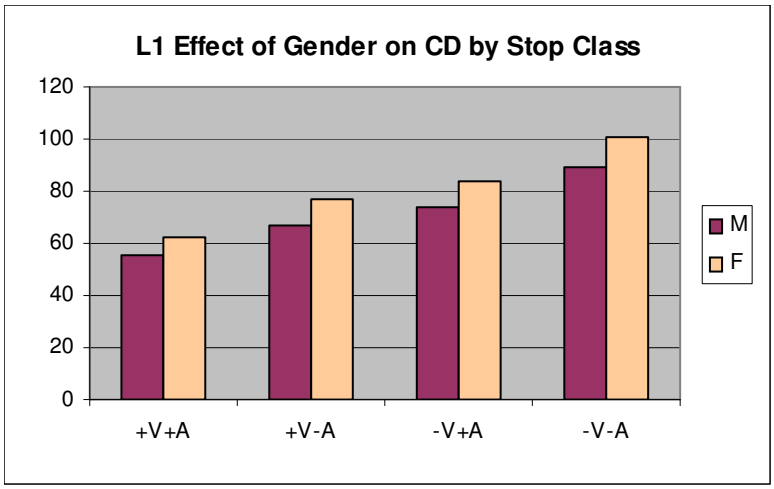


Figure 41. L1 Gender Effect on Closure Duration Means (ms) by Stop Class
[Key: M = Male. F = Female.]

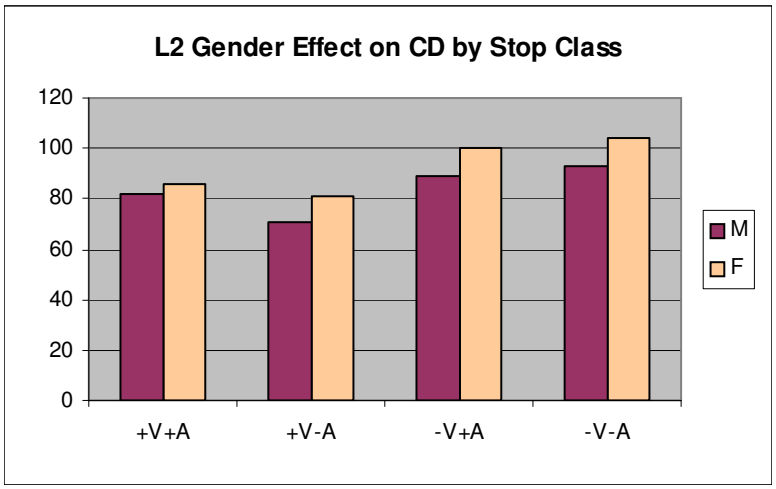


Figure 42. L2 Gender Effect on Closure Duration Means (ms) by Stop Class
[Key: M = Male. F = Female.]

As Figure 43 shows, the female mean for L1 CV was smaller than that of the males for the +V+A stops, and this difference was significant ($p = 0.003$).

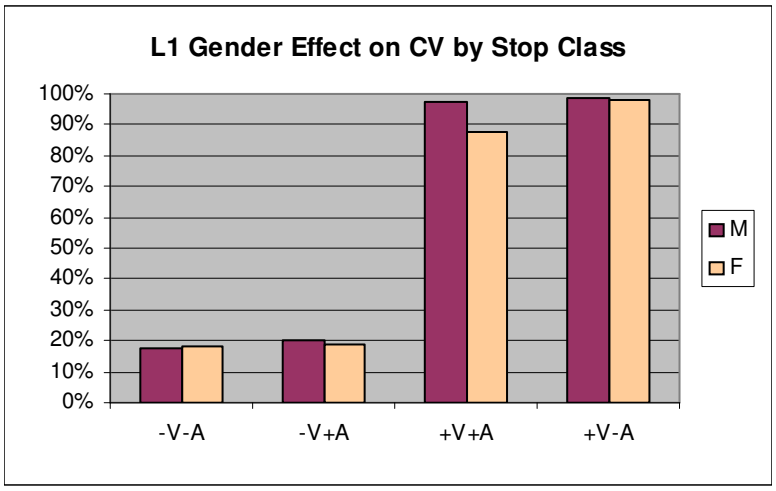


Figure 43. L1 Gender Effect on Closure Voicing Means (as % of Closure Duration) by Stop Class [Key: M = Male. F = Female.]

There were no major effects of Gender on the L2 CV means across the four stop classes, as Figure 44 shows, but the differences in each of the voiceless classes were significant ($p \leq 0.004$).

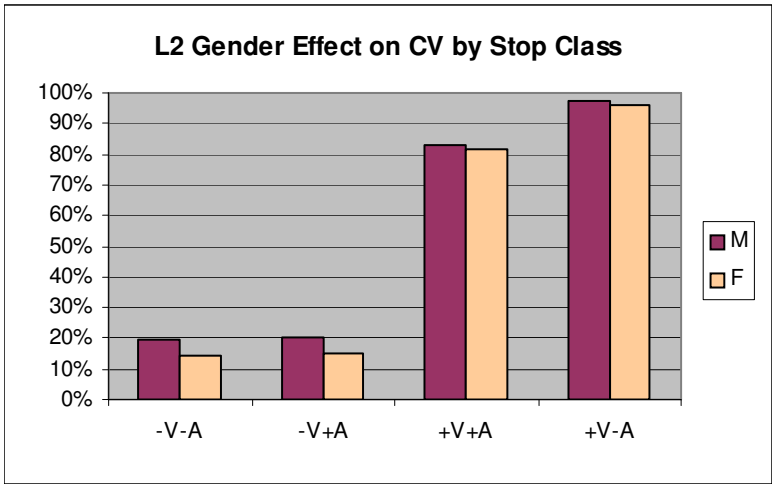


Figure 44. L2 Gender Effect on Closure Voicing Means (as % of Closure Duration) by Stop Class [Key: M = Male. F = Female.]

There was a significant effect ($p \leq 0.015$) of Gender on L1 PVD for each of the stop classes, as shown in Figure 45, with the female means being consistently larger than those of the males.

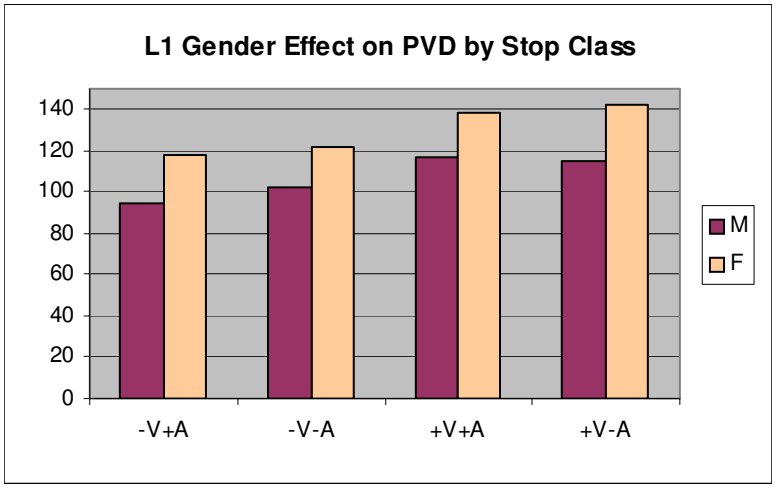


Figure 45. L1 Gender Effect on Preceding Vowel Duration Means (ms) by Stop Class [Key: M = Male. F = Female.]

As illustrated in Figure 46, for the L2 PVD means the only significant effect ($p = 0.001$) of Gender was in the –V–A stop class, where the female means were noticeably smaller than the

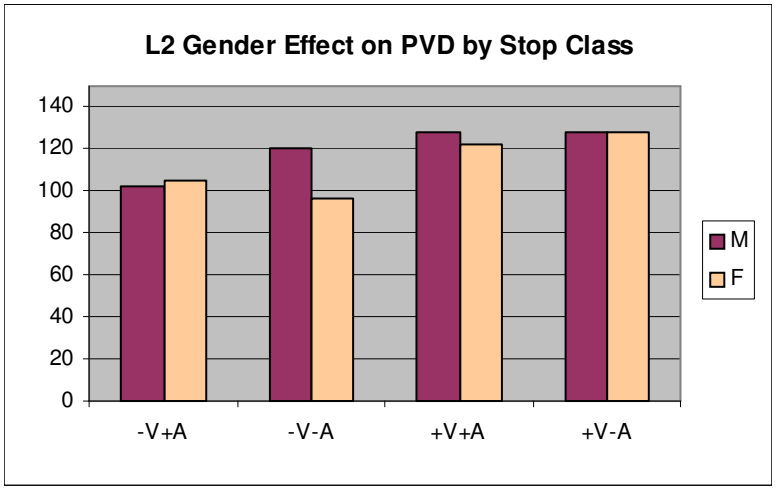


Figure 46. L2 Gender Effect on Preceding Vowel Duration Means (ms) by Stop Class [Key: M = Male. F = Female.]

male means. This is quite different from the corresponding L1 results (cf. Figure 45), where the female means were consistently and equally larger than the male means for each of the stop classes.

For the L1 speakers, there was a significant effect of Gender ($p \leq 0.034$) on duration of aspiration noise, as seen in Figure 47. The female means were larger than those of the males, and the greatest gender difference was among the +V+A stops. The effect of Stop Class on aspiration noise duration was significant for the females ($p = 0.001$), but insignificant for the males ($p = 0.235$). For the aspirated stop classes, the L2 group exhibited an effect of Gender similar to that of the L1 speakers, as shown in Figure 48, but the only difference that was significant ($p = 0.030$) was that within the +V+A class. Additionally, for the –V–A stops the female mean was somewhat smaller than the male mean, a difference that was insignificant ($p = 0.360$).

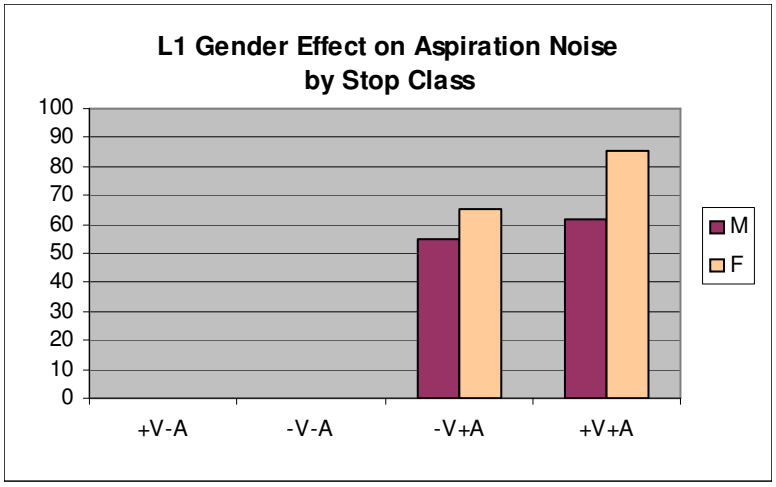


Figure 47. L1 Gender Effect on Aspiration Noise Duration Means (ms) by Stop Class [Key: M = Male. F = Female.]

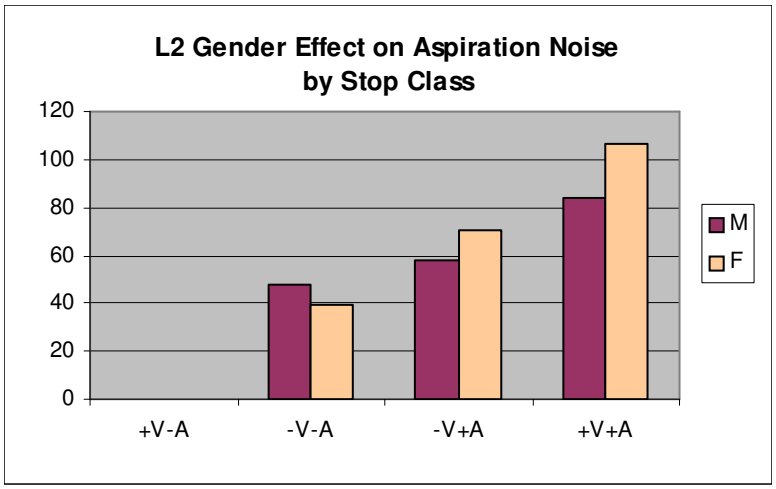


Figure 48. L2 Gender Effect on Aspiration Noise Duration Means (ms) by Stop Class [Key: M = Male. F = Female.]

CHAPTER 3

DISCUSSION AND IMPLICATIONS

In this chapter a summary of the L1 results is given, including a look at the contributions of this study to the literature on stops in Bangla and other Indo-Aryan languages. The L2 results are then compared to those of the L1 group, accompanied by a brief summary of the effect of Gender within each of the stop classes, and the negative correlation between Duration of Exposure to Bangla on the one hand, and the standard deviation of VVOT values on the other hand. Then, a short acoustic phonetic characterization of each of the four stop classes is constructed based on observations in this study, and the implications of the research results for L2 pronunciation acquisition are explored. Finally, some suggestions for further research are given.

3.1 Summary of Main Results

Several key effects were observed in the results of the L1 speaker group. Stop Class, Voice, and Aspiration each had a significant effect on vowel voicing onset time, with Aspiration having the main effect, indicating that VVOT is a primary acoustic correlate of Aspiration in Bangla. Both Voice and Aspiration had a significant effect on closure duration, with the voiced and the aspirated groups tending to have smaller CD means. There was clear and consistent evidence that closure voicing is the primary acoustic correlate of Voice in Bangla, and an interesting finding was that CV and VVOT interact to clearly distinguish each of the four stop classes from each other (cf. Figure 17). It was also observed that aspiration noise was not an

obligatory feature of Aspiration, although its presence was quite common. Additionally, in more than 10% of the L1 voiced aspirated tokens, there was at least some portion of the aspiration noise that was not superimposed on a periodic signal. These findings emphasize that a relatively long VVOT is a primary acoustic cue of the aspirated stops and that a significant amount of closure voicing is a primary acoustic cue of the voiced stops. This in turn indicates that for the +V+A stops the term 'voiced aspirated' is more accurate than the term 'breathy voiced' since it more clearly expresses not only the unity this stop class has with the voiced unaspirated stops in regards to Voice but also the unity it has with the voiceless aspirated stops in regards to Aspiration.

The L1 VVOT results in this study also support the work of Davis (1994), Mikuteit and Reetz (2007), and others who have suggested that a [breathy voice] feature is not necessary for making a phonological distinction between the voiced unaspirates and aspirates of the four-category stop systems of the Indo-Aryan languages. In her study of initial velar stops in Hindi, Davis showed that her noise-offset time (NOT) measure successfully distinguished all four stop classes, and Mikuteit and Reetz used their ACT measure with the medial velar stops (singletons only) in Bangla to achieve comparable results. In Table 3 the VVOT means for the L1 initial and medial velar stops in this study are compared with both of their results. Even though a precise comparison is not possible because of differences in how these studies were conducted, the similar pattern of each set of results is worth noting.

For ease of reference in the following discussion, comparisons of the L1 and L2 summary results for each stop class at each position in the word are illustrated in Figures 49-52. The L2 speakers were similar to the L1 group in that Stop Class and Aspiration had a significant effect on VVOT; however, the effect of Voice was not significant for the L2 group. In the initial position, the +V+A mean was about the same as that of the L1 speakers, but it was quite a bit

larger in the medial position, and even larger in the final position. The –V+A means, however, were somewhat smaller than those of the L1 speakers in both initial and medial positions, but then were larger in final position. Additionally, the L2 CD means for the aspirated stops in medial and final position were larger than those of the L1 speakers, even though their unaspirated counterparts had means similar to the L1 speakers. In initial position a similar pattern was followed, except that the +V–A mean was also greater than that of the L1 group. The main difference for the PVD means was that they were consistently smaller than those of the L1 group in final position.

Table 3. Comparison of Vowel Voicing Onset Time Means by Stop Class with Similar Data from Davis (1994) and Mikuteit and Reetz (2007)

Phoneme	Davis (1994) (NOT)	Miller (VVOT)	M&R (2007) (ACT)
k	25.85 (10.64) [139]	21.49 (5.77) [24]	36.06 (6.95) [60]
kʰ	74.24 (17.58) [158]	86.58 (31.45) [23]	72.83 (13.42) [60]
g	11.58 (11.06) [156]	8.92 (7.77) [24]	22.42 (9.18) [60]
gʱ	34.26 (21.30) [147]	57.03 (38.45) [24]	43.82 (23.57) [60]

Key: NOT = Noise offset time. VVOT = Vowel voicing onset time. M&R = Mikuteit and Reetz. ACT = After closure time. Standard deviations in (parentheses) and sample sizes in [brackets].

In regards to aspiration noise duration, the L2 group produced some of the –V–A stops with aspiration noise, and they had a larger mean than the L1 speakers for the +V+A stops. The L2 speakers also had slightly fewer instances of aspiration noise in the initial position than the L1 group, slightly more in the final position, and about the same in the medial position. Twenty-four (17%) of the –V–A tokens were aspirated and their aspiration noise duration mean (44 ms) was smaller than that of the –V+A (65 ms) and +V+A (98 ms) tokens. One male and three females did not aspirate these tokens at all; of the remaining eight subjects, the range of aspirated

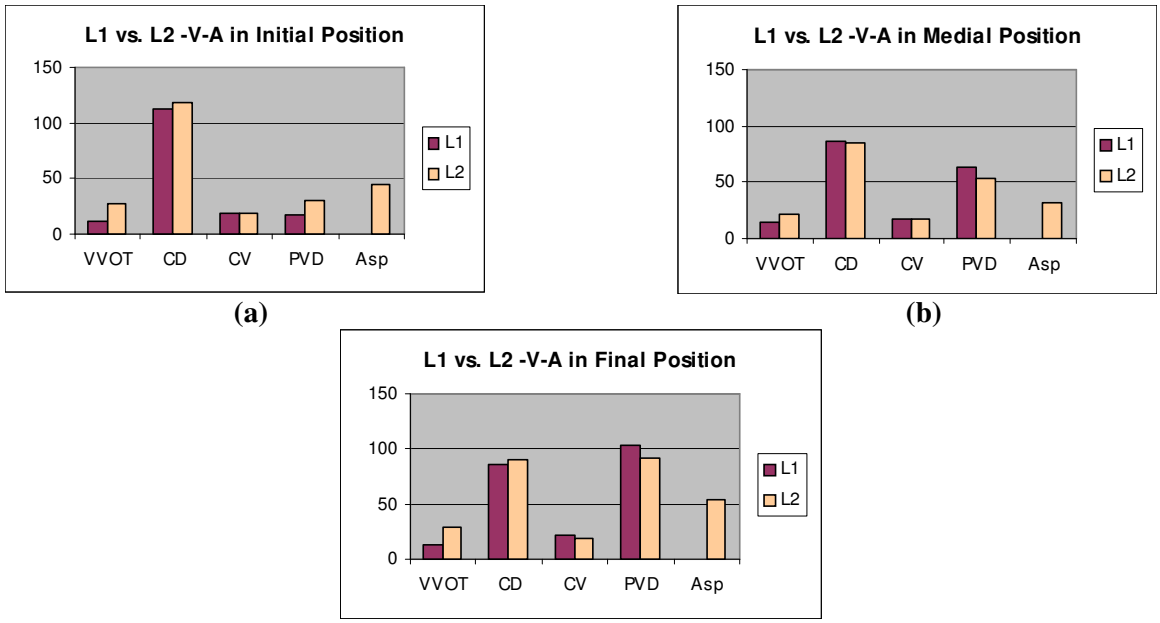


Figure 49. L1 vs. L2 Realizations of the Voiceless Unaspirated Stops in: (a) Initial, (b) Medial, and (c) Final Positions in the Word [Key: VVOT = Vowel Voicing Onset Time Means (ms). CD = Closure Duration Means (ms). CV = Closure Voicing Means (as % of Closure Duration). PVD = Preceding Vowel Duration Means (less 50 ms). Asp = Aspiration Noise Duration Means (ms).]

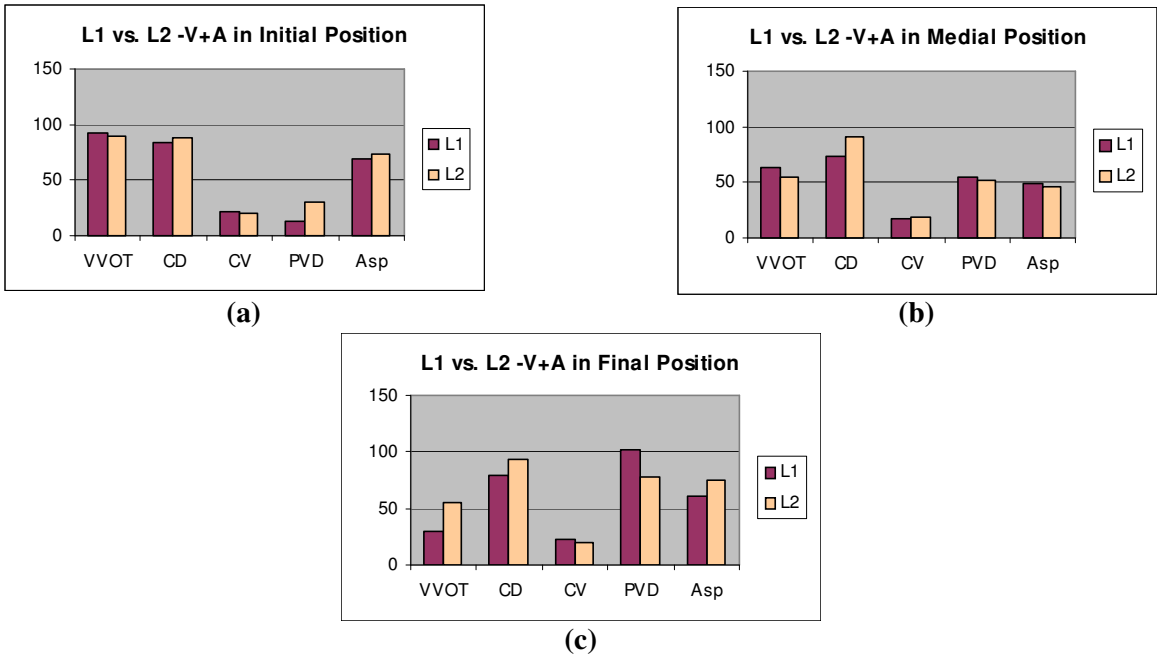


Figure 50. L1 vs. L2 Realizations of the Voiceless Aspirated Stops in: (a) Initial, (b) Medial, and (c) Final Positions in the Word [Key: VVOT = Vowel Voicing Onset Time Means (ms). CD = Closure Duration Means (ms). CV = Closure Voicing Means (as % of Closure Duration). PVD = Preceding Vowel Duration Means (less 50 ms). Asp = Aspiration Noise Duration Means (ms).]

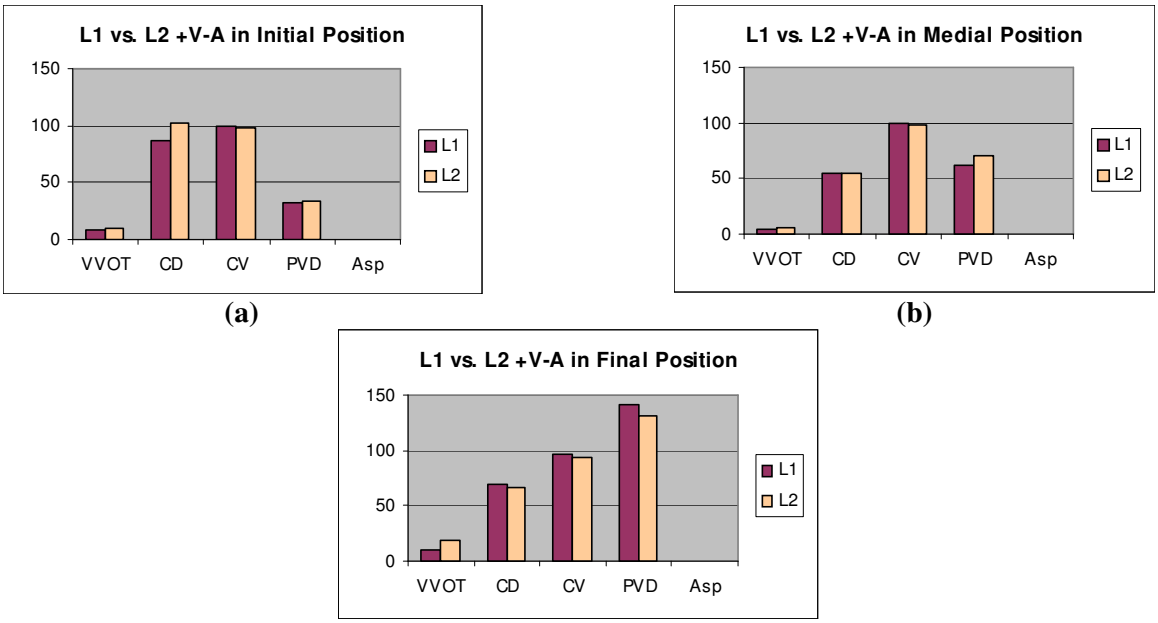


Figure 51. L1 vs. L2 Realizations of the Voiced Unaspirated Stops in: (a) Initial, (b) Medial, and (c) Final Positions in the Word [Key: VVOT = Vowel Voicing Onset Time Means (ms). CD = Closure Duration Means (ms). CV = Closure Voicing Means (as % of Closure Duration). PVD = Preceding Vowel Duration Means (less 50 ms). Asp = Aspiration Noise Duration Means (ms).]

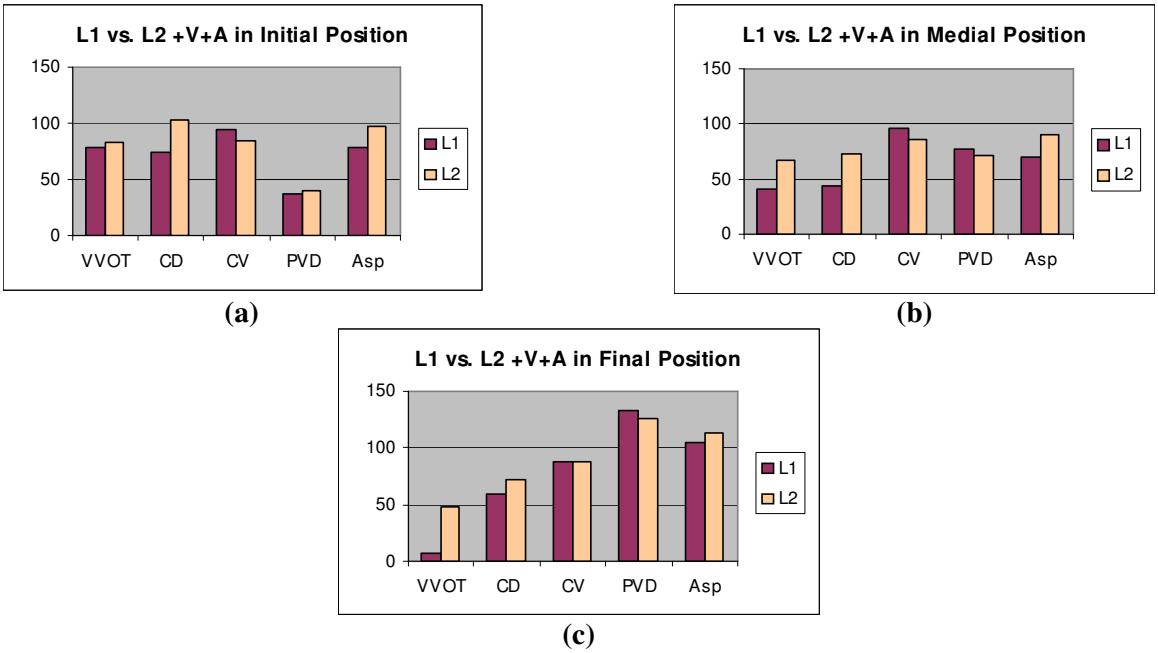


Figure 52. L1 vs. L2 Realizations of the Voiced Aspirated Stops in: (a) Initial, (b) Medial, and (c) Final Positions in the Word [Key: VVOT = Vowel Voicing Onset Time Means (ms). CD = Closure Duration Means (ms). CV = Closure Voicing Means (as % of Closure Duration). PVD = Preceding Vowel Duration Means (less 50 ms). Asp = Aspiration Noise Duration Means (ms).]

tokens per subject was 1–7 (out of a possible total of twelve) and the range of subject means was 18–56 milliseconds.

Duration of Exposure to Bangla (as measured by years of residence in Bangladesh) did not have a significant effect ($p = 0.452$) on the aspiration noise duration of the L2 –V–A tokens in which this acoustic variable was present, and also did not show a clear influence on the frequency of its occurrence, as illustrated in Figure 53. Of the four subjects who did not have any instances of aspiration noise in their productions of this class of stops, three had lived in Bangladesh for 0–4 years and one had lived there for more than four years. This suggests that there may not be a direct correlation between the amount of time native English speakers are exposed to Bangla and their acquisition of this aspect of a native-like production of this class of stops. However, there was a negative correlation between the standard deviation of aspiration noise duration values and subjects' Duration of Exposure to Bangla, as shown in Table 4 and Figure 54. The aspiration noise values of all subjects were divided into three sets by: L2 speakers who had lived in Bangladesh for 0–4 years, those who had lived there for more than four years,

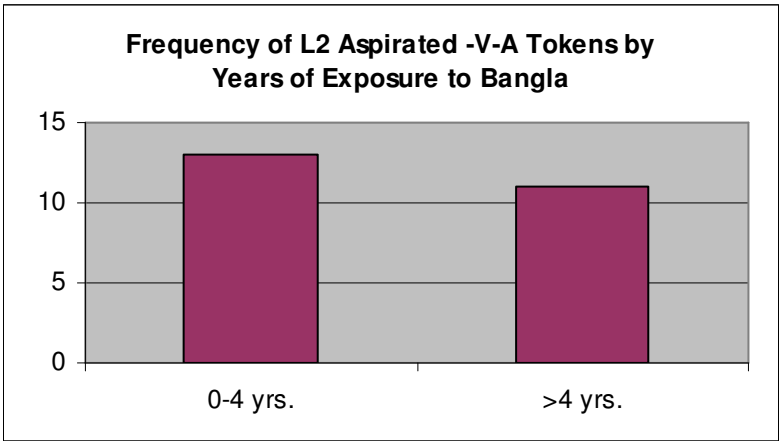


Figure 53. Relationship Between Frequencies of Aspiration Noise in L2 Productions of Voiceless Unaspirated Stops and Duration of Exposure to Bangla (As Measured by Years of Residence in Bangladesh)

and the L1 speakers, who had all lived in Bangladesh their entire lives. The standard deviation of the aspiration noise values for each of these groups was calculated and was found to be largest for L2 speakers who had been in the country the least amount of time, smaller for those who had been there for a greater amount of time, and smallest for the native Bangla speakers.

Table 4. Relationship Between Aspiration Noise Standard Deviation and Duration of Exposure to Bangla (As Measured by Years of Residence in Bangladesh) by Stop Class

	–V–A	**–V+A**	**+V–A**	**+V+A**	**All**
L2 (6): 0–4 yrs.	25.7 [13]	48.7 [53]	N/A	49.5 [42]	50.6 [108]
L2 (6): >4 yrs.	19.1 [11]	38.5 [46]	N/A	31.7 [31]	38.2 [88]
L1 (12): Life	N/A	24.3 [95]	N/A	27.4 [61]	26.9 [156]

Number of subjects in (parentheses). Sample sizes in [brackets].

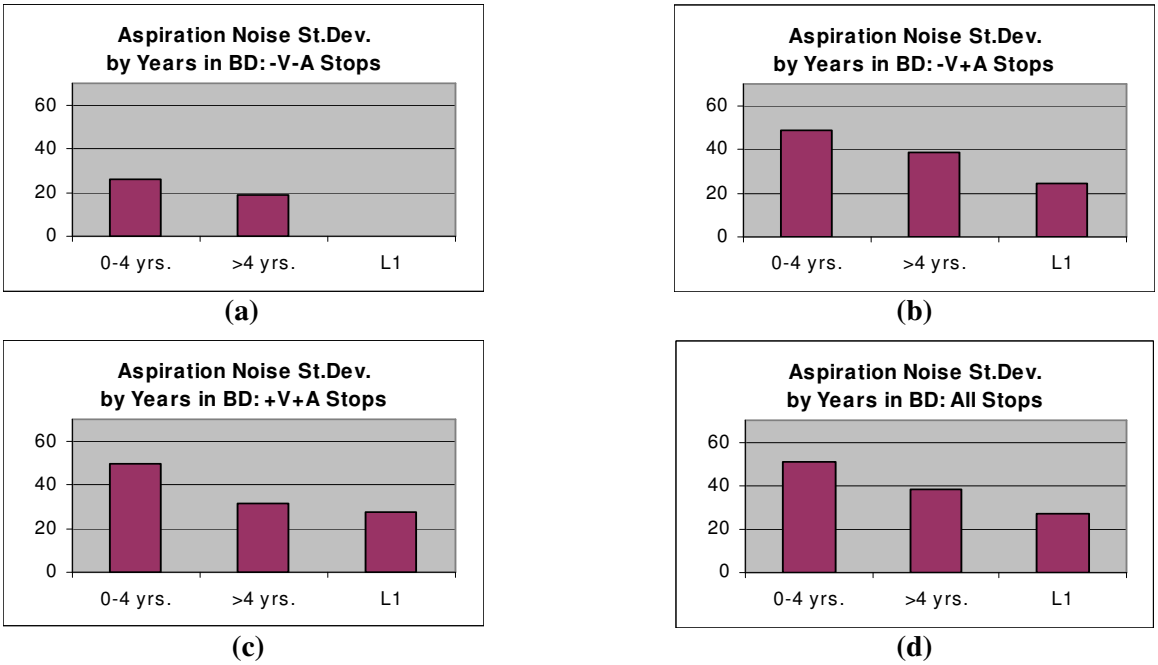


Figure 54. Relationship Between Aspiration Noise Standard Deviation (St.Dev.) and Duration of Exposure to Bangla (As Measured by Years of Residence in Bangladesh [BD]) for: (a) Voiceless Unaspirates; (b) Voiceless Aspirates; (c) Voiced Aspirates; (d) Stops of All Classes

Using the same type of grouping and calculation, evidence was also found of a similar negative correlation between the standard deviation of VVOT values and subjects' Duration of Exposure to Bangla, as detailed in Table 5 and further illustrated in Figure 55. As can be seen, the +V–A results are an exception to this tendency, which is possibly due to the fact that these stops are unaspirated and are perceived by L2 speakers as familiar sounds, thus decreasing the potential for their exaggeration or overgeneralization, which are factors that would likely affect standard deviation size. It is also likely that some or all of this effect on the VVOT values is a product of the similar effect on the aspiration noise values mentioned earlier. This effect was not noticeable for CD, CV, and PVD, the other three acoustic variables in this study.

A graphical summary and comparison of the L1 and L2 effects of Gender within each of the four stop classes is given in Figure 56 for ease of reference in the following discussion. For the +V+A stops, the L1 male VVOT mean was half that of the females, as well as being about half that of their own mean for the -V+A stops, an interesting relationship that was not observed for the L1 females or either of the L2 gender groups. The females in both subject groups had larger means than the males for the two aspirated classes, but this effect was strongest for the +V+A stops. The CD means for the females from both groups were consistently larger than the male means. For PVD, Gender had a considerably different effect in each of the subject groups. The L1 females had consistently larger means than the L1 males, whereas the L2 females had similar means to the L2 males, except for the –V–A stops, where their mean was significantly smaller than that of the males. For aspiration noise duration, the females of both groups had larger means than the males, which likely contributed to the effect on VVOT observed earlier. Further research would need to be done in order to determine the cause and significance of these effects of Gender.

Table 5. Relationship Between Vowel Voicing Onset Time Standard Deviation and Duration of Exposure to Bangla (As Measured by Years of Residence in Bangladesh) by Stop Class

	–V–A	–V+A	+V–A	+V+A	All
L2 (6): 0–4 yrs.	20.4 [59]	55.6 [53]	11.4 [64]	65.3 [60]	53.8 [236]
L2 (6): >4 yrs.	17.5 [57]	38.9 [47]	16.6 [56]	53.4 [54]	40.6 [214]
L1 (12): Life	7.6 [129]	31.9 [93]	11.2 [128]	42.8 [115]	36.6 [465]

Number of subjects in (parentheses). Sample sizes in [brackets].

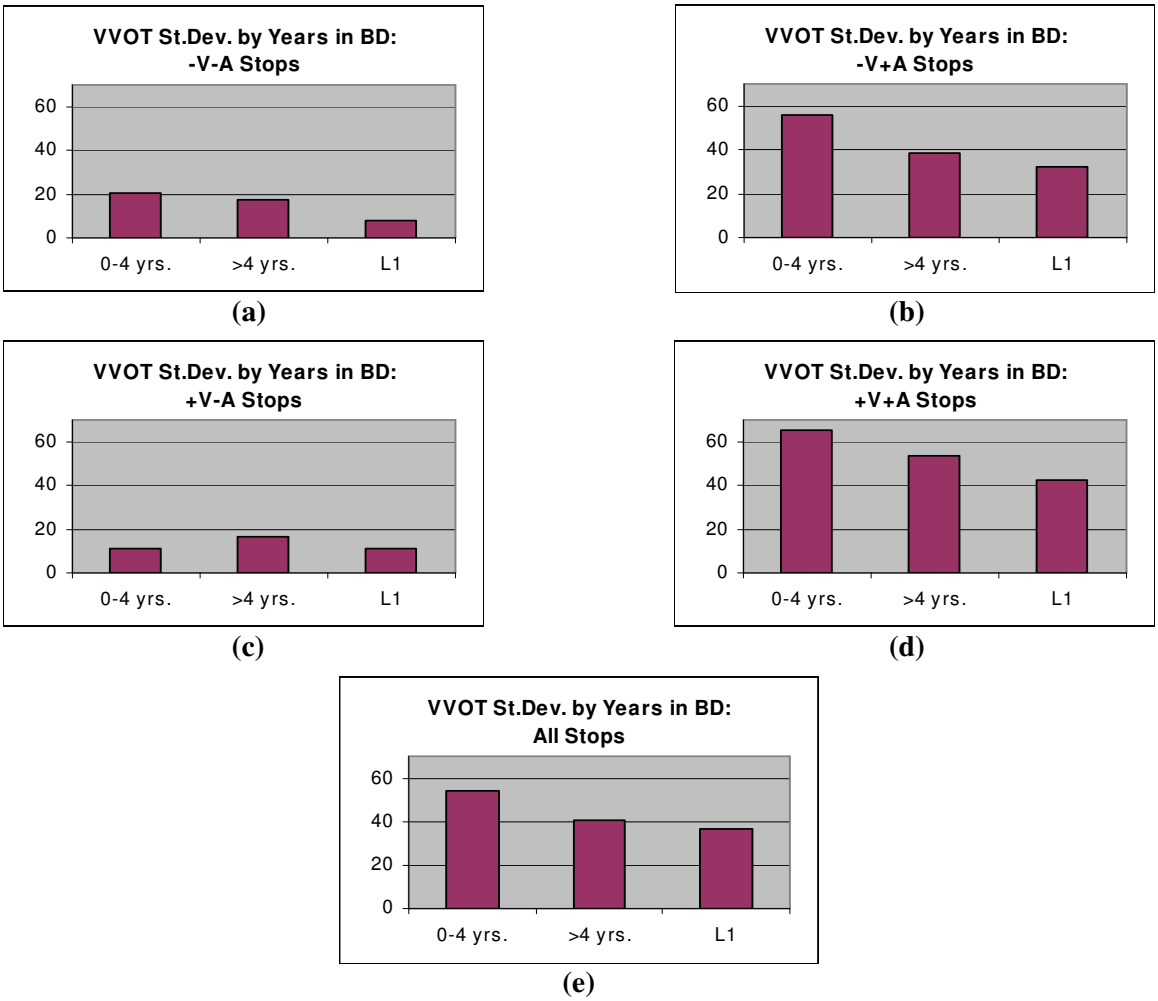
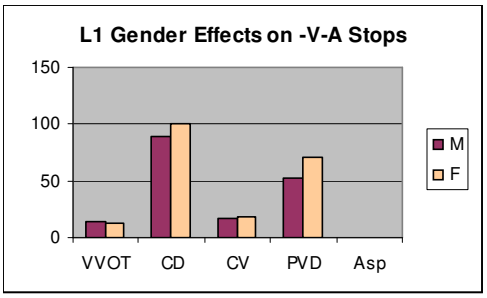
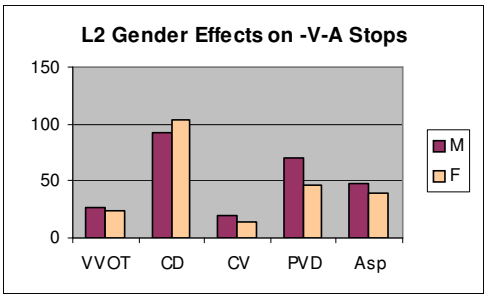


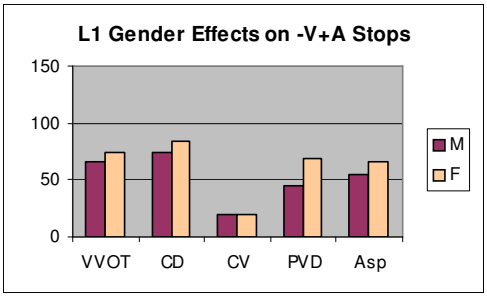
Figure 55. Relationship Between Vowel Voicing Onset Time Standard Deviation (St.Dev.) and Duration of Exposure to Bangla (As Measured by Years of Residence in Bangladesh [BD]) for: (a) Voiceless Unaspirates; (b) Voiceless Aspirates; (c) Voiced Unaspirates; (d) Voiced Aspirates; (e) Stops of All Classes



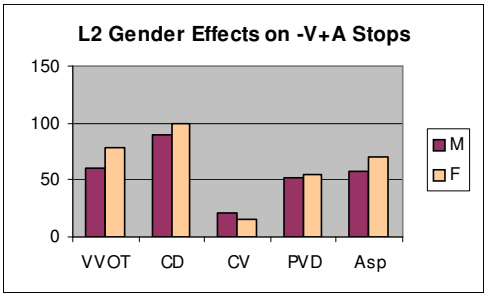
(a)



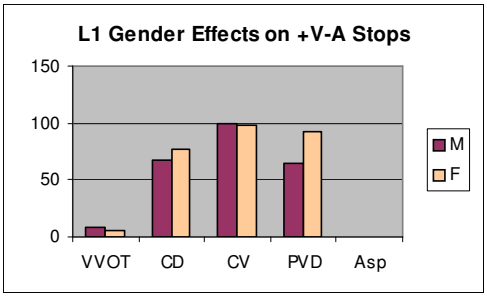
(b)



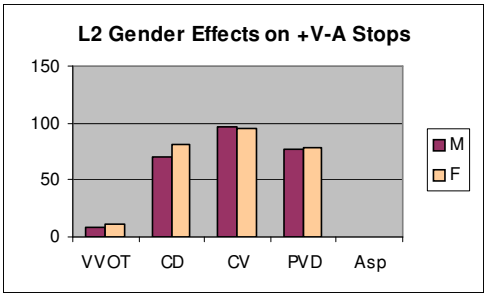
(c)



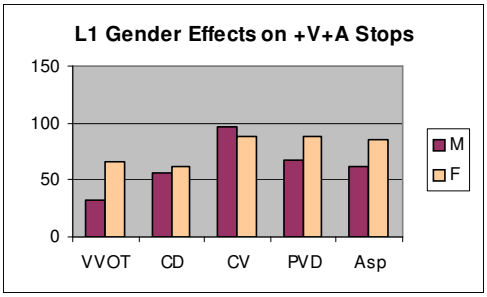
(d)



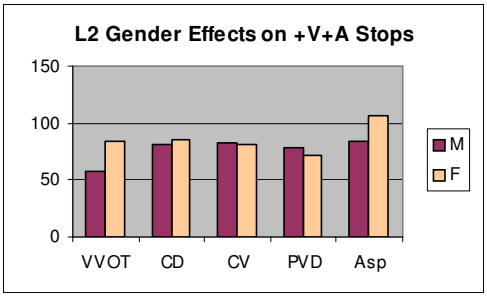
(e)



(f)



(g)



(h)

Figure 56. Gender Effects on the Means of Vowel Voicing Onset Time (VVOT) (ms), Closure Duration (CD) (ms), Closure Voicing (CV) (as % of Closure Duration), Preceding Vowel Duration (PVD) (less 50 ms), and Aspiration Noise Duration (Asp) (ms) for the: (a) L1 Voiceless Unaspirated (–V–A) Stops; (b) L2 Voiceless Unaspirated (–V–A) Stops; (c) L1 Voiceless Aspirated (–V+A) Stops; (d) L2 Voiceless Aspirated (–V+A) Stops; (e) L1 Voiced Unaspirated (+V–A) Stops; (f) L2 Voiced Unaspirated (+V–A) Stops; (g) L1 Voiced Aspirated (+V+A) Stops; and (h) L2 Voiced Aspirated (+V+A) Stops

Seven of the L2 subjects were speakers of American English, two either grew up in Canada or had a large amount of exposure to Canadian English, two were speakers of British English, and one was a speaker of South African English. Slight differences in their English accent were noticeable, primarily as the result of variation in vowel production, but in some cases in their production of word-medial *t* as well. In order to test whether these differences in accent had an effect on the five acoustic variables in this study, the L2 speakers were divided into those who spoke American English and those who did not. One-way ANOVAs were used to test for significant differences of these two groups for each of the variables in the following sample sets: all plosives, all word-medial plosives, and all word-medial coronal plosives. No significant effect was found except for aspiration noise, a result considered to be due to one subject whose mean was twice that of the next highest subject mean, which caused the mean of the group containing that subject to be skewed. Thus, it was concluded that the variation in the English accent of the L2 speakers did not have a significant effect on the results in this study.

3.2 Characterization of the Four Stop Classes of Bangla

A key area of difference between the results of the two speaker groups was in the phonetic realization of the Voice and Aspiration features. One evidence of this difference can be seen in that none of the L1 unaspirated tokens contained aspiration noise, whereas twenty-four (17%) of the L2 voiceless unaspirated tokens did. The successful L2 user must not only acquire the control of the features of Voice and Aspiration, but also that of their specific interactions represented by the distinct stop classes.

Based on the general effects of Voice and Aspiration on the five acoustic variables in this study as observed in the results of the native Bangla speakers, the prototypical characteristics of each of the four stop classes can be described as follows:

1. The –V–A stops will have the following acoustic characteristics:
    a) Absence of a periodic signal of significant duration during closure.
    b) The longest CD among the stop classes.
    c) A VVOT somewhere in the range between that of the +V–A and –V+A stop classes.
    d) A PVD somewhere in the range between that of the +V–A and –V+A stop classes.
    e) Absence of aspiration noise.
2. The –V+A stops will have the following acoustic characeristics:
    a) Absence of a periodic signal of significant duration during closure.
    b) A CD somewhere in the range between that of the +V+A and –V–A stop classes.
    c) The longest VVOT among the stop classes.
    d) The shortest PVD among the stop classes.
    e) Presence of aspiration noise.
3. The +V–A stops will have the following acoustic characteristics:
    a) Presence of a periodic signal of significant duration during closure.
    b) A CD somewhere in the range between that of the +V+A and –V–A stop classes.
    c) The shortest VVOT among the stop classes.
    d) The longest PVD among the stop classes.
    e) Absence of aspiration noise.
4. The +V+A stops will have the following acoustic characteristics:
    a) Presence of a periodic signal of significant duration during closure.
    b) The shortest CD among the stop classes.
    c) A VVOT somewhere in the range between that of the +V–A and –V+A stop classes.
    d) A PVD somewhere in the range between the +V–A and –V+A stop classes.
    e) Presence of aspiration noise.

The results of this study indicate that for the L1 speakers the characterizations mentioned above generally hold true, but that there are some exceptions that are correlated with Position in

the Word, and could possibly be caused by influences of stress or other factors. As might be expected, the productions of the L2 speakers show even less conformity to the characterizations mentioned above; however, it is interesting that in some instances, their productions do match what would be predicted, even though they are different from the productions of native speakers. This suggests that there are certain L1 phonetic production skills and/or phonological processes that the L2 speakers have not fully acquired. (See Figures 49-52 for an L1 vs. L2 comparison of summarized results for each stop class at each position in the word.)

3.3 Implications for L2 Pronunciation Acquisition

There was a rank order among the four stop classes in regards to how much the L2 productions were different from the L1 norm. The speaker group differences in their respective means for each of the five acoustic variables were examined for each stop class at each position in the word, resulting in a total of sixty different comparisons (cf. Figures 49-52). For VVOT, CD, CV, and PVD, the absolute value of the L1-L2 difference in means was divided by the L1 mean and then multiplied by 100 to obtain a value representing the percentage of L2 difference from the L1 norm. For aspiration noise, the percentage of tokens containing aspiration noise was calculated for each comparative context. The absolute value of the resulting L1-L2 difference was then multiplied by 100 to obtain a similar value as that of the other variables. For each of the stop classes all of these values were then totaled, with larger numbers representing the greatest L2 difference from the L1 norm. As shown in Figure 57, the rank order of the four stop classes in regards to how different the L2 productions are from those of the L1 is as follows, from most similar to most dissimilar: voiced unaspirated, voiceless aspirated, voiceless unaspirated, and voiced aspirated. The large value for the +V+A class is primarily due to the large value (563) for VVOT in the final position, which is most likely due to overgeneralization by some or all of the

L2 speakers. Not surprisingly, this rank order also correlates with each stop class's degree of similarity to English phonological patterns. Whether this also represents the rank order of difficulty for a native English speaker learning Bangla would depend on how important each of the acoustic correlates in this study is to an L1 speaker's perception.

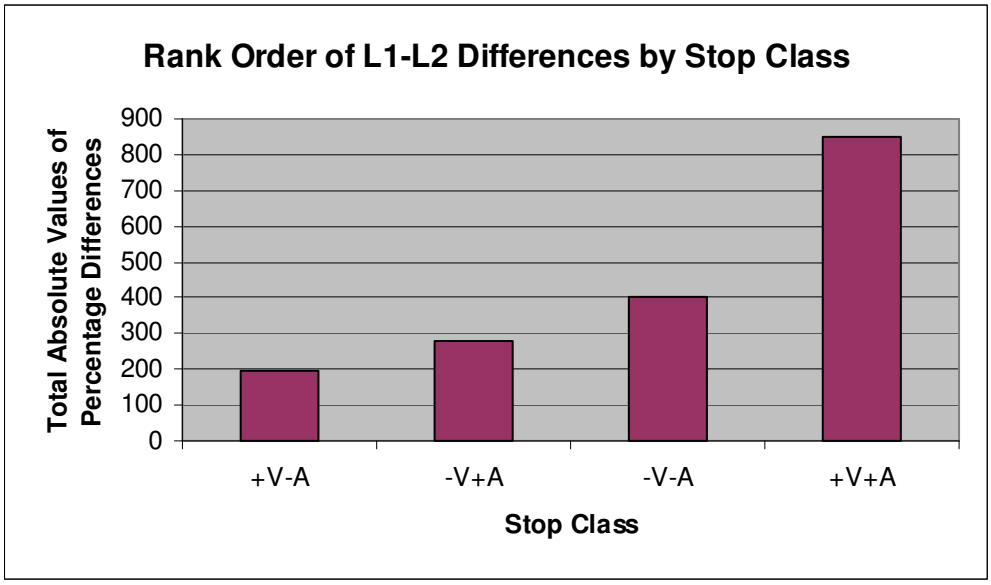


Figure 57. Rank Order of Bangla's Four Stop Classes by Degree of L2 Production Difference from the L1 Norm

Compared to the other stop classes the L2 CD and/or VVOT means for the voiced aspirates are relatively greater than those of the L1 speakers, which when added together results in a total stop duration that is significantly longer than the L1 norm. This could be because these segments are being exaggerated by the L2 speakers, possibly in an effort to make sure they pronounce them 'correctly.' In the final position the L1 VVOT mean for the +V+A stops is quite small, slightly smaller even than that of the +V–A stops, suggesting the possibility of a voiced unaspirated allophone of $b^{\hbar}$. In spite of this significant effect of Position in the Word for the L1 speakers, the corresponding L2 VVOT mean is quite large in comparison, suggesting that they are overgeneralizing in this context.

It is clear from the results in this study that the L2 speakers are aware that they are using four stop classes and have some concept as to how each of the classes is different from the others. However, it is also evident they do not control them in the same ways or with the same consistency as the L1 speakers. The following are some pertinent questions related to the successful pronunciation acquisition of these stops:

1. What is the importance of each of these acoustic correlates in achieving a 'native-like' pronunciation?
2. Does the ranking of this importance change depending on Position in the Word and Place of Articulation?
3. What is the relative importance of approximating the L1 norm for each stop class versus developing a similar distributional pattern across the classes?
4. What are practical ways in which pronunciation can be learned by current L2 users so as to bring the phonetic realization of segments closer to the L1 norm?
5. What are ways in which language teachers can present Bangla's four-way stop contrast so as to enable new language learners in working towards a more native-like pronunciation?
6. Which differences in the realization of acoustic correlates are related to transfer (positive or negative), and which are due to exaggeration or overgeneralization?
7. Which acoustic variations can be described as 'errors,' and with which phonemes do L2 users tend to make the most errors?
8. How do the acoustic patterns of the L2 group compare with general acoustic patterns for English? How do they compare with their own individual acoustic patterns for English?
9. Are there particular acoustic patterns or contexts in English that are similar to unfamiliar phonemes used contrastively by L1 speakers?

10. How do the conceptualizations of both L1 and L2 speakers for the four stop classes in Bangla relate to the acoustic realizations of their productions?
11. How do the individual L2 subjects differ from one another in their L2 productions, and what can be learned from this?
12. How do L2 learners with a more native-like pronunciation conceptualize the four stop classes?
13. What role does perception play in acquiring native-like production?
14. How can Major's Ontogeny Phylogeny Model be used to better understand the process of phonological acquisition for an L2 learner?
15. In what ways are the suprasegmental features of the speech of L2 speakers different than those of L1 speakers, and what effect does this have on the pronunciation of the various Bangla stops?

In the following section, several suggestions are given for future research that could build on the current study by addressing some of the above-mentioned questions.

3.4 Suggestions for Further Research

Future production studies investigating other acoustic variables related to these stops could be useful. These variables might include burst quality, pitch of the following vowel, measures of breathiness, and the duration and quality of the following vowel. This would help to better understand the differences in the clusters of features associated with each stop at the various positions in the word as produced by both L1 and L2 speakers. Intonation and stress patterns could also be compared to see what effect they might have on segmental pronunciation.

However, there is also a need for several different types of perceptual studies involving one or both speaker groups, which would focus on the acoustic correlates of the Voice and Aspiration features as well as the four stop classes derived from these features. Productions of

the Bangla stops by L1 speakers could be manipulated in various ways to determine how well they are perceived by native English speakers in isolation or in the context of a real word or sentence. In order to control for the recognition of a stop by a prior knowledge of the word it is in, nonce words could be used to provide sets of minimal pairs that could then appear in a sentence frame such as, "I don't know what _____ means." Perceptual studies like this would shed more light on the role that phonetic perception of segments plays in the corresponding productions of L2 speakers, especially if those who demonstrate better perception also demonstrate better production. It might also be interesting to use the stop productions of L2 subjects in an experiment testing their perception of their own pronunciation, which would provide information on whether it is possible for L2 users to produce a phonemic contrast they cannot perceive, simply by knowing the articulatory mechanics involved in producing the segments. Speech manipulation techniques and perceptual testing could also be used with both L1 and L2 stop productions and L1 subjects to see what acoustic correlates are most important to their perception, and what differences there might be between the speaker groups in regards to which acoustic cues are most salient to each of them. These results could then be used to develop methods of improving both the perception and production of the L2 speakers.

A related type of perceptual study necessary is one in which L2 stop productions are used with L1 subjects, in an investigation of the intelligibility and/or accentedness of L2 speakers. Rating scales could be used on which a subject could indicate the understandability and accentedness of the L2 speaker whose pronunciation they are listening to. These ratings could then be compared to the results of acoustic analysis for the same L2 productions in order to see which acoustic variables are the most important to a native speaker's perception. This could then provide a helpful prioritization as to which acoustic variables should be focused on when seeking to develop practical methods for more successful pronunciation acquisition. The work of Flege,

Munro, and Skelton (1992) with Spanish and English could be used as a model for this type of study.

A limitation of the methodology chosen for this study is that there is no stop production data occurring after silence. This has the effect of masking any similarities and differences in the lead voicing of stops, an acoustic variable known to be often absent in English and always present and of a relatively long duration in Bangla. Because of this difference it is assumed that native English speakers will face challenges in producing the Bangla voiced stops in this position. Production and perceptual studies involving both speaker groups and using stimuli in which these same stops are in an utterance-initial position is necessary for two reasons. First, it would provide data that would be more comparable with much of the research that has been focused on the traditional VOT measurement, including that done on other Indo-Aryan languages, which would provide the potential of contributing to a better understanding of lead voicing in these languages. Secondly, it would provide some crucial data contributing to a better understanding of the challenges native English speakers face in producing stops in this not-as-frequent but still important verbal context, as well as pointing towards possible solutions for an improved pronunciation.

Once there is a better understanding of which acoustic variables are most influential in accurate production and perception, some studies are necessary whose aim is to develop a practical way for language learners to improve their segmental pronunciation. From a cognitive standpoint, it might be useful to investigate and compare the ways in which L1 and L2 speakers conceptualize the stop classes, as well as individual stops. For instance, how would each one describe the segments, and how would they describe the productions of the other speaker group? If there are consistent conceptual differences associated with differences in acoustic variables, a follow-up study could be done on whether changing the way a person thinks about a segment will

also change their production of it. This type of study might also be useful if an L2 speaker who has excellent pronunciation is used instead of an L1 speaker. Another approach is to use basic phonetic training to help a language learner understand the mechanics behind a particular sound. With any of these types of studies, a longitudinal approach would be needed in order to test the effectiveness of the different methods. The methods that prove the most effective could then be integrated into a broader language-learning program.

Longitudinal studies of native English speakers who are just beginning to learn Bangla would likely be most beneficial over a period of 3–5 years, but shorter studies could also be useful. An example of this type of research based on the content of the current study could include the following elements. First, the same type of data (focusing on the four stop classes) could be elicited from the same language learners every three months over a period of two years, every six months in the third and fourth years, and once a year thereafter. For any of the first data elicitations for which the subjects have not yet learned to read the Bangla script, a collection of visual stimuli and a simpler sentence frame could be integrated with the initial language acquisition process and then used at the time of recording as well. It would also be good to include a short verbal interview at each elicitation in which the language learner is asked questions such as:

1. Which of these sounds are you finding easier to pronounce, and which are more difficult?
2. Which sounds are the same as English sounds, which are somewhat similar, and which are completely different?
3. Which sounds do you find the hardest to understand when a native Bangla speaker says them, and which ones have you noticed seem hard for a native Bangla speaker to understand when you say them?

4. How would you describe the voiceless unaspirated sounds? How would you describe the voiced aspirated sounds?
5. For the sounds you find most difficult, what are some of the strategies you've been using to try to improve your pronunciation? Have you found these to be helpful?
6. Do you see any system or pattern in these sixteen stops? What are some of the similarities and differences you see among them?

These types of questions could provide insights into the ways that language learners are thinking about their pronunciation acquisition of these stops along with their conceptualizations of the four-way stop system. As part of a longitudinal study, the relationship between any changes in these psycholinguistic aspects and changes in the acoustic realizations of the stop productions could then be investigated with the potential for a number of interesting results and applications in working towards a more efficient and effective approach to acquisition of Bangla pronunciation. Observations of changes in the language learner's interlanguage (including error analysis) using Major's Ontogeny Phylogeny Model as a guiding framework could also be valuable in understanding transfer issues and the processes involved in a native English speaker's acquisition of Bangla stop pronunciation. If more than one subject is used in a study such as this, it could also be useful to compare the results of the various subjects with each other in an effort to understand both the universal and variable tendencies for this particular L2 learner group, as well as to highlight particular factors that seem to contribute to more successful or less successful stop pronunciation acquisition.

Native English speakers are by no means the only expatriates who desire to learn and use the Bangla language, and depending on what their mother tongue is, they may face somewhat similar or quite different challenges in achieving a native-like production of the Bangla stops. Thus, more research focusing on native speakers of other languages than English is necessary.

The current study along with the follow-up studies mentioned above could be used as a starting point and necessary adjustments could then be made according to the language of focus. In this way, the learning and teaching of Bangla stop production and perception could be tailored to fit the needs of each mother-tongue group of language learners.

3.5 Conclusion

In this study, an introduction to key issues of L2 pronunciation acquisition was given, and a brief contrastive analysis of English and Bangla was carried out to emphasize the likelihood of native English speakers having difficulty learning the four-way stop system of Bangla, mainly due to its interacting phonemic contrasts of both Voice and Aspiration. An acoustic analysis of the stop productions of L1 speakers was done, looking at the acoustic variables of vowel voicing onset time, closure duration, closure voicing, preceding vowel duration, and aspiration noise duration, and focusing mainly on distinctions related to the four stop classes: voiceless unaspirated, voiceless aspirated, voiced unaspirated, and voiced aspirated. The effects of Position in the Word and Gender on these variables were also investigated. This analysis served two purposes. First, it supported the findings of other researchers that the two voiced stop classes are distinguished by a significant difference in mean duration of post-release vowel onset time, thus making a [breathy voice] feature unnecessary for a phonological description of the stop system of Bangla and other Indo-Aryan languages. Secondly, the L1 means provided a standard by which to evaluate the corresponding stop productions of the L2 speakers, of which a similar analysis was done. The main ways in which the L2 phonetic realizations were different from those of the L1 were discussed, and a short characterization of each of the four stop classes was presented. Both cross-language and language-specific effects of Gender were observed, and a negative correlation was found between subjects' Duration of Exposure to Bangla and the standard

deviation of their vowel voicing onset time values. A number of key questions were posed relating to the acquisition of Bangla stops by native English speakers, and suggestions were made in regards to other avenues of research that could serve to answer these questions. Thus this study contributes to a fuller understanding of the acoustic phonetic characteristics of Voice and Aspiration and their interaction in four-way stop systems such as that of Bangla and other Indo-Aryan languages. It also provides an acoustic phonetic description of both negative and positive transfer in a production context, highlighting important characteristics of L2 pronunciation acquisition. Lastly, it serves as a foundation and starting point for the development of more effective Bangla stop pronunciation acquisition strategies for native speakers of English and other languages.

APPENDICES

Appendix 1
Target Words

Stop	Word-Initial	Word-Medial	Word-Final
ক *k*	কান *kan* 'ear'	টাকা *ʈaka* 'money'	কাক *kak* 'crow'
খ *k^{h}*	খাবার *k^{h}abar* 'food'	শাখা *ʃak^{h}a* 'branch'	মুখ *muk^{h}* 'face'
গ *g*	গাড়ি *gaɽi* 'car'	বাগান *bagan* 'garden'	দাগ *ɖag* 'spot'
ঘ *g^{ɦ}*	ঘড়ি *g^{ɦ}oɽi* 'watch'	আঘাত *ag^{ɦ}aʈ* 'wound'	বাঘ *bag^{ɦ}* 'tiger'
চ *t͡ʃ*	চাকা *t͡ʃaka* 'wheel'	আচার *at͡ʃar* 'pickle'	কাচ *kat͡ʃ* 'glass'
ছ *t͡ʃ^{h}*	ছাতা *t͡ʃ^{h}aʈa* 'umbrella'	গাছা *gat͡ʃ^{h}a* 'lampstand'	গাছ *gat͡ʃ^{h}* 'tree'
জ *d͡ʒ*	জামা *d͡ʒama* 'shirt'	রাজা *ɾad͡ʒa* 'king'	কাজ *kad͡ʒ* 'work'
ঝ *d͡ʒ^{ɦ}*	ঝাড়ু *d͡ʒ^{ɦ}aɽu* 'broom'	মাঝি *mad͡ʒ^{ɦ}i* 'boatman'	বুঝ *bud͡ʒ^{ɦ}* 'understanding'
ট *ʈ*	টাকা *ʈaka* 'money'	আটা *aʈa* 'flour'	ঘাট *g^{ɦ}aʈ* 'dock'
ঠ *ʈ^{h}*	ঠাণ্ডা *ʈ^{h}anɖa* 'cold'	আঠা *aʈ^{h}a* 'glue'	কাঠ *kaʈ^{h}* 'wood'
ড *ɖ*	ডানা *ɖana* 'wing'	সোডা *soɖa* 'soda'	রড *ɾɔɖ* 'rod'
ঢ *ɖ^{ɦ}*	ঢাকনা *ɖ^{ɦ}akna* 'cover'	অঢেল *ɔɖ^{ɦ}el* 'abundant'	N/A
ত *t̪*	তারা *t̪ara* 'star'	বাতাস *bat̪af* 'air'	ভাত *b^{ɦ}at̪* 'rice'
থ *t̪^{h}*	থালা *t̪^{h}ala* 'plate'	মাথা *mat̪^{h}a* 'head'	রথ *ɾɔt̪^{h}* 'chariot'
দ *d̪*	দাড়ি *d̪aɽi* 'beard'	বাদাম *bad̪am* 'nut'	ছাদ *t͡ʃ^{h}ad̪* 'roof'
ধ *d̪^{ɦ}*	ধারা *d̪^{ɦ}ara* 'stream'	গাধা *gad̪^{ɦ}a* 'donkey'	সাধ *ʃad̪^{ɦ}* 'desire'
প *p*	পাতা *paʈa* 'leaf'	আপা *apa* 'big sister'	ধাপ *d̪^{ɦ}ap* 'step'
ফ *p^{h}*	ফাঁকা *p^{h}ãka* 'gap'	তফাত *t̪ɔp^{h}at̪* 'difference'	লাফ *lap^{h}* 'jump'
ব *b*	বাসা *baʃa* 'house'	থাবা *t̪^{h}aba* 'paw'	ডাব *ɖab* 'green coconut'
ভ *b^{ɦ}*	ভাষা *b^{ɦ}aʃa* 'language'	সভা *ʃɔb^{ɦ}a* 'meeting'	লাভ *lab^{ɦ}* 'profit'

Appendix 2

Stimulus Sentences

Stop	Stimulus (target word in bold)	Stop	Stimulus (target word in bold)
ক Init. *k*	ওখানে একটা **কান** আছে। *okʰane ɛkʈa* ***kan*** *at͡ʃʰe* 'There is an **ear** over there.'	ড Init. *ɖ*	ওখানে একটা **ডানা** আছে। *okʰane ɛkʈa* ***ɖana*** *at͡ʃʰe* 'There is a **wing** over there.'
ক Med. *k*	ওখানে কিছু **টাকা** আছে। *okʰane kit͡ʃʰu* ***ʈaka*** *at͡ʃʰe* 'There is some **money** over there.'	ড Med. *ɖ*	ওখানে কিছু **সোডা** আছে। *okʰane kit͡ʃʰu* ***soɖa*** *at͡ʃʰe* 'There is some **soda** over there.'
ক Fin. *k*	ওখানে একটা **কাক** আছে। *okʰane ɛkʈa* ***kak*** *at͡ʃʰe* 'There is a **crow** over there.'	ড Fin. *ɖ*	ওখানে একটা **রড** আছে। *okʰane ɛkʈa* ***rɔɖ*** *at͡ʃʰe* 'There is a **rod** over there.'
খ Init. *kʰ*	ওখানে কিছু **খাবার** আছে। *okʰane kit͡ʃu* ***kʰabar*** *at͡ʃʰe* 'There is some **food** over there.'	ঢ Init. *ɖʱ*	ওখানে একটা **ঢাকনা** আছে। *okʰane ɛkʈa* ***ɖʱakna*** *at͡ʃʰe* 'There is a **cover** over there.'
খ Med. *kʰ*	ওখানে একটা **শাখা** আছে। *okʰane ɛkʈa* ***ʃakʰa*** *at͡ʃʰe* 'There is a **branch** over there.'	ঢ Med. *ɖʱ*	ওখানে **অঢেল** আনন্দ আছে। *okʰane* ***ɔɖʱel*** *anɔnɖo at͡ʃʰe* 'There is **abundant** joy there.'
খ Fin. *kʰ*	ওখানে একটা **মুখ** আছে। *okʰane ɛkʈa* ***mukʰ*** *at͡ʃʰe* 'There is a **face** over there.'	ঢ Fin. *ɖʱ*	N/A
গ Init. *ɡ*	ওখানে একটা **গাড়ি** আছে। *okʰane ɛkʈa* ***gaɽi*** *at͡ʃʰe* 'There is a **car** over there.'	ত Init. *t̪*	ওখানে একটা **তারা** আছে। *okʰane ɛkʈa* ***t̪ara*** *at͡ʃʰe* 'There is a **star** over there.'
গ Med. *ɡ*	ওখানে একটা **বাগান** আছে। *okʰane ɛkʈa* ***bagan*** *at͡ʃʰe* 'There is a **garden** over there.'	ত Med. *t̪*	ওখানে কিছু **বাতাস** আছে। *okʰane kit͡ʃʰu* ***bat̪aʃ*** *at͡ʃʰe* 'There is some **air** over there.'
গ Fin. *ɡ*	ওখানে একটা **দাগ** আছে। *okʰane ɛkʈa* ***d̪ag*** *at͡ʃʰe* 'There is a **spot** over there.'	ত Fin. *t̪*	ওখানে কিছু **ভাত** আছে। *okʰane kit͡ʃʰu* ***bʱat̪*** *at͡ʃʰe* 'There is some **rice** over there.'

Stop	Stimulus (target word in bold)	Stop	Stimulus (target word in bold)
ঘ Init. $g^{ɦ}$	ওখানে একটা **ঘড়ি** আছে। *ok^{h}ane ɛkʈa* ***gˠ^{ɦ}oʈi*** *atʃ^{h}e* 'There is a **watch** over there.'	থ Init. $ʈ^{h}$	ওখানে একটা **থালা** আছে। *ok^{h}ane ɛkʈa* ***ʈ^{h}ala*** *atʃ^{h}e* 'There is a **plate** over there.'
ঘ Med. $g^{ɦ}$	ওখানে একটা **আঘাত** আছে। *ok^{h}ane ɛkʈa* ***ag^{ɦ}aʈ*** *atʃ^{h}e* 'There is a **wound** there.'	থ Med. $ʈ^{h}$	ওখানে একটা **মাথা** আছে। *ok^{h}ane ɛkʈa* ***maʈ^{h}a*** *atʃ^{h}e* 'There is a **head** over there.'
ঘ Fin. $g^{ɦ}$	ওখানে একটা **বাঘ** আছে। *ok^{h}ane ɛkʈa* ***bag^{ɦ}*** *atʃ^{h}e* 'There is a **tiger** over there.'	থ Fin. $ʈ^{h}$	ওখানে একটা **রথ** আছে। *ok^{h}ane ɛkʈa* ***rɔʈ^{h}*** *atʃ^{h}e* 'There is a **chariot** over there.'
চ Init. $\widehat{tʃ}$	ওখানে একটা **চাকা** আছে। *ok^{h}ane ɛkʈa* ***ʧaka*** *atʃ^{h}e* 'There is a **wheel** over there.'	দ Init. $ɖ$	ওখানে একটা **দাড়ি** আছে। *ok^{h}ane ɛkʈa* ***ɖaʈi*** *atʃ^{h}e* 'There is a **beard** over there.'
চ Med. $\widehat{tʃ}$	ওখানে কিছু **আচার** আছে। *ok^{h}ane kiʧ^{h}u* ***aʧar*** *atʃ^{h}e* 'There is some **pickle** over there.'	দ Med. $ɖ$	ওখানে একটা **বাদাম** আছে। *ok^{h}ane ɛkʈa* ***baɖam*** *atʃ^{h}e* 'There is a **nut** over there.'
চ Fin. $\widehat{tʃ}$	ওখানে কিছু **কাচ** আছে। *ok^{h}ane kiʧ^{h}u* ***kaʧ*** *atʃ^{h}e* 'There is some **glass** over there.'	দ Fin. $ɖ$	ওখানে একটা **ছাদ** আছে। *ok^{h}ane ɛkʈa* ***ʧ^{h}aɖ*** *atʃ^{h}e* 'There is a **roof** over there.'
ছ Init. $\widehat{tʃ^{h}}$	ওখানে একটা **ছাতা** আছে। *ok^{h}ane ɛkʈa* ***ʧ^{h}aʈa*** *atʃ^{h}e* 'There is an **umbrella** over there.'	ধ Init. $ɖ^{ɦ}$	ওখানে একটা **ধারা** আছে। *ok^{h}ane ɛkʈa* ***ɖ^{ɦ}ara*** *atʃ^{h}e* 'There is a **stream** over there.'
ছ Med. $\widehat{tʃ^{h}}$	ওখানে একটা **গাছা** আছে। *ok^{h}ane ɛkʈa* ***gaʧ^{h}a*** *atʃ^{h}e* 'There is a **lampstand** over there.'	ধ Med. $ɖ^{ɦ}$	ওখানে একটা **গাধা** আছে। *ok^{h}ane ɛkʈa* ***gaɖ^{ɦ}a*** *atʃ^{h}e* 'There is a **donkey** over there.'
ছ Fin. $\widehat{tʃ^{h}}$	ওখানে একটা **গাছ** আছে। *ok^{h}ane ɛkʈa* ***gaʧ^{h}*** *atʃ^{h}e* 'There is a **tree** over there.'	ধ Fin. $ɖ^{ɦ}$	আমার একটা **সাধ** আছে। *amar ɛkʈa* ***ʃaɖ^{ɦ}*** *atʃ^{h}e* 'I have a **desire**.'
জ Init. $\widehat{dʒ}$	ওখানে একটা **জামা** আছে। *ok^{h}ane ɛkʈa* ***ʤama*** *atʃ^{h}e* 'There is a **shirt** over there.'	প Init. p	ওখানে একটা **পাতা** আছে। *ok^{h}ane ɛkʈa* ***paʈa*** *atʃ^{h}e* 'There is a **leaf** over there.'
জ Med. $\widehat{dʒ}$	ওখানে একটি **রাজা** আছে। *ok^{h}ane ɛkʈi* ***raʤa*** *atʃ^{h}e* 'There is a **king** over there.'	প Med. p	ওখানে আমার একটা **আপা** আছে। *ok^{h}ane amar ɛkʈa* ***apa*** *atʃ^{h}e* 'One of my **big sisters** is there.'

Stop	Stimulus (target word in bold)	Stop	Stimulus (target word in bold)
জ Fin. *d͡ʒ*	ওখানে একটা **কাজ** আছে। *okʰane ɛkʈa* ***kad͡ʒ*** *at͡ʃʰe* ‘There is a **job** over there.’	প Fin. *p*	ওখানে একটা **ধাপ** আছে। *okʰane ɛkʈa* ***d̪ʱap*** *at͡ʃʰe* ‘There is a **step** over there.’
ঝ Init. *d͡ʒʱ*	ওখানে একটা **ঝাড়ু** আছে। *okʰane ɛkʈa* ***d͡ʒʱaɽu*** *at͡ʃʰe* ‘There is a **broom** over there.’	ফ Init. *pʰ*	ওখানে একটা **ফাঁকা** আছে। *okʰane ɛkʈa* ***pʰãka*** *at͡ʃʰe* ‘There is a **gap** there.’
ঝ Med. *d͡ʒʱ*	ওখানে একটি **মাঝি** আছে। *okʰane ɛkʈi* ***mad͡ʒʱi*** *at͡ʃʰe* ‘There is a **boatman** over there.’	ফ Med. *pʰ*	ওখানে একটা **তফাত** আছে। *okʰane ɛkʈa* ***ʈɔpʰaʈ*** *at͡ʃʰe* ‘There is a **difference** there.’
ঝ Fin. *d͡ʒʱ*	তার একটা **বুঝ** আছে। *ʈar ɛkʈa* ***budʒʱ*** *at͡ʃʰe* ‘He has an **idea**.’	ফ Fin. *pʰ*	ওখানে একটা **লাফ** আমি দিয়েছি। *okʰane ɛkʈa* ***lapʰ*** *ami d̪ijet͡ʃʰi* ‘I **jumped** over there.’
ট Init. *ʈ*	ওখানে কিছু **টাকা** আছে। *okʰane kit͡ʃʰu* ***ʈaka*** *at͡ʃʰe* ‘There is some **money** over there.’	ব Init. *b*	ওখানে একটা **বাসা** আছে। *okʰane ɛkʈa* ***baʃa*** *at͡ʃʰe* ‘There is a **house** over there.’
ট Med. *ʈ*	ওখানে কিছু **আটা** আছে। *okʰane kit͡ʃʰu* ***aʈa*** *at͡ʃʰe* ‘There is some **flour** over there.’	ব Med. *b*	ওখানে একটা **থাবা** আছে। *okʰane ɛkʈa* ***ʈʰaba*** *at͡ʃʰe* ‘There is a **paw** over there.’
ট Fin. *ʈ*	ওখানে একটা **ঘাট** আছে। *okʰane ɛkʈa* ***gʱaʈ*** *at͡ʃʰe* ‘There is a **dock** over there.’	ব Fin. *b*	ওখানে একটা **ডাব** আছে। *okʰane ɛkʈa* ***ɖab*** *at͡ʃʰe* ‘There is a **green coconut** over there.’
ঠ Init. *ʈʰ*	ওখানে কিছু **ঠাণ্ডা** আছে। *okʰane kit͡ʃʰu* ***ʈʰanɖa*** *at͡ʃʰe* ‘There is some **cold** [weather] there.’	ভ Init. *bʱ*	ওখানে একটা **ভাষা** আছে। *okʰane ɛkʈa* ***bʱaʃa*** *at͡ʃʰe* ‘There is a **language** there.’
ঠ Med. *ʈʰ*	ওখানে কিছু **আঠা** আছে। *okʰane kit͡ʃʰu* ***aʈʰa*** *at͡ʃʰe* ‘There is some **glue** over there.’	ভ Med. *bʱ*	ওখানে একটা **সভা** আছে। *okʰane ɛkʈa* ***ʃɔbʱa*** *at͡ʃʰe* ‘There is a **meeting** over there.’
ঠ Fin. *ʈʰ*	ওখানে কিছু **কাঠ** আছে। *okʰane kit͡ʃʰu* ***kaʈʰ*** *at͡ʃʰe* ‘There is some **wood** over there.’	ভ Fin. *bʱ*	ওখানে কিছু **লাভ** আছে। *okʰane kit͡ʃʰu* ***labʱ*** *at͡ʃʰe* ‘There is some **profit** over there.’

Appendix 3

Subject Information (Gender, Age, Education, Languages)

Subject	Gender	Age	Education	Mother Tongue	Dialect?	Other Languages
S01	M	37	Masters	Bangla	N/A	English, Hindi
S03	M	28	Masters in progress	Bangla	N/A	Some English, some Hindi
S04	M	23	Higher secondary school in progress	Bangla	N/A	Some English, a little Hindi
S05	M	23	Higher secondary school in progress	Bangla	N/A	Some English
S06	M	28	Masters in progress	Bangla	N/A	English
S07	M	38	Some higher secondary school	Bangla	N/A	Some English
S21	F	61	Some higher secondary school	Bangla	N/A	Some English, a little Hindi
S22	F	30	Masters in progress	Bangla	N/A	English, some Hindi
S23	F	37	Undergraduate degree	Bangla	N/A	Some English, some Hindi and Urdu
S24	F	31	Undergraduate degrees	Bangla	N/A	English
S25	F	34	Undergraduate degree	Bangla	N/A	English
S26	F	44	Higher secondary school	Bangla	N/A	English, a little Hindi
S40	M	31	Masters (Linguistics) in progress	English	N/A	
S41	M	35	Masters degree	English	N/A	Some Arabic
S42	M	52	Undergraduate degree	English	N/A	
S43	M	22	Higher secondary school	English	N/A	
S44	M	29	Some undergraduate study	German dialect; bilingual in English	N/A	Some Spanish
S45	M	39	Masters degree	English	N/A	

Subject	Gender	Age	Education	Mother Tongue	Dialect?	Other Languages
S61	F	44	Undergraduate degree	English	N/A	Some Spanish
S62	F	53	Undergraduate degree	English	N/A	Some Spanish
S63	F	25	Higher secondary school	English	N/A	
S64	F	26	Masters degree	English	(Southern England)	Some French, some Spanish
S65	F	32	Masters degree	English	N/A	Dutch
S66	F	28	Undergraduate degrees	English	N/A	Afrikaans, Xhosa, Tswana

Subject Information (Work, Residence, Other Comments)

Subject	Profession	Current Occupation	Current Residence* (years/months)	Residence Before Age 21	Comments
S01	'Private service'	NGO Worker	(10/0)	Kushtia, Bangladesh	Brother of S03, husband of S22
S03	Student	Student	(1/6)	Kushtia, Bangladesh	Brother of S01
S04	Student	Student	(4/0)	Kushtia, Bangladesh	
S05	Office assistant	Office assistant	(15/0)	Norshingdi, Bangladesh	
S06	Student	Teaching	(7/0)	Faridpur, Bangladesh	
S07	Electronics	Electronics	(32/0)	Gopalganj District, Bangladesh	Husband of S25
S21	Housewife	Housewife	(61/0)	Dhaka, Bangladesh	
S22	Teaching	Teaching	(9/0)	Kushtia, Bangladesh	Wife of S01
S23	'Private service'	'Private Service'	(17/0)	Gazipur, Bangladesh	Sister of S24
S24	Housewife	Housewife / Tutor	(10–14/0)	Gazipur, Bangladesh	Sister of S23
S25	Office assistant	Tutoring	(34/0)	Dhaka, Bangladesh	Wife of S07
S26	Housewife	Housewife	(3/0)	Faridpur, Bangladesh	
S40	Development	Development	(7/0)	United States; Canada	Researcher
S41	Engineer	Treasurer	(1/6)	Africa	Husband of S65
S42	Engineer	Engineer	Kushtia, Bangladesh (4/0)	United States	Husband of S62
S43	Teacher	Student	(0/8)	United States	Husband of S63
S44	IT / Software	CEO (IT business)	(0/8)	United States	
S45	Development	Development	(5/6)	England	
S61	Housewife	Housewife	(5/0)	United States	
S62	Registered nurse	Registered nurse	Kushtia, Bangladesh (4/0)	United States	Wife of S42
S63	Teacher	Student	(0/8)	United States	Wife of S43
S64	English teacher	English teacher	(1/2)	England	
S65	Physiotherapist	Physiotherapist	(0/10.5)	Canada; Indonesia	Wife of S41
S66	Teacher	Teacher	(1/7)	South Africa	

*Current residence is Dhaka, Bangladesh unless otherwise noted.

Subject Information (Language Learning Experience of Native English Speakers)

Subject	Formal Bangla Classes	Pronunciation Training	Time In Bangladesh	Total Bangla Learning
S40	1 year	2-month Articulatory Phonetics course	7 years	2 years
S41	Less than 6 months		5 years	1 year
S42	5 months	1 day or so out of 2-week language acquisition training	6 years	9 months
S43	6 months		8 months	8 months
S44	5 months	10–15 hrs of 2-week language acquisition training	8 months	6 months
S45	5 months		6 years	1.5 years
S61	5 months	Part of 2-week language acquisition training	18 years	1.5 years
S62	5 months	Some comprehension-based activities for undergraduate Spanish; 1 day of 2-week language acquisition training	6 years	About 3 years
S63	6 months		8 months	8.5 months
S64	1 year	A little in various language classes taken, some in Bangla as part of singing lessons	3.5 years	1 year, 2 months
S65	6 months	1 day of 2-week language acquisition training	10.5 months	11.5 months
S66	6 months		1 year, 7 months (plus 3 months four years earlier)	9 months

Appendix 4

List of Stimulus Sentences

1. ওখানে একটা বাজার আছে।*
2. ওখানে একটা চাকা আছে।
3. ওখানে একটা দাগ আছে।
4. ওখানে কিছু খাবার আছে।
5. ওখানে একটা গাছা আছে।
6. তার একটা বুঝ আছে।
7. ওখানে একটা বাদাম আছে।
8. ওখানে কিছু বাতাস আছে।
9. ওখানে আমার একটা আপা আছে।
10. ওখানে কিছু ঠাণ্ডা আছে।
11. ওখানে একটা ঘড়ি আছে।
12. ওখানে একটা কান আছে।
13. ওখানে একটা ঝাড়ু আছে।
14. ওখানে কিছু সোডা আছে।
15. আমার একটা সাধ আছে।
16. ওখানে একটা বাসা আছে।
17. ওখানে কিছু ভাত আছে।
18. ওখানে কিছু লাভ আছে।
19. ওখানে একটা ফাঁকা আছে।
20. ওখানে একটা রথ আছে।
21. ওখানে একটা ডাব আছে।
22. ওখানে একটা ঢাকনা আছে।
23. ওখানে একটি রাজা আছে।
24. ওখানে একটা বাঘ আছে।
25. ওখানে একটা শাখা আছে।
26. ওখানে কিছু কাচ আছে।
27. ওখানে একটা ঘাট আছে।
28. ওখানে কিছু কাঠ আছে।
29. ওখানে একটা থালা আছে।
30. ওখানে একটা ধারা আছে।
31. ওখানে একটা লাফ আমি দিয়েছি।
32. ওখানে অঢেল আনন্দ আছে।
33. ওখানে কিছু আটা আছে।
34. ওখানে একটা ছাতা আছে।
35. ওখানে একটা গাড়ি আছে।
36. ওখানে একটা কাজ আছে।
37. ওখানে একটা ডানা আছে।
38. ওখানে একটা গাধা আছে।
39. ওখানে একটা সভা আছে।
40. ওখানে একটা ধাপ আছে।
41. ওখানে একটা রড আছে।
42. ওখানে একটা গাছ আছে।
43. ওখানে কিছু টাকা আছে।
44. ওখানে একটা বাগান আছে।
45. ওখানে একটি মাঝি আছে।
46. ওখানে একটা মাথা আছে।
47. ওখানে একটা পাতা আছে।
48. ওখানে একটা তফাত আছে।
49. ওখানে একটা মুখ আছে।
50. ওখানে কিছু আচার আছে।
51. ওখানে কিছু আঠা আছে।
52. ওখানে একটা দাড়ি আছে।
53. ওখানে একটা কাক আছে।
54. ওখানে একটা ছাদ আছে।
55. ওখানে একটা ভাষা আছে।
56. ওখানে একটা তারা আছে।
57. ওখানে একটা আঘাত আছে।
58. ওখানে একটা জামা আছে।
59. ওখানে একটা থাবা আছে।
60. ওখানে একটা বাসা আছে।

*This line represents the approximate size of the text used for data elicitation. A different font, smaller font size and a few minor layout changes were made to the original list for presentation here.

List of Stimulus Sentences (Reversed)

1. ওখানে একটা বাসা আছে।*
2. ওখানে একটা থাবা আছে।
3. ওখানে একটা জামা আছে।
4. ওখানে একটা আঘাত আছে।
5. ওখানে একটা তারা আছে।
6. ওখানে একটা ভাষা আছে।
7. ওখানে একটা ছাদ আছে।
8. ওখানে একটা কাক আছে।
9. ওখানে একটা দাড়ি আছে।
10. ওখানে কিছু আঠা আছে।
11. ওখানে কিছু আচার আছে।
12. ওখানে একটা মুখ আছে।
13. ওখানে একটা তফাত আছে।
14. ওখানে একটা পাতা আছে।
15. ওখানে একটা মাথা আছে।
16. ওখানে একটি মাঝি আছে।
17. ওখানে একটা বাগান আছে।
18. ওখানে কিছু টাকা আছে।
19. ওখানে একটা গাছ আছে।
20. ওখানে একটা রড আছে।
21. ওখানে একটা ধাপ আছে।
22. ওখানে একটা সভা আছে।
23. ওখানে একটা গাধা আছে।
24. ওখানে একটা ডানা আছে।
25. ওখানে একটা কাজ আছে।
26. ওখানে একটা গাড়ি আছে।
27. ওখানে একটা ছাতা আছে।
28. ওখানে কিছু আটা আছে।
29. ওখানে অঢেল আনন্দ আছে।
30. ওখানে একটা লাফ আমি দিয়েছি।
31. ওখানে একটা ধারা আছে।
32. ওখানে একটা থালা আছে।
33. ওখানে কিছু কাঠ আছে।
34. ওখানে একটা ঘাট আছে।
35. ওখানে কিছু কাচ আছে।
36. ওখানে একটা শাখা আছে।
37. ওখানে একটা বাঘ আছে।
38. ওখানে একটি রাজা আছে।
39. ওখানে একটা ঢাকনা আছে।
40. ওখানে একটা ডাব আছে।
41. ওখানে একটা রথ আছে।
42. ওখানে একটা ফাঁকা আছে।
43. ওখানে কিছু লাভ আছে।
44. ওখানে কিছু ভাত আছে।
45. ওখানে একটা বাসা আছে।
46. আমার একটা সাধ আছে।
47. ওখানে কিছু সোডা আছে।
48. ওখানে একটা ঝাড়ু আছে।
49. ওখানে একটা কান আছে।
50. ওখানে একটা ঘড়ি আছে।
51. ওখানে কিছু ঠাণ্ডা আছে।
52. ওখানে আমার একটা আপা আছে।
53. ওখানে কিছু বাতাস আছে।
54. ওখানে একটা বাদাম আছে।
55. তার একটা বুঝ আছে।
56. ওখানে একটা গাছা আছে।
57. ওখানে কিছু খাবার আছে।
58. ওখানে একটা দাগ আছে।
59. ওখানে একটা চাকা আছে।
60. ওখানে একটা বাজার আছে।

*This line represents the approximate size of the text used for data elicitation. A different font, smaller font size and a few minor layout changes were made to the original list for presentation here.

Appendix 5

Consent Forms (English and Bangla)

CONSENT FORM

FOR TAKING PART IN A STUDY OF LANGUAGE

You are invited to take part in a study that seeks to find out differences in Bengali pronunciation between native English speaking language learners of Bengali and native Bengali speakers themselves. It is hoped that knowing more about these differences will help language learners to produce more natural Bengali pronunciation.

The person who is doing this study is Dustin Miller, who is a graduate student in the linguistics program of the University of North Dakota in Grand Forks, North Dakota, USA. He is doing this study to get some information he needs to write a research paper. He also hopes that this study can eventually be used to help language learners of Bengali improve their pronunciation.

If you choose to take part in this study, you will be given a list of Bengali sentences to read. You will be given time to read over them and become familiar with them. When you are ready to begin, a recording will be made of you reading these sentences aloud. This will probably take from thirty minutes to two hours. If you make a mistake during the recording you will pause and then read that sentence again before going on to the next one. When the person who is doing this study has listened to the recording and it sounds okay, your part in this study will be complete. If, in the future, this recording is lost or damaged in some way, you may be contacted again to see if you would be willing to make another recording. Otherwise, this will be the end of your part in this study.

Because of the kind of study this is, you may be asked to take part several times so that there will be a good recording. It is hoped that you will not mind if this happens.

This recording will be kept for future use by other people who may want to do studies of language. So, there is a very slight possibility that someone who listens to it will be able to recognize your voice and know that it was you reading these sentences. This may cause you embarrassment if you feel that your pronunciation or way of speaking while reading these sentences was not good in some way. However, the chance of this happening is very small.

If you are a language learner of Bengali, this study may lead to useful ways of improving your Bengali pronunciation. This will help you to communicate better in Bengali. If you are a native speaker of Bengali this study will not really help you in any way, except that it may help you to better understand the Bengali of native English speaking foreigners who have been helped by this study.

Any information that is received through this study and that can be identified with you will remain secret and will be shown to others only with your permission. At the end of this study the research data and the consent forms will be stored in separate locked locations for at least three years. The research data will be

stored in the house of the person doing this study, which is currently in Bangladesh. The consent forms will be stored in the house of the parents of the person doing this study, which is currently in Ohio, USA. Only the researcher, the adviser, and people who audit IRB procedures will have access to the data. Three years from the completion of this study all research data that can be identified with you, along with the consent forms, may be destroyed by shredding or burning. The above statements are not meant to refer to the audio recording of you reading the Bengali sentences, which may be made available in the future to other language researchers, especially for more study in learning another language. While it is possible that someone you know would be able to recognize your voice if they listened to this recording, the chance and consequences of this happening are small.

You may have certain costs because of taking part in this study, such as transportation costs. While your help with this study is meant to be voluntary, your transportation costs will be given to you if you ask for them, up to the amount of 50 taka ($0.85). If you will have costs above this amount that make it difficult for you to take part in this study, please tell the person who is doing the study to see if anything else can be done.

Your decision whether or not to take part in this study will not change your future relationship with the person doing the study or the people working with him. If you decide to take part, you are free to stop at any time without it being held against you. To stop taking part in this study, please tell the person doing the study that you want to stop, and then you will be free to leave whenever you want to.

The person doing this study is available to answer any questions you may have concerning it. Also, you are encouraged to ask any questions concerning this study that you may have in the future. Questions may be asked by calling Dustin Miller (when in Bangladesh) at 02-812-1386 (office) or 0171-563-492 (mobile) or (when in the U.S.) at 330-674-2497. You may also reach Dustin Miller by email at dustbiz@softhome.net . You may also contact Dustin's Adviser, Joan Baart, at University of North Dakota, Grand Forks, ND 58202 USA. If you have any other questions or concerns, please call the University of North Dakota Office of Research and Program Development (in the U.S.) at 701-777-4279.

If you want to be told about the results of this study, please give a written request to Dustin Miller. You may deliver this in person, use email (dustbiz@softhome.net) or send it by mail to 2/22 Block B Lalmatia, Dhaka 1207 Bangladesh. If he receives your written request Dustin will do his best to tell you about the results once the study is complete.

If you choose to take part in this study, you will be asked to sign two of these consent forms, one of which will be given to you.

If you are hurt while the study is being done, medical treatment will be as available as it is to a member of the general public in similar circumstances. Payment for any such treatment must be provided by you and your third party payer, if any (such as health insurance, Medicare, and so forth).

ALL OF MY QUESTIONS ABOUT THIS STUDY HAVE BEEN ANSWERED AND I FEEL FREE TO ASK ANY QUESTIONS THAT I MAY HAVE CONCERNING THIS STUDY IN THE FUTURE.

I have read all of the above and willingly agree to participate in this study explained to me by _____.

Subject's Signature_____

Date_____

সম্মতি-পত্র

একটি ভাষা বিষয়ক অধ্যয়নে অংশগ্রহণের জন্য

সহজাত বাংলাভাষী এবং সহজাত ইংরেজীভাষী বাংলা শিক্ষার্থীদের মধ্যেকার বাংলা উচ্চারণের পার্থক্য অনুসন্ধানের লক্ষ্যে পরিচালিত একটি অধ্যয়নে অংশগ্রহণের জন্য আপনি আমন্ত্রিত। আশা করা হচ্ছে যে, এই পার্থক্য সম্পর্কে আরো বেশী অবগতি ভাষা শিক্ষার্থীদেরকে আরো ভালভাবে স্বাভাবিক বাংলা উচ্চারণে সাহায্য করবে।

যে ব্যক্তি এই অধ্যয়ন করছেন তাঁর নাম ডাস্টিন মিলার (Dustin Miller), যিনি আমেরিকার নর্থ ডাকোটা রাজ্যের গ্রান্ড ফোর্কস-এ নর্থ ডাকোটা বিশ্ববিদ্যালয়ের ভাষা কার্যক্রম (linguistics program) এর একজন স্নাতক ছাত্র। তাঁর একটি গবেষণাপত্র লেখার জন্য কিছু তথ্য সংগ্রহের প্রয়োজনে তিনি এই অধ্যয়নটি করছেন। তিনি অবশ্য এটাও আশা করেন যে, পরিশেষে বাংলা শিক্ষার্থীদের উচ্চারণ উন্নয়নের সাহায্যের জন্য এই অধ্যয়নটি ব্যবহৃত হবে।

আপনি যদি এই অধ্যয়নে অংশগ্রহণ করতে চান, তবে বাংলা বাক্যের একটি তালিকা আপনাকে পড়তে দেওয়া হবে। বাক্যগুলি পড়া এবং সেগুলির সঙ্গে পরিচিত হওয়ার জন্য আপনাকে সময় দেওয়া হবে। আপনি যখন শুরু করার জন্য প্রস্তুত হবেন, তখন এই বাক্যগুলি আপনাকে দিয়ে শব্দ করে পড়িয়ে সেটা রেকর্ড করা হবে। এই কাজে সম্ভবত ত্রিশ মিনিট থেকে দু'ঘণ্টা সময় লাগবে। রেকর্ড করার সময় আপনি যদি কোন ভুল করে ফেলেন, তবে তখন আপনি থামবেন এবং পরবর্তী বাক্যটি পড়ার আগে সেই বাক্যটি আবার পড়বেন। যে ব্যক্তি এই অধ্যয়নটি করছেন, এই রেকর্ডিং শুনে যদি তাঁর কাছে ভাল লাগে, তবে তখন এই অধ্যয়নে আপনার কাজ সম্পূর্ণ হবে। ভবিষ্যতে যদি এই রেকর্ডিংটি হারিয়ে যায় বা কোনোভাবে ক্ষতিগ্রস্ত হয়, তবে আরেকবার এই রেকর্ডিং করাতে আপনি ইচ্ছুক কিনা সেটা জানার জন্য আপনার সঙ্গে যোগাযোগ করা হতে পারে। আর তা নাহলে, এই অধ্যয়নে আপনার ভূমিকা সেখানেই শেষ হয়ে যাবে।

কাজের ধরণটি এই রকম হওয়ার কারণে, যাতে একটি সুন্দর রেকর্ডিং হয় সেজন্য এই কাজে আপনাকে কয়েকবার ডাকা হতে পারে। এরকম যদি ঘটে, আশা করি আপনি তাতে কিছু মনে করবেন না।

ভবিষ্যতে ভাষা অধ্যয়ন করতে পারেন কেবল এমন ব্যক্তিদের ব্যবহারের জন্যই এই রেকর্ডিং সংরক্ষণ করা হবে। সুতরাং, কেউ এটা শুনে আপনার কণ্ঠ চিনে ফেলবে এবং আপনি যে এই বাক্যগুলি পড়ছেন সেটা জানতে পারবে, এমন সম্ভাবনা খুবই ক্ষীণ। আপনার যদি মনে হয় যে, এই বাক্যগুলি পড়ার সময় আপনার উচ্চারণ বা কথা বলার ভঙ্গি ভাল হয় নি তবে এতে আপনি অন্যের সামনে বিব্রত বোধ করতে পারেন। কিন্তু এরকমটি ঘটার সম্ভাবনা খুবই কম।

আপনি যদি একজন বাংলা শিক্ষার্থী হয়ে থাকেন, তবে এই অধ্যয়ন বা পাঠ আপনাকে বাংলা উচ্চারণের উপায় সম্বন্ধে দিকনির্দেশনা দিতে পারে। বাংলায় আরো ভালভাবে যোগাযোগ স্থাপনে এটা আপনাকে সাহায্য করবে। আর আপনি যদি এদেশীয় বাংলা-ভাষী (নেটিভ) হন, তবে যেসব বিদেশী ইংরেজীভাষী এই পাঠ থেকে সাহায্য পেয়েছেন কেবল তাদের বাংলা আরো ভালভাবে বুঝতে পারা ছাড়া এটা আসলেই আপনার তেমন কোন কাজে আসবে না।

এই পাঠের মাধ্যমে প্রাপ্ত যে-কোন তথ্য যা আপনার পরিচয়-বাহক হতে পারে, সেসব গোপন রাখা হবে এবং কেবল আপনার অনুমতিক্রমেই সেগুলি অন্যদের কাছে দেওয়া হবে। এই অধ্যয়ন বা পাঠের শেষে এই গবেষণা উপাত্তসমূহ এবং এই সম্মতি-পত্র পৃথক একটি তালাবন্ধ স্থানে অন্তত তিন বছরের জন্য সংরক্ষিত থাকবে। গবেষণার উপাত্তসমূহ এই অধ্যয়নকারীর বাড়ি, যা বর্তমানে বাংলাদেশে অবস্থিত, সেখানে সংরক্ষণ করা হবে। আর এই সম্মতি-পত্র অধ্যয়নকারীর পিতা-মাতার বাড়িতে, যা বর্তমানে আমেরিকার ওহিও রাজ্যে অবস্থিত, সেখানে সংরক্ষণ করা হবে। একমাত্র গবেষণাকারী, উপদেষ্টা, এবং যারা আই আর বি প্রক্রিয়াসমূহ নিরীক্ষা করেন, তাদেরই কেবল এই উপাত্তসমূহে প্রবেশাধিকার থাকবে। এই অধ্যয়নের সমাপ্তিকাল থেকে তিন বছর পর সম্মতিপত্রসহ আপনার পরিচয়বাহক সকল গবেষণা উপাত্ত সম্ভবত ছিঁড়ে বা পুড়িয়ে বিনষ্ট করা হবে। উপরের বক্তব্যে আপনার পঠিত বাংলা বাক্যগুলির রেকর্ডিংএর কথা বোঝানো হয়নি, যা ভবিষ্যতে অন্যান্য ভাষা গবেষকদের জন্য করা হতে পারে, বিশেষ করে আরেকটি ভাষা শিক্ষায় আরো অধ্যয়নের জন্য। এটা হয়তো সম্ভব যে, আপনার পরিচিত কেউ যদি এই রেকর্ডিং শোনে তবে সে আপনার কণ্ঠস্বর চিনে ফেলবে, কিন্তু এরকম ঘটার এবং ঘটার পরিপ্রেক্ষিতে কিছু হবার সম্ভবনা অত্যন্ত ক্ষীণ।

এই অধ্যয়নে অংশগ্রহণের জন্য আপনার কিছু খরচ হতে পারে, যেমন পরিবহন খরচ বা যাতায়াত। যদিও এই অধ্যয়ন কাজে আপনার সহযোগিতা স্বেচ্ছাসেবামূলক হিসাবে বিবেচিত, তবুও আপনি চাইলে আপনাকে ৳ ৫০/= ($ ০.৮৫) পর্যন্ত আপনার যাতায়াত খরচ বাবদ প্রদান করা হবে। আপনার খরচ যদি এই পরিমাণের চেয়ে বেশী হয়, এবং তাতে যদি এই অধ্যয়নে অংশগ্রহণ করা আপনার পক্ষে কঠিন হয়ে পড়ে, তবে দয়া করে, এবিষয়ে আর কিছু করা যায় কিনা, সেটা এই অধ্যয়নকারী ব্যক্তিকে জিজ্ঞাসা করুন।

এই অধ্যয়নে আপনি অংশগ্রহণ করবেন, কি করবেন না সে বিষয়ে আপনার সিদ্ধান্তের ফলে এই অধ্যয়নকারী ব্যক্তি বা তার সঙ্গে কাজে নিয়োজিত ব্যক্তিবর্গের সঙ্গে আপনার ভবিষ্যৎ সম্পর্কের কোন পরিবর্তন হবে না। আপনি যদি কাজে অংশ নিয়ে সেটা আবার বন্ধ করতে চান, তবে যে-কোন সময় আপনি সেটা করতে পারেন; এতে আপনার বিরুদ্ধে কিছু করা হবে না। তবে এই অধ্যয়নে আপনার অংশগ্রহণ বন্ধ করার জন্য, দয়া করে এই অধ্যয়নকারী ব্যক্তিকে বলুন যে, আপনি বন্ধ করতে চান, এবং তারপর যখন চাইবেন, তখনই আপনি এই কাজ ছেড়ে চলে যেতে পারবেন।

এই অধ্যয়নকারী ব্যক্তি এ সংক্রান্ত আপনার যেকোন প্রশ্নের উত্তর দেওয়ার জন্য প্রস্তুত আছেন। অবশ্য, এই অধ্যয়ন সম্পর্কে ভবিষ্যতেও আপনার কোন প্রশ্ন থাকতে পারে এবং সে সম্পর্কেও প্রশ্ন জিজ্ঞাসা করার জন্য আপনাকে উৎসাহিত করা হচ্ছে। ডাস্টিন মিলার, যখন তিনি বাংলাদেশে থাকেন, তখন ০২-৮১২-১৩৮৬ (অফিস) ও ০১৭১-৫৬৩-৪৯২ (মোবাইল) নম্বর টেলিফোনে অথবা যখন তিনি আমেরিকায় থাকেন, তখন ৩৩০-৬৭৪-২৪৯৭ নম্বর টেলিফোনের মাধ্যমেও তাকে প্রশ্ন করা যেতে পারে। এছাড়াও আপনি dustbiz@softhome.net নম্বরে ই-মেইলের মাধ্যমে ডাস্টিন মিলারের সঙ্গে যোগাযোগ করতে পারেন। আপনি অবশ্য ডাস্টিন মিলারের উপদেষ্টা (Joan Baart, University of North Dakota, Grand Forks, ND 58202 USA) - এই ঠিকানায়ও যোগাযোগ করতে পারেন। এ সংক্রান্ত আপনার যদি কোন প্রশ্ন বা জানার কোন বিষয় থাকে, তবে অনুগ্রহ করে, আমেরিকায় অবস্থিত নর্থ ডাকোটা বিশ্ববিদ্যালয়ের গবেষণা ও কার্যক্রম উন্নয়ন কার্যালয়ে (University of North Dakota Office of Research and Program Development, U.S.A) ৭০১-৭৭৭-৪২৭৯ নম্বর টেলিফোনে যোগাযোগ করুন।

আপনি যদি এই অধ্যয়নের ফলাফল সম্পর্কে অবগত হতে চান, তবে দয়া করে ডাস্টিন মিলারকে লিখিতভাবে অনুরোধ করুন। এই অনুরোধটি আপনি ব্যক্তিগতভাবে, ই-মেইলের মাধ্যমে (dustbiz@softhome.net)

অথবা ২/২২ ব্লক বি লালমাটিয়া, ঢাকা ১২০৭ বাংলাদেশ, ঠিকানায় ডাকযোগে পাঠতে পারেন। আপনার লিখিত অনুরোধ পাওয়ার পর, এই অধ্যয়ন শেষ হলে তার ফলাফল সম্বন্ধে আপনাকে অবগত করানোর জন্য ডাস্টিন মিলার যথাসাধ্য চেষ্টা করবেন।

আপনি যদি এই অধ্যয়নে অংশগ্রহণ করতে সম্মত হন, তবে এই সম্মতি-পত্রের দুটি অনুলিপিতে আপনাকে স্বাক্ষর করার জন্য বলা হবে, যার একটি আপনাকে দেওয়া হবে।

এই অধ্যয়ন চলাকালে আপনি যদি আহত হন, তবে, সেই একই পরিস্থিতিতে একজন সাধারণ ব্যক্তিকে যে চিকিৎসা ব্যবস্থা দেওয়া হয়, সেই ব্যবস্থা আপনার জন্যেও করা হবে। তবে এই ধরণের চিকিৎসার জন্য প্রয়োজনীয় ব্যয় অবশ্যই আপনাকে অথবা আপনার যদি কোন তৃতীয় পক্ষ থাকে (যেমন, স্বাস্থ্য বীমা, মেডিকেয়ার ইত্যাদি) তাকে পরিশোধ করতে হবে।

এই অধ্যয়ন সংক্রান্ত সকল প্রশ্নের উত্তর আমাকে দেওয়া হয়েছে, এবং ভবিষ্যতে যদি এই অধ্যয়ন সংক্রান্ত কোন প্রশ্ন দেখা দেয়, তবে বিনা দ্বিধায় আমি সে বিষয়ে জিজ্ঞাসা করবো।

উপর্যুক্ত সমস্ত বিষয় আমি পড়েছি, এবং প্রথম পক্ষ কর্তৃক আমার কাছে বিবৃত এই অধ্যয়নে অংশগ্রহণের জন্য স্বেচ্ছায় সম্মত হয়েছি।

প্রথম পক্ষ: _____

দ্বিতীয় পক্ষ (অংশগ্রহণকারী) এর স্বাক্ষর: _____

তারিখ: _____

Appendix 6

Segmentation Criteria

1. **Closure:** The positive zero crossing closest to the end of the second and third formants and to a sharp decrease in the amplitude of the preceding vowel accompanied by the waveform becoming more sinusoidal in shape.
2. **Zero Crossing:** A point on the waveform at which the signal crosses the 0 dB mark. Generally, boundaries were set at positive zero crossings, but where there was a need for greater precision, negative zero crossings were used or a boundary was set on the basis of factors other than a zero crossing, such as the beginning or end of a burst.
3. **Closure Voicing End:** The point, if any, within the closure interval when low-amplitude periodicity ended, which may have been associated with the previous vowel (its 'voice tail') or the closure voicing of the consonant. This was taken to be the zero crossing closest to the point at which the signal periodicity ended. The point at which the voicing pulses at the bottom of the spectrogram began to fade was also considered, as well as the pitch-marking feature in Praat.
4. **Burst:** The clear beginning of marked aperiodicity (distinct from the closure interval's own characteristics) on the waveform in conjunction with one or more strong striations or spikes in the spectrogram. If there seemed to be a prior period of slight aperiodicity, a subjective decision was made about its inclusion based on its proximity to the burst and its level of energy as indicated by its darkness on the spectrogram. For the burst boundary zero crossings were not considered.
5. **Aspiration Noise:** The boundary for the beginning of aspiration noise was set after the end of the burst, any frication, and any periodicity related to an inserted mini-vowel, where a sustained period of frication-like noise in the lower frequencies began. Its endpoint was considered to be the onset of the following vowel, or the point at which its acoustic features were no longer evident.

6. **Frication:** Continuous energy (noise), associated with the production of a fricative, or the frication-phase of an affricate; distinguished from aspiration noise by having more energy in the higher frequencies.
7. **Vowel Onset:** The positive zero crossing closest to the beginning of regular vertical striations in the spectrogram and the first cycle period in the waveform associated with the following vowel. As much of the periodic cycle as possible was included, which may have sometimes necessitated using a negative zero crossing as a boundary point. However, periodic cycles with burst activity on them were kept with the burst. In cases where there was ambiguity as to where the vocal fold vibrations associated with the vowel began, the following factors were considered: the pattern of the vowel's cycle, presence of voicing as indicated by voicing pulses at the bottom of the spectrogram and/or a pitch contour, and impressionistic listening.
8. **Vowel:** A vowel before a consonant was considered to stop at the beginning of that consonant. A vowel before another vowel when there was continuous voicing was considered to stop at the midpoint of their transition. If there was silence between two vowels, the boundaries of the vowels in the target word were determined by the criteria for vowel onset or closure.
9. **Smooth transitions:** When there was a smooth transition between two segments for which it was difficult to apply the above criteria with any degree of certainty, then a temporal midpoint was chosen at which to put the boundary. This process was guided by considering the characteristics of each segment in its purest form as well as impressionistic listening.
10. **Frication and vowels:** When there were transitions between fricatives and vowels and vice-versa, the change in amplitude and wave structure of the vowel and the degree of frication were both used to determine a boundary point, giving similar weight to both indicators. When there was no clear boundary point, the principles for smooth transitions were applied.

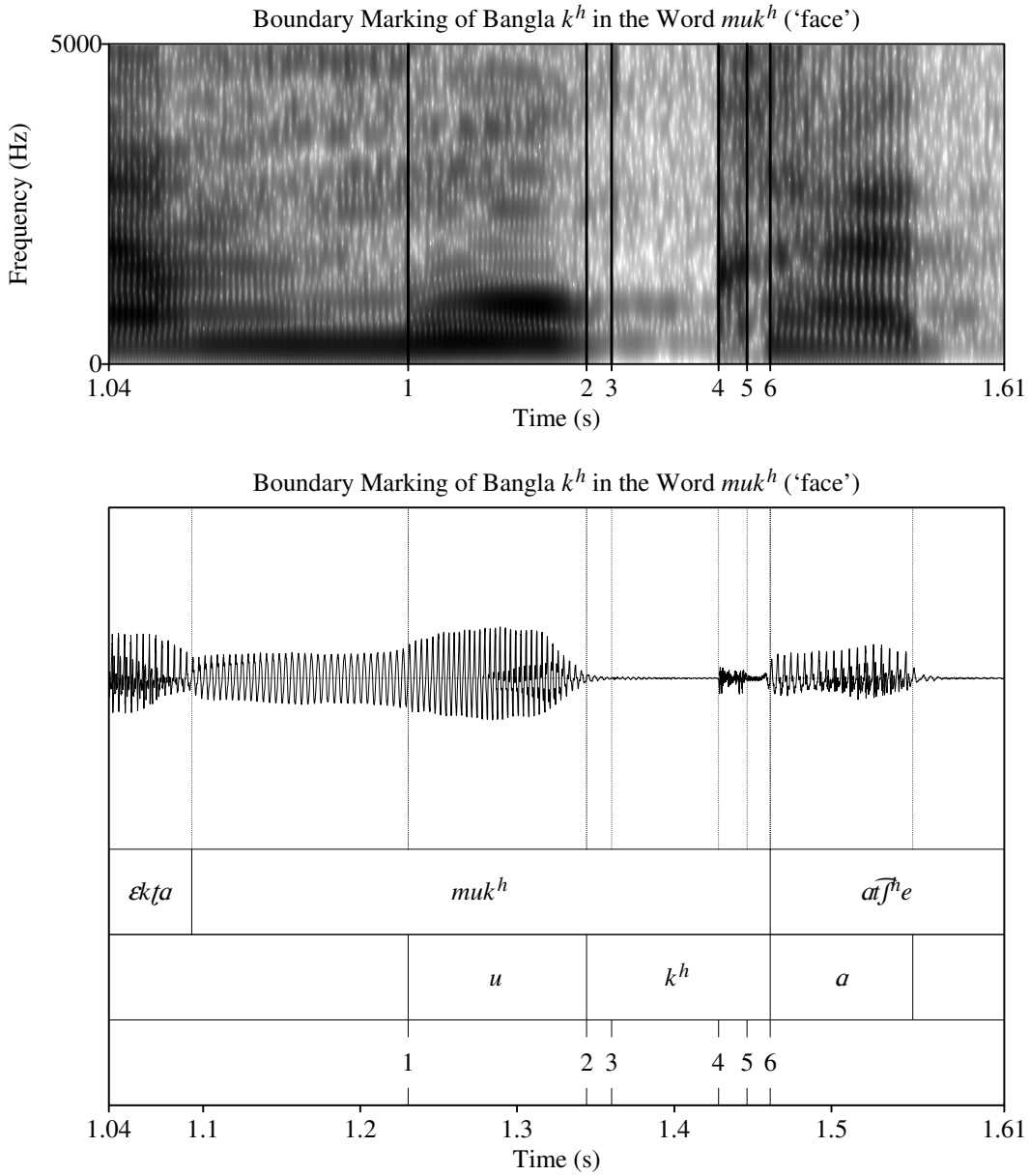


Figure 58. Example of the Application of Segmentation Criteria in Marking Boundaries [Key: 1 = Beginning of preceding vowel. 2 = Closure and beginning of stop. 3 = End of low-amplitude periodicity during closure. 4 = Beginning of burst. 5 = Beginning of aspiration noise. 6 = Vowel onset.]

Appendix 7

Details of One-Way ANOVA Significance Tests

Statistical Data for L1 (Native Bangla Speaker) Production Values by Stop Class

		ANOVA Results			–V–A			–V+A			+V–A			+V+A		
Sample Group	**AV**	***f***	***p***	***df***	**Mn**	**SD**	**N**	**Mn**	**SD**	**N**	**Mn**	**SD**	**N**	**Mn**	**SD**	**N**
L1	VVOT	138.5	0.000	3, 461	13.3	7.6	129	70.1	31.9	93	7.1	11.2	128	49.4	42.8	115
L1-In	VVOT	176.1	0.000	3, 178	12.2	6.8	48	91.9	23.8	40	8.1	5.8	48	78.1	36.6	46
L1-Me	VVOT	60.2	0.000	3, 177	14.3	7.9	48	63.2	21.9	38	4.3	3.4	48	41.2	38.4	47
L1-Fi	VVOT	6.8	0.000	3, 98	13.6	8.4	33	29.2	24.0	15	9.7	20.5	32	7.2	6.2	22
L1-M	VVOT	73.5	0.000	3, 232	14.1	8.2	66	66.5	26.1	47	8.5	14.9	65	32.9	33.5	58
L1-F	VVOT	88.3	0.000	3, 225	12.5	7.0	63	73.7	36.9	46	5.6	4.8	63	66.2	45.0	57
L1	CD	41.7	0.000	3, 496	94.9	31.4	138	78.5	23.5	118	72.0	26.7	121	59.0	22.6	123
L1-In	CD	18.6	0.000	3, 177	111.8	32.0	47	83.3	23.3	40	87.1	28.0	48	73.9	15.3	46
L1-Me	CD	41.4	0.000	3, 162	86.1	23.5	46	73.3	22.0	38	54.7	14.9	37	43.6	16.1	45
L1-Fi	CD	7.3	0.000	3, 149	86.3	31.1	45	78.9	24.6	40	69.8	23.1	36	59.2	25.0	32
L1-M	CD	17.1	0.000	3, 243	89.3	32.3	68	73.6	24.0	61	66.7	26.1	59	55.6	24.9	59
L1-F	CD	27.1	0.000	3, 249	100.4	29.7	70	83.9	21.9	57	77.1	26.5	62	62.1	19.8	64
L1	CV	2105.9	0.000	3, 493	19.4	10.9	138	20.5	9.9	115	98.6	5.4	121	93.6	14.4	123
L1-In	CV	1058.3	0.000	3, 175	19.2	11.6	47	21.0	9.0	38	99.1	3.7	48	94.6	10.3	46
L1-Me	CV	1379.5	0.000	3, 162	17.0	7.8	46	17.4	10.1	38	100.0	0.0	37	96.5	9.7	45
L1-Fi	CV	323.4	0.000	3, 148	21.9	12.6	45	23.1	10.0	39	96.6	8.6	36	88.1	22.1	32
L1-M	CV	1771.7	0.000	3, 241	18.6	9.2	68	21.5	10.6	59	98.8	4.7	59	97.7	7.8	59
L1-F	CV	774.0	0.000	3, 248	20.1	12.4	70	19.5	9.1	56	98.4	6.0	62	89.9	17.8	64
L1	PVD	7.5	0.000	3, 560	111.6	44.8	144	106.4	46.4	144	128.8	55.3	144	127.6	48.8	132
L1-In	PVD	13.8	0.000	3, 188	67.0	18.1	48	62.5	18.7	48	82.4	21.8	48	86.7	27.7	48
L1-Me	PVD	5.4	0.001	3, 188	113.7	26.2	48	105.0	27.8	48	112.2	25.6	48	126.8	28.8	48
L1-Fi	PVD	13.6	0.000	3, 176	154.1	35.0	48	151.7	37.0	48	191.7	40.7	48	183.2	36.3	36
L1-M	PVD	4.4	0.005	3, 278	101.8	42.7	72	94.7	39.3	72	115.0	46.3	72	117.3	43.6	66
L1-F	PVD	3.8	0.011	3, 278	121.3	45.0	72	118.1	50.2	72	142.5	60.3	72	137.9	51.9	66
L1	Asp.	17.8	0.000	1, 154	N/A	N/A	N/A	59.9	24.3	95	N/A	N/A	N/A	77.6	27.4	61
L1-In	Asp.	3.1	0.084	1, 80	N/A	N/A	N/A	69.7	22.3	44	N/A	N/A	N/A	78.8	25.2	38
L1-Me	Asp.	18.5	0.000	1, 56	N/A	N/A	N/A	48.5	17.7	39	N/A	N/A	N/A	69.5	16.8	19
L1-Fi	Asp.	3.1	0.101	1, 14	N/A	N/A	N/A	61.0	34.7	12	N/A	N/A	N/A	104.3	64.0	4
L1-M	Asp.	1.4	0.235	1, 66	N/A	N/A	N/A	54.7	21.4	48	N/A	N/A	N/A	61.5	21.9	20
L1-F	Asp.	12.9	0.001	1, 86	N/A	N/A	N/A	65.2	26.2	47	N/A	N/A	N/A	85.4	26.5	41

Key: –V–A = Voiceless unaspirated. **–V+A** = Voiceless aspirated. **+V–A** = Voiced unaspirated. **+V+A** = Voiced aspirated. **AV** = Acoustic variable. ***f*** = *f*-value of one-way ANOVA. ***p*** = *p*-value of one-way ANOVA. ***df*** = Degrees of freedom. **Mn** = Mean. **SD** = Standard deviation. **N** = Sample size. **In** = Word-initial position. **Me** = Word-medial position. **Fi** = Word-final position. **M** = Male. **F** = Female. **VVOT** = Vowel voicing onset time (ms). **CD** = Closure duration (ms). **CV** = Closure voicing (as % of closure duration). **PVD** = Preceding vowel duration (ms). **Asp.** = Aspiration noise duration (ms).

Statistical Data for L2 (Native English Speaker) Production Values by Stop Class

		ANOVA Results			–V–A			–V+A			+V–A			+V+A		
Sample Group	**AV**	***f***	***p***	***df***	**Mn**	**SD**	**N**	**Mn**	**SD**	**N**	**Mn**	**SD**	**N**	**Mn**	**SD**	**N**
L2	VVOT	66.4	0.000	3, 446	25.1	18.9	116	68.7	49.1	100	10.0	14.0	120	70.7	60.8	114
L2-In	VVOT	44.9	0.000	3, 181	27.5	23.3	48	88.9	49.5	41	9.3	6.9	48	83.5	60.1	48
L2-Me	VVOT	28.3	0.000	3, 180	21.0	12.6	47	54.5	38.6	42	6.2	5.2	48	66.9	60.6	47
L2-Fi	VVOT	3.2	0.027	3, 77	28.7	18.8	21	55.0	56.5	17	18.8	27.3	24	47.7	57.7	19
L2-M	VVOT	29.4	0.000	3, 233	26.5	17.9	61	60.3	44.3	54	9.0	7.5	63	57.8	53.5	59
L2-F	VVOT	39.8	0.000	3, 209	23.6	20.1	55	78.6	53.0	46	11.0	18.8	57	84.6	65.4	55
L2	CD	13.4	0.000	3, 520	98.4	33.3	143	94.1	29.6	122	75.7	32.7	133	83.7	34.2	126
L2-In	CD	3.9	0.011	3, 181	118.4	32.3	48	97.5	29.8	41	102.5	27.3	48	102.3	35.7	48
L2-Me	CD	18.9	0.000	3, 168	85.8	23.2	47	91.1	29.4	42	54.1	16.7	39	72.3	25.5	44
L2-Fi	CD	8.2	0.000	3, 163	90.9	34.1	48	93.9	30.1	39	66.0	29.4	46	72.3	30.8	34
L2-M	CD	6.1	0.001	3, 258	93.0	35.9	71	89.1	28.8	65	70.6	30.9	66	81.5	34.9	60
L2-F	CD	8.1	0.000	3, 258	103.8	29.9	72	99.9	29.8	57	80.7	34.0	67	85.8	33.7	66
L2	CV	1041.7	0.000	3, 513	18.4	12.4	137	19.7	12.6	121	97.0	8.7	133	85.2	22.4	126
L2-In	CV	329.6	0.000	3, 178	18.8	12.6	46	20.0	11.4	40	98.7	4.7	48	83.6	25.2	48
L2-Me	CV	325.7	0.000	3, 167	17.6	12.6	46	19.4	14.0	42	98.3	8.6	39	85.3	22.3	44
L2-Fi	CV	394.1	0.000	3, 160	18.8	12.2	45	19.8	12.3	39	94.3	11.2	46	87.2	18.3	34
L2-M	CV	457.1	0.000	3, 253	22.0	14.8	67	22.8	13.3	64	97.7	9.2	66	87.0	21.8	60
L2-F	CV	626.9	0.000	3, 256	14.9	8.2	70	16.3	10.8	57	96.4	8.2	67	83.5	22.9	66
L2	PVD	9.8	0.000	3, 560	108.6	43.6	144	103.8	34.8	144	128.0	51.7	144	125.0	50.4	132
L2-In	PVD	0.7	0.548	3, 188	80.5	39.5	48	80.7	32.9	48	83.1	32.5	48	90.4	45.5	48
L2-Me	PVD	9.7	0.000	3, 188	102.9	22.2	48	102.4	21.3	48	120.2	21.3	48	121.7	28.8	48
L2-Fi	PVD	20.7	0.000	3, 176	142.4	41.6	48	128.3	31.5	48	180.8	40.7	48	175.7	36.4	36
L2-M	PVD	4.5	0.005	3, 278	120.4	49.4	72	102.4	35.6	72	127.8	53.8	72	128.0	51.8	66
L2-F	PVD	8.5	0.000	3, 278	96.8	33.4	72	105.2	34.2	72	128.2	49.9	72	122.1	49.2	66
L2	Asp.	20.3	0.000	2, 193	43.8	22.7	24	64.7	44.3	99	N/A	N/A	N/A	97.5	43.1	73
L2-In	Asp.	7.6	0.001	2, 81	45.1	25.9	12	73.5	44.6	40	N/A	N/A	N/A	97.6	42.0	32
L2-Me	Asp.	11.8	0.000	2, 64	31.1	13.7	6	46.2	35.1	33	N/A	N/A	N/A	90.4	46.5	28
L2-Fi	Asp.	5.1	0.011	2, 42	53.7	19.7	6	74.5	48.3	26	N/A	N/A	N/A	112.6	36.8	13
L2-M	Asp.	5.7	0.005	2, 83	48.1	20.1	12	58.1	42.4	45	N/A	N/A	N/A	84.1	34.6	29
L2-F	Asp.	14.2	0.000	2, 107	39.4	25.1	12	70.2	45.4	54	N/A	N/A	N/A	106.4	46.2	44

Key: –V–A = Voiceless unaspirated. **–V+A** = Voiceless aspirated. **+V–A** = Voiced unaspirated. **+V+A** = Voiced aspirated. **AV** = Acoustic variable. ***f*** = *f*-value of one-way ANOVA. ***p*** = *p*-value of one-way ANOVA. ***df*** = Degrees of freedom. **Mn** = Mean. **SD** = Standard deviation. **N** = Sample size. **In** = Word-initial position. **Me** = Word-medial position. **Fi** = Word-final position. **M** = Male. **F** = Female. **VVOT** = Vowel voicing onset time (ms). **CD** = Closure duration (ms). **CV** = Closure voicing (as % of closure duration). **PVD** = Preceding vowel duration (ms). **Asp.** = Aspiration noise duration (ms).

Statistical Data for L1 (Native Bangla Speaker) Production Values by Voice

		ANOVA Results			Voiceless			Voiced		
Sample Group	**AV**	***f***	***p***	***df***	**Mn**	**SD**	**N**	**Mn**	**SD**	**N**
L1	VVOT	8.8	0.003	1, 463	37.1	35.3	222	27.1	37.1	243
L1-In	VVOT	0.9	0.348	1, 180	48.4	43.3	88	42.3	43.6	94
L1-Me	VVOT	8.3	0.004	1, 179	35.9	29.0	86	22.6	32.7	95
L1-Fi	VVOT	9.2	0.003	1, 100	18.5	16.5	48	8.7	16.2	54
L1	CD	79.9	0.000	1, 498	87.4	29.1	256	65.4	25.5	244
L1-In	CD	19.3	0.000	1, 179	98.7	31.6	87	80.6	23.5	94
L1-Me	CD	100.3	0.000	1, 164	80.3	23.6	84	48.6	16.5	82
L1-Fi	CD	17.2	0.000	1, 151	82.8	28.3	85	64.8	24.4	68
L1	CV	794.3	0.000	1, 495	16.2	7.6	253	62.4	25.0	244
L1-In	CV	3091.9	0.000	1, 177	20.0	10.5	85	96.9	7.9	94
L1-Me	CV	4090.7	0.000	1, 164	17.2	8.9	84	98.1	7.4	82
L1-Fi	CV	935.8	0.000	1, 150	22.4	11.4	84	92.6	16.8	68
L1	PVD	21.7	0.000	1, 562	109.0	45.6	288	128.2	52.2	276
L1-In	PVD	39.3	0.000	1, 190	64.8	18.5	96	84.6	24.9	96
L1-Me	PVD	6.5	0.012	1, 190	109.3	27.2	96	119.5	28.1	96
L1-Fi	PVD	39.9	0.000	1, 178	152.9	35.8	96	188.1	38.9	84
L1	Asp.	17.8	0.000	1, 154	59.9	24.3	95	77.6	27.4	61
L1-In	Asp.	3.1	0.084	1, 80	69.7	22.3	44	78.8	25.2	38
L1-Me	Asp.	18.5	0.000	1, 56	48.5	17.7	39	69.5	16.8	19
L1-Fi	Asp.	3.1	0.101	1, 14	61.0	34.7	12	104.3	64	4
L1 +A	Asp.	17.8	0.000	1, 154	59.9	24.3	95	77.6	27.4	61

Key: AV = Acoustic variable. ***f*** = *f*-value of one-way ANOVA. ***p*** = *p*-value of one-way ANOVA. ***df*** = Degrees of freedom. **Mn** = Mean. **SD** = Standard deviation. **N** = Sample size. **In** = Word-initial position. **Me** = Word-medial position. **Fi** = Word-final position. **+A** = Aspirated. **VVOT** = Vowel voicing onset time (ms). **CD** = Closure duration (ms). **CV** = Closure voicing (as % of closure duration). **PVD** = Preceding vowel duration (ms). **Asp.** = Aspiration noise duration (ms).

Statistical Data for L2 (Native English Speaker) Production Values by Voice

		ANOVA Results			Voiceless			Voiced		
Sample Group	**AV**	***f***	***p***	***df***	**Mn**	**SD**	**N**	**Mn**	**SD**	**N**
L2	VVOT	1.6	0.208	1, 448	45.3	42.2	216	39.6	53.1	234
L2-In	VVOT	1.5	0.228	1, 183	55.8	48.5	89	46.4	56.6	96
L2-Me	VVOT	0.01	0.929	1, 182	36.8	32.6	89	36.3	52.4	95
L2-Fi	VVOT	0.8	0.363	1, 79	40.5	41.8	38	31.6	45.3	43
L2	CD	34.9	0.000	1, 522	96.5	31.7	265	79.6	33.6	259
L2-In	CD	1.8	0.179	1, 183	108.8	32.7	89	102.4	31.6	96
L2-Me	CD	41.4	0.000	1, 170	88.3	26.3	89	63.8	23.5	83
L2-Fi	CD	23.8	0.000	1, 165	92.2	32.2	87	68.7	30.0	80
L2	CV	2859.1	0.000	1, 515	19.0	12.5	258	91.3	17.8	259
L2-In	CV	866.2	0.000	1, 180	19.3	12.0	86	91.1	19.6	96
L2-Me	CV	893.4	0.000	1, 169	18.5	13.3	88	91.4	18.4	83
L2-Fi	CV	1151.3	0.000	1, 162	19.3	12.1	84	91.3	15.0	80
L2	PVD	28.3	0.000	1, 562	106.2	39.5	288	126.6	51.0	276
L2-In	PVD	1.3	0.266	1, 190	80.6	36.2	96	86.7	39.5	96
L2-Me	PVD	29.1	0.000	1, 190	102.6	21.7	96	120.9	25.2	96
L2-Fi	PVD	57.9	0.000	1, 178	135.4	37.4	96	178.6	38.8	84
L2	Asp.	35.0	0.000	1, 194	60.6	41.7	123	97.5	43.1	73
L2-In	Asp.	10.4	0.002	1, 82	67.0	42.6	52	97.6	42.0	32
L2-Me	Asp.	22.9	0.000	1, 65	43.9	33.1	39	90.4	46.5	28
L2-Fi	Asp.	8.9	0.005	1, 43	70.6	44.8	32	112.6	36.8	13
L2 +A	Asp.	23.6	0.000	1, 170	64.7	44.3	99	97.5	43.1	73

Key: AV = Acoustic variable. ***f*** = *f*-value of one-way ANOVA. ***p*** = *p*-value of one-way ANOVA. ***df*** = Degrees of freedom. **Mn** = Mean. **SD** = Standard deviation. **N** = Sample size. **In** = Word-initial position. **Me** = Word-medial position. **Fi** = Word-final position. **+A** = Aspirated. **VVOT** = Vowel voicing onset time (ms). **CD** = Closure duration (ms). **CV** = Closure voicing (as % of closure duration). **PVD** = Preceding vowel duration (ms). **Asp.** = Aspiration noise duration (ms).

Statistical Data for L1 (Native Bangla Speaker) Production Values by Aspiration

		ANOVA Results			Unaspirated			Aspirated		
Sample Group	**AV**	***f***	***p***	***df***	**Mn**	**SD**	**N**	**Mn**	**SD**	**N**
L1	VVOT	356.1	0.000	1, 463	10.2	10.0	257	58.6	39.6	208
L1-In	VVOT	498.8	0.000	1, 180	10.1	6.7	96	84.5	31.9	86
L1-Me	VVOT	138.0	0.000	1, 179	9.3	7.8	96	51.0	33.8	85
L1-Fi	VVOT	1.6	0.211	1, 100	11.7	15.6	65	16.1	19.2	37
L1	CD	37.8	0.000	1, 498	84.2	31.4	259	68.6	25.0	241
L1-In	CD	27.2	0.000	1, 179	99.3	32.4	95	78.2	19.9	86
L1-Me	CD	15.1	0.000	1, 164	72.1	25.4	83	57.2	24.0	83
L1-Fi	CD	3.9	0.051	1, 151	79.0	28.9	81	70.1	26.5	72
L1	CV	0.3	0.591	1, 495	56.4	40.6	259	58.3	38.7	238
L1-In	CV	0.1	0.769	1, 177	59.6	41.0	95	61.3	38.1	84
L1-Me	CV	1.0	0.328	1, 164	54.0	41.9	83	60.3	40.9	83
L1-Fi	CV	0.2	0.661	1, 150	55.1	38.9	81	52.4	36.5	71
L1	PVD	0.7	0.390	1, 562	120.2	51.0	288	116.6	48.7	276
L1-In	PVD	0.0	0.982	1, 190	74.7	21.4	96	74.6	26.5	96
L1-Me	PVD	0.6	0.459	1, 190	112.9	25.8	96	115.9	30.2	96
L1-Fi	PVD	1.6	0.213	1, 178	172.9	42.2	96	165.2	39.7	84

Key: AV = Acoustic variable. ***f*** = *f*-value of one-way ANOVA. ***p*** = *p*-value of one-way ANOVA. ***df*** = Degrees of freedom. **Mn** = Mean. **SD** = Standard deviation. **N** = Sample size. **In** = Word-initial position. **Me** = Word-medial position. **Fi** = Word-final position. **VVOT** = Vowel voicing onset time (ms). **CD** = Closure duration (ms). **CV** = Closure voicing (as % of closure duration). **PVD** = Preceding vowel duration (ms).

Statistical Data for L2 (Native English Speaker) Production Values by Aspiration

		ANOVA Results			Unaspirated			Aspirated		
Sample Group	**AV**	***f***	***p***	***df***	**Mn**	**SD**	**N**	**Mn**	**SD**	**N**
L2	VVOT	188.0	0.000	1, 448	17.4	18.2	236	69.8	55.5	214
L2-In	VVOT	127.2	0.000	1, 183	18.4	19.4	96	86.0	55.2	89
L2-Me	VVOT	76.4	0.000	1, 182	13.5	12.1	95	61.1	51.5	89
L2-Fi	VVOT	8.9	0.004	1, 79	23.4	24.0	45	51.2	56.5	36
L2	CD	0.2	0.642	1, 522	87.5	34.9	276	88.8	32.4	248
L2-In	CD	4.8	0.029	1, 183	110.4	30.8	96	100.1	33.0	89
L2-Me	CD	5.7	0.018	1, 170	71.5	25.9	86	81.5	28.9	86
L2-Fi	CD	1.0	0.324	1, 165	78.7	34.1	94	83.8	32.1	73
L2	CV	1.4	0.245	1, 515	57.1	40.8	270	53.1	37.5	247
L2-In	CV	0.7	0.404	1, 180	59.6	41.2	94	54.7	37.6	88
L2-Me	CV	0.1	0.806	1, 169	54.6	41.9	85	53.1	38.0	86
L2-Fi	CV	0.9	0.344	1, 162	57.0	39.7	91	51.2	37.1	73
L2	PVD	1.2	0.268	1, 562	118.3	48.7	288	114.0	44.2	276
L2-In	PVD	0.5	0.494	1, 190	81.8	36.0	96	85.5	39.7	96
L2-Me	PVD	0.02	0.886	1, 190	111.5	23.4	96	112.0	27.0	96
L2-Fi	PVD	4.0	0.046	1, 178	161.6	45.2	96	148.6	41.0	84

Key: AV = Acoustic variable. ***f*** = *f*-value of one-way ANOVA. ***p*** = *p*-value of one-way ANOVA. ***df*** = Degrees of freedom. **Mn** = Mean. **SD** = Standard deviation. **N** = Sample size. **In** = Word-initial position. **Me** = Word-medial position. **Fi** = Word-final position. **VVOT** = Vowel voicing onset time (ms). **CD** = Closure duration (ms). **CV** = Closure voicing (as % of closure duration). **PVD** = Preceding vowel duration (ms).

Statistical Data for L1 (Native Bangla Speaker) Production Values by Position in the Word

		ANOVA Results			Word-Initial			Word-Medial			Word-Final		
Sample Group	**AV**	***f***	***p***	***df***	**Mn**	**SD**	**N**	**Mn**	**SD**	**N**	**Mn**	**SD**	**N**
L1	VVOT	29.1	0.000	2, 462	45.3	43.4	182	28.9	31.6	181	13.3	17.0	102
L1	CD	34.9	0.000	2, 497	89.3	29.1	181	64.6	25.8	166	74.8	28.0	153
L1	CV	1.1	0.326	2, 494	60.4	39.6	179	57.1	41.4	166	53.8	37.7	152
L1	PVD	417.1	0.000	2, 561	74.7	24.0	192	114.4	28.0	192	169.3	41.1	180
L1	Asp.	9.2	0.000	2, 153	73.9	24.0	82	55.4	19.9	58	71.8	45.6	16

Key: AV = Acoustic variable. ***f*** = *f*-value of one-way ANOVA. ***p*** = *p*-value of one-way ANOVA. ***df*** = Degrees of freedom. **Mn** = Mean. **SD** = Standard deviation. **N** = Sample size. **VVOT** = Vowel voicing onset time (ms). **CD** = Closure duration (ms). **CV** = Closure voicing (as % of closure duration). **PVD** = Preceding vowel duration (ms). **Asp.** = Aspiration noise duration (ms).

Statistical Data for L2 (Native English Speaker) Production Values by Position in the Word

		ANOVA Results			Word-Initial			Word-Medial			Word-Final		
Sample Group	**AV**	***f***	***p***	***df***	**Mn**	**SD**	**N**	**Mn**	**SD**	**N**	**Mn**	**SD**	**N**
L2	VVOT	5.1	0.006	2, 447	50.9	52.9	185	36.5	43.8	184	35.7	43.6	81
L2	CD	45.1	0.000	2, 521	105.5	32.2	185	76.5	27.8	172	80.9	33.2	167
L2	CV	0.4	0.692	2, 514	57.2	39.5	182	53.9	39.9	171	54.4	38.6	164
L2	PVD	184.9	0.000	2, 561	83.7	37.9	192	111.8	25.2	192	155.5	43.7	180
L2	Asp.	3.1	0.045	2, 193	78.7	44.7	84	63.3	45.3	67	82.7	46.4	45

Key: AV = Acoustic variable. ***f*** = *f*-value of one-way ANOVA. ***p*** = *p*-value of one-way ANOVA. ***df*** = Degrees of freedom. **Mn** = Mean. **SD** = Standard deviation. **N** = Sample size. **VVOT** = Vowel voicing onset time (ms). **CD** = Closure duration (ms). **CV** = Closure voicing (as % of closure duration). **PVD** = Preceding vowel duration (ms). **Asp.** = Aspiration noise duration (ms).

Statistical Data for All Stop Production Values by L1/L2 Subject Groups

		ANOVA Results			L1 Speakers			L2 Speakers		
Sample Group	**AV**	***f***	***p***	***df***	**Mn**	**SD**	**N**	**Mn**	**SD**	**N**
All	VVOT	13.7	0.000	1, 913	31.9	36.6	465	42.3	48.2	450
All	CD	33.3	0.000	1, 1022	76.7	29.5	500	88.1	33.7	524
All	CV	0.72	0.398	1, 1012	57.3	39.7	497	55.2	39.3	517
All	PVD	0.6	0.439	1, 1126	118.4	49.9	564	116.2	46.6	564
All	Asp.	3.3	0.070	1, 350	66.8	26.9	156	74.4	45.8	196

Key: AV = Acoustic variable. ***f*** = *f*-value of one-way ANOVA. ***p*** = *p*-value of one-way ANOVA. ***df*** = Degrees of freedom. **Mn** = Mean. **SD** = Standard deviation. **N** = Sample size. **VVOT** = Vowel voicing onset time (ms). **CD** = Closure duration (ms). **CV** = Closure voicing (as % of closure duration). **PVD** = Preceding vowel duration (ms). **Asp.** = Aspiration noise duration (ms).

Statistical Data for L2 (Native English Speaker) Production Values by Duration of Exposure to Bangla

		ANOVA Results			0–4 years			>4 years		
Sample Group	**AV**	***f***	***p***	***df***	**Mn**	**SD**	**N**	**Mn**	**SD**	**N**
L2 –V–A	Asp.	0.6	0.452	1, 22	47.1	25.7	13	39.9	19.1	11

Key: AV = Acoustic variable. ***f*** = *f*-value of one-way ANOVA. ***p*** = *p*-value of one-way ANOVA. ***df*** = Degrees of freedom. **Mn** = Mean. **SD** = Standard deviation. **N** = Sample size. **–V–A** = Voiceless unaspirated. **Asp.** = Aspiration noise duration (ms).

Statistical Data for L1 (Native Bangla Speaker) Production Values by Gender

		ANOVA Results			Male			Female		
Sample Group	**AV**	***f***	***p***	***df***	**Mn**	**SD**	**N**	**Mn**	**SD**	**N**
L1 –V–A	VVOT	1.4	0.236	1, 127	14.1	8.2	66	12.5	7.0	63
L1 –V+A	VVOT	1.2	0.276	1, 91	66.5	26.1	47	73.7	36.9	46
L1 +V–A	VVOT	2.2	0.138	1, 126	8.5	14.9	65	5.6	4.8	63
L1 +V+A	VVOT	20.4	0.000	1, 113	32.9	33.5	58	66.2	45.0	57
L1 –V–A	CD	4.5	0.036	1, 136	89.3	32.3	68	100.4	29.7	70
L1 –V+A	CD	5.9	0.017	1, 116	73.6	24.0	61	83.9	21.9	57
L1 +V–A	CD	4.7	0.031	1, 119	66.7	26.1	59	77.1	26.5	62
L1 +V+A	CD	2.6	0.108	1, 121	55.6	24.9	59	62.1	19.8	64
L1 –V–A	CV	0.6	0.450	1, 136	18.6	9.2	68	20.1	12.4	70
L1 –V+A	CV	1.2	0.275	1, 113	21.5	10.6	59	19.5	9.1	56
L1 +V–A	CV	0.2	0.687	1, 119	98.8	4.7	59	98.4	6.0	62
L1 +V+A	CV	9.5	0.003	1, 121	97.7	7.8	59	89.9	17.8	64
L1 –V–A	PVD	7.1	0.008	1, 142	101.8	42.7	72	121.3	45.0	72
L1 –V+A	PVD	9.7	0.002	1, 142	94.7	39.3	72	118.1	50.2	72
L1 +V–A	PVD	9.4	0.003	1, 142	115.0	46.3	72	142.5	60.3	72
L1 +V+A	PVD	6.1	0.015	1, 130	117.3	43.6	66	137.9	51.9	66
L1 –V+A	Asp.	4.6	0.034	1, 93	54.7	21.4	48	65.2	26.2	47
L1 +V+A	Asp.	12.1	0.001	1, 59	61.5	21.9	20	85.4	26.5	41

Key: AV = Acoustic variable. ***f*** = *f*-value of one-way ANOVA. ***p*** = *p*-value of one-way ANOVA. ***df*** = Degrees of freedom. **Mn** = Mean. **SD** = Standard deviation. **N** = Sample size. **–V–A** = Voiceless unaspirated. **–V+A** = Voiceless aspirated. **+V–A** = Voiced unaspirated. **+V+A** = Voiced aspirated. **VVOT** = Vowel voicing onset time (ms). **CD** = Closure duration (ms). **CV** = Closure voicing (as % of closure duration). **PVD** = Preceding vowel duration (ms). **Asp.** = Aspiration noise duration (ms).

Statistical Data for L2 (Native English Speaker) Production Values by Gender

		ANOVA Results			Male			Female		
Sample Group	**AV**	***f***	***p***	***df***	**Mn**	**SD**	**N**	**Mn**	**SD**	**N**
L2 –V–A	VVOT	0.7	0.411	1, 114	26.5	17.9	61	23.6	20.1	55
L2 –V+A	VVOT	3.6	0.062	1, 98	60.3	44.3	54	78.6	53.0	46
L2 +V–A	VVOT	0.6	0.438	1, 118	9.01	7.45	63	11	18.8	57
L2 +V+A	VVOT	5.8	0.018	1, 112	57.8	53.5	59	84.6	65.4	55
L2 –V–A	CD	3.9	0.050	1, 141	93.0	35.9	71	103.8	29.9	72
L2 –V+A	CD	4.2	0.042	1, 120	89.1	28.8	65	99.9	29.8	57
L2 +V–A	CD	3.2	0.076	1, 131	70.6	30.9	66	80.7	34.0	67
L2 +V+A	CD	0.5	0.483	1, 124	81.5	34.9	60	85.8	33.7	66
L2 –V–A	CV	12.1	0.001	1, 135	22	14.8	67	14.9	8.2	70
L2 –V+A	CV	8.4	0.004	1, 119	22.8	13.3	64	16.3	10.8	57
L2 +V–A	CV	0.8	0.388	1, 131	97.7	9.2	66	96.4	8.2	67
L2 +V+A	CV	0.8	0.386	1, 124	87.0	21.8	60	83.5	22.9	66
L2 –V–A	PVD	11.3	0.001	1, 142	120.4	49.4	72	96.8	33.4	72
L2 –V+A	PVD	0.2	0.623	1, 142	102.4	35.6	72	105.2	34.2	72
L2 +V–A	PVD	0.00	0.961	1, 142	127.8	53.8	72	128.2	49.9	72
L2 +V+A	PVD	0.5	0.505	1, 130	128.0	51.8	66	122.1	49.2	66
L2 –V–A	Asp.	0.9	0.360	1, 22	48.1	20.1	12	39.4	25.1	12
L2 –V+A	Asp.	1.9	0.175	1, 97	58.1	42.4	45	70.2	45.4	54
L2 +V+A	Asp.	4.9	0.030	1, 71	84.1	34.6	29	106.4	46.2	44

Key: AV = Acoustic variable. ***f*** = *f*-value of one-way ANOVA. ***p*** = *p*-value of one-way ANOVA. ***df*** = Degrees of freedom. **Mn** = Mean. **SD** = Standard deviation. **N** = Sample size. **–V–A** = Voiceless unaspirated. **–V+A** = Voiceless aspirated. **+V–A** = Voiced unaspirated. **+V+A** = Voiced aspirated. **VVOT** = Vowel voicing onset time (ms). **CD** = Closure duration (ms). **CV** = Closure voicing (as % of closure duration). **PVD** = Preceding vowel duration (ms). **Asp.** = Aspiration noise duration (ms).

REFERENCES

Bada, Erdogan. 2001. Native language influence on the production of English sounds by Japanese learners. *Reading Matrix: An International Online Journal* 1, no. 2 (September), http://www.readingmatrix.com/articles/bada/article.pdf (accessed June 16, 2008). 15 pages.

Benguerel, André-Pierre, and Tej K. Bhatia. 1980. Hindi stop consonants: An acoustic and fiberscopic study. *Phonetica* 37:134–148.

Bhattacharya, Tanmoy. 2001. Bangla. In *Facts about the world's languages: An encyclopedia of the world's major languages: Past and present*, edited by Jane Garry and Carl Rubino. New York: H. W. Wilson.

Biswas, Sailendra. 2000. *Samsad Bengali-English dictionary*. 3d ed. Kolkata: Sahitya Samsad.

Bloomfield, Leonard. 1933. *Language*. New York: Henry Holt.

Boersma, Paul, and David Weenink. 2008. *Praat: Doing phonetics by computer* (Version 5.0.25) [Computer program]. Retrieved June 4, 2008 from http://www.praat.org/.

Byrd, Dani. 1993. 54,000 American Stops. *UCLA Working Papers in Phonetics* 83:97–115.

Catford, J. C. 1977. *Fundamental problems in phonetics*. Bloomington: Indiana University Press.

Chatterji, Suniti Kumar. 1921. Bengali phonetics. *Bulletin of the School of Oriental and African Studies* 2:1–25.

Chen, Matthew. 1970. Vowel length variation as a function of the voicing of the consonant environment. *Phonetica* 22:129–59.

Cho, Taehong, and Peter Ladefoged. 1999. Variations and universals in VOT: Evidence from 18 languages. *Journal of Phonetics* 27, no. 2:207–229.

Chomsky, Noam. 1968. *Language and mind*. New York: Harcourt Brace & World, Inc.

Corder, S. Pit. 1973. *Introducing applied linguistics*. Harmondsworth: Penguin Books.

Crystal, David, ed. 1997. *The Cambridge encyclopedia of language*. 2d ed. Cambridge: Cambridge University Press.

Davis, Katharine. 1994. Stop voicing in Hindi. *Journal of Phonetics* 22, no. 2:177–193.

Davis, Katharine. 1995. Phonetic and phonological contrasts in the acquisition of voicing: Voice onset time production in Hindi and English. *Journal of Child Language* 22:275–305.

Davis, Katharine, and Mary Beckman. 1983. Production and perception of the voicing contrast in Indian and American English. *Working Papers of the Cornell Phonetics Laboratory* 1:77–90.

De Bot, Kees, Wander Lowie, and Marjorlijn Verspoor. 2005. *Second language acquisition: An advanced resource book*. Routledge Applied Linguistics. New York: Routledge.

Dixit, R. Prakash. 1987. In defense of the phonetic adequacy of the traditional term "voiced aspirated". *University of California Working Papers in Phonetics* 67 (June): 103–111.

Dixit, R. Prakash. 1989. Glottal gestures in Hindi plosives. *Journal of Phonetics* 17:213–237.

Dutta, Indranil. 2007. *Four-way stop contrasts in Hindi: An acoustic study of voicing, fundamental frequency and spectral tilt*. University of Illinois dissertation.

Dutta Majumder, Dwijesh, Asoke Kumar Dutta, and Nihar Ranjan Ganguli. 1973. Some studies on acoustic features of human speech in relation to Hindi speech sounds. *Indian Journal of Physics and Proceedings of the Indian Association for the Cultivation of Science* 47:598–613.

Ferguson, Charles A., and Munier Chowdhury. 1960. The phonemes of Bengali. *Language* 36:22–59.

Flege, James Emil, Murray J. Munro, and Laurie Skelton. 1992. Production of the word-final English /t/-/d/ contrast by native speakers of Mandarin and Spanish. *Journal of the Acoustic Society of America* 92:128–43.

Gass, Susan M., and Larry Selinker. 1994. *Second language acquisition: An introductory course*. Hillsdale, N.J.: L. Erlbaum Associates.

Ghosh, Aditi. 2003. *The sounds of Bengali and French*. Kolkata: The Asiatic Society.

Hai, Muhammad Abdul, and W. J. Ball. 1961. *The sound structures of English and Bengali*. East Pakistan: Dacca University Press.

Halle, Morris. 1959. *The sound pattern of Russian*. The Hague: Mouton.

Hammond, Michael. 1999. *The phonology of English*. Oxford: Oxford University Press.

Hancin-Bhatt, Barbara. 1994. Segment transfer: A consequence of a dynamic system. *Second Language Research* 10, no. 3:241–269.

Hayes, Bruce, and Aditi Lahiri. 1991. Bengali intonational phonology. *Natural Language and Linguistic Theory* 9:47–96.

Ingemann, Frances, and Ramawatar Yadav. 1978. Voiced aspirated consonants. In *Papers from the 1977 Mid-America Linguistics Conference*, edited by Donald M. Lance and Daniel E. Gulstad. Columbia, Missouri: University of Missouri.

Islam, Sirajul. 2003. International Mother Language Day. In *Banglapedia: National encyclopedia of Bangladesh*, edited by Sirajul Islam and Sajahan Miah. 10 vols. Dhaka: Asiatic Society of Bangladesh.

Kagaya, Ryohei, and Hajime Hirose. 1975. Fiberoptic electromyographic and acoustic analyses of Hindi stop consonants. *Annual Bulletin, Research Institute of Logopedics and Phoniatrics* 9:27–46.

Kostić, Djordje, and Rhea S. Das. 1972. *A short outline of Bengali phonetics*. Calcutta: Statistical Publishing Society.

Ladefoged, Peter. 1971. *Preliminaries to linguistic phonetics.* Chicago: University of Chicago Press.

Ladefoged, Peter. 1973. The features of the larynx. *Journal of Phonetics* 1:78–83.

Ladefoged, Peter. 1975. *A course in phonetics.* New York: Harcourt Brace Jovanovich.

Ladefoged, Peter. 2005. *Vowels and consonants: An introduction to the sounds of languages*. 2d ed. Oxford: Blackwell.

Ladefoged, Peter, and Ian Maddieson. 1996. *The sounds of the world's languages*. Oxford: Blackwell.

Lado, Robert. 1957. *Linguistics across cultures*. Ann Arbor: University of Michigan Press.

Larsen-Freeman, Diane. 1997. Chaos/complexity science and second language acquisition. *Applied Linguistics* 18:141–65.

Lisker, Leigh. 1986. "Voicing" in English: A catalogue of acoustic features signaling /b/ versus /p/ in trochees. *Language and Speech* 29, no. 1:3–11.

Lisker, Leigh, and Arthur S. Abramson. 1964. A cross-language study of voicing in initial stops: Acoustical measurements. *Word* 20, no. 3:384–422.

Major, Roy. 2002. The phonology of the L2 user. In *Portraits of the L2 user*, edited by Vivian J. Cook. Clevedon: Buffalo Multilingual Matters.

Masica, Colin P. 1991. *The Indo-Aryan languages*. Cambridge: Cambridge University Press.

Mazzoni, Dominic, Matt Brubeck, James Crook, Roger Dannenberg, Vaughan Johnson, Markus Meyer, et al. 2008. *Audacity: A Free, Cross-Platform Digital Audio Editor* (Version 1.2.6) [Computer Program]. Retrieved from http://audacity.sourceforge.net/ June 4, 2008.

Michaels, David. 1973. Sinhalese sound replacements and feature hierarchies. *Linguistics* 107:14–22.

Mikuteit, Simone, and Henning Reetz. 2007. Caught in the ACT: The timing of aspiration and voicing in East Bengali. *Language and Speech* 50, no. 2:247–277.

Odlin, Terence. 1989. *Language transfer: Cross-linguistic influence in language learning*. Cambridge: Cambridge University Press.

Odlin, Terence. 2003. Cross-linguistic influence. In *The handbook of second language acquisition*, edited by Catherine. J. Doughty and Michael. H. Long. Malden, MA: Blackwell.

Pandit, P. B. 1964. Indian readjustments in the English consonant system. *Indian Linguistics* 25:202–205.

Poon, Pamela G., and Catherine A. Mateer. 1985. A study of VOT in Nepali stop consonants. *Phonetica* 42:39–47.

Radice, William. 1994. *Bengali: A complete course for beginners*. Teach Yourself Series. London: Hodder Headline.

Rafique, Ahmed. Ekushey February. 2003. In *Banglapedia: National encyclopedia of Bangladesh*, edited by Sirajul Islam and Sajahan Miah. 10 vols. Dhaka: Asiatic Society of Bangladesh.

Ritchie, William C. 1968. On the explanation of phonic interference. *Language Learning* 18:183–197.

Roach, Peter. 2000. *English phonetics and phonology*. 3d ed. Cambridge: Cambridge University Press.

Selinker, Larry. 1972. Interlanguage. *International Review of Applied Linguistics* 10, no. 3:209–231.

Sharf, Donald J. 1962. Duration of post-stress intervocalic stops and preceding vowels. *Language and Speech* 5:26–31.

SIL International. 2008. *Speech analyzer: A speech analysis tool* (Version 3.0.1) [Computer Program]. Retrieved from http://www.sil.org/computing/speechtools/ June 4, 2008.

Stockwell, Robert P., and J. Donald Bowen. 1965. *The sounds of English and Spanish.* Contrastive Structure Series. Chicago: University of Chicago Press.

Syntrillium Software Corporation. 2000. *Cool Edit 2000* [Computer Program]. Retrieved from http://www.syntrillium.com/. (Note: Syntrillium was acquired by Adobe. On June 4, 2008, this web address was redirected to http://www.adobe.com/products/soundbooth/, where similar software is available.)

Trubetzkoy, N. S. 1969. *Principles of phonology*. Translated by Christiane A. M. Baltaxe. Berkeley, Calif.: University of California Press.

Weinreich, Uriel. 1957. On the description of phonic interference. *Word* 13, no. 1:1–11.

Weinreich, Uriel. 1974. *Languages in contact*. Publications of the Linguistic Circle of New York, no. 1. New York: Linguistic Circle of New York, 1953. Reprint, with a preface by André Martinet, The Hague: Mouton.

Yadav, Ramawatar. 1984. Voicing and aspiration in Maithili: A fiberoptic and acoustic study. *Indian Linguistics: Journal of the Linguistic Society of India* 45, no. 1–4:1–30.