A Cross-Language Study of Voicing Contrasts of Stop Consonants in Asian Languages

by

Katsumasa Shimizu

Thesis submitted for the degree of Ph.D.
University of Edinburgh
1990

ABSTRACT

This thesis examines some of the cross-language differences and similarities of voicing contrasts of stop consonants in six Asian languages and explores ways of explaining their phonetic characteristics. The languages investigated are Japanese, Mandarin Chinese, Korean, Burmese, Thai, and Hindi, and the examination is mainly based on the acoustic analysis of initial stops in these languages.

The thesis consists of ten chapters. Chapter 1 deals with the background for the study, the scope of the study, and general experimental procedure. Some of the articulatory, acoustic, and perceptual characteristics of stop consonants are examined.

Chapters 2 to 7 present the results of acoustic analysis in each language. In each chapter, the general properties of stop consonants and linguistic materials analyzed are presented, and acoustic characteristics such as voice onset time (VOT), fundamental frequency (Fo) and the curve, spectral properties, and onset frequency of the first formant (F1) are examined in detail.

Chapter 8 examines cross-language characteristics based on the acoustic analysis. It can be said that languages in the present study use several acoustic features for distinguishing voicing categories in different ways, and "same" or similar sounds in these languages show some language-specific properties as well as features which are common to many languages. VOT functions for distinguishing voicing categories of stops if they are based on the timing events of glottal and supralaryngeal movements, and if other laryngeal features are involved, other acoustic dimensions are necessary for making a distinction. Furthermore, cross-language characteristics in acoustic dimensions such as Fo and the curve, spectral properties and the F1 onset frequency are examined.

Chapter 9 examines some theoretical issues in cross-language phonetics. A model was proposed in which cross-language differences can be expressed as differences in phonological features. Some generalized phonetic patterns which are shared by these languages are presented, and the underlying articulatory mechanisms and the implications are examined.

DECLARATION

This thesis is my original work and of my own execution and authorship.

ACKNOWLEDGEMENTS

I first wish to express my sincere gratitude to Mr. Alan J. Kemp, my supervisor, for his invaluable suggestions and constant encouragement to complete this research work, without which this work would not have been accomplished. I have benefited on numerous occasions from his guidance and insightful remarks.

I am grateful to my committee members, Ms. Jody Higgs, Ms. Sandy Hutcheson, and Dr. Gerald J. Docherty for their suggestions and encouragement with the study. I am also indebted to Prof. John Laver and Dr. Robert D. Ladd, Jr. for their ideas which I learned from their advanced phonetics theory class and phonology class.

All of the instrumental work in the present study was done at the Phonetics Laboratory, Department of Linguistics, University of Edinburgh. I wish to thank all of those who had assisted and supported me along the way; Norman Dryden, Morag Brown, Irene Macleod, Stewart Smith and Cedric MacMartin. I also wish to thank the informants of the languages investigated; P. V. Dhamija (Hindi), S. Luksaneeyanawin (Thai), and Myint-Su (Burmese).

And above all I wish to thank my family for their love, support, and encouragement.

TABLE OF CONTENTS

ABSTRACT i

DECLARATION ii

ACKNOWLEDGEMENTS iii

TABLE OF CONTENTS iv

LIST OF TABLES viii

LIST OF FIGURES x

CHAPTER 1 INTRODUCTION 1
- 1.1. Introductory Remarks 1
- 1.2. The Scope of the Present Study 3
- 1.3. Review of Literature on Cross-Language Studies ✓ 4
- 1.4. Current View on Voicing of Stops ✓ 7
- 1.5. Acoustic Properties of Voicing Contrasts of Stops ✓ 9
- 1.6. Perceptual Cues for Initial-Stop Voicing 12
- 1.7. Physiological Aspects of the Voicing Contrasts of Stops ✓ 15
- 1.8. Structure of the Present Study 16
- 1.9. General Experimental Procedure 17
  - 1.9.1. Subjects 17
  - 1.9.2. Languages investigated and their voicing categories 17
  - 1.9.3. Acoustic analysis 18

CHAPTER 2 STOPS IN JAPANESE 20
- 2.1. Introduction 20
- 2.2. Experimental Procedure 22
  - 2.2.1. Subjects 22
  - 2.2.2. Linguistic Materials 22
  - 2.2.3. Acoustic Analysis 23
- 2.3. Results 23
  - 2.3.1. Voice Onset Time 23
  - 2.3.2. Fundamental Frequency (Fo) and its Contour 29
  - 2.3.3. Spectral Analysis 34
  - 2.3.4. The Onset Frequency of the First Formant 37
- 2.4. Discussion on Acoustic Analysis 40
- 2.5. Oral Airflow 44
- 2.6. Electropalatographic (EPG) Analysis of Japanese Stops 47
- 2.7. Summary 52

CHAPTER 3 STOPS IN MANDARIN CHINESE 54
- 3.1. Introduction 54
- 3.2. Experimental Procedure 56
  - 3.2.1. Subjects 56
  - 3.2.2. Linguistic Materials 56
  - 3.2.3. Acoustic Analysis 57

3.3. Results 57
    3.3.1. Voice Onset Time (VOT) 57
    3.3.2. Fundamental Frequency (Fo) and its Contour 60
    3.3.3. Spectral Analysis 64
    3.3.4. The Onset Frequency of the First Formant 68
3.4. Discussion 71
3.5. Summary 73

CHAPTER 4 STOPS IN KOREAN 74
4.1. Introduction 74
4.2. Experimental Procedure 76
    4.2.1. Subjects 76
    4.2.2. Linguistic Materials 76
    4.2.3. Acoustic Analysis 77
4.3. Results 77
    4.3.1. Voice Onset Time (VOT) 77
    4.3.2. Fundamental Frequency (Fo) and its Contour 79
    4.3.3. Spectral Analysis 81
    4.3.4. The Onset Frequency of the First Formant 86
4.4. Oral Airflow 87
4.5. Discussion 89
4.6. Summary 93

CHAPTER 5 STOPS AND NASALS IN BURMESE 95
5.1. Introduction 95
5.2. Stops in Burmese 96
    5.2.1. Subjects 96
    5.2.2. Linguistic Materials 96
    5.2.3. Acoustic Analysis 98
    5.2.4. Results 98
        5.2.4.1. Voice Onset Time (VOT) 98
        5.2.4.2. Fundamental Frequency (Fo) and its Contour 100
        5.2.4.3. Spectral Analysis 104
        5.2.4.4. The Onset Frequency of the First Formant 107
    5.2.5. Discussion 108
5.3. Nasals in Burmese 110
    5.3.1. Nasals 110
    5.3.2. Linguistic Materials and Acoustic Analysis 110
    5.3.3. Results 111
    5.3.4. Discussion 117
5.4. Summary 118

CHAPTER 6 STOPS IN THAI 120
6.1. Introduction 120
6.2. Experimental Procedure 121
    6.2.1. Subjects 121
    6.2.2. Linguistic Materials 121
    6.2.3. Acoustic Analysis 123
6.3. Results 123
    6.3.1. Analysis of Waveforms 123
    6.3.2. Voice Onset Time (VOT) 126
    6.3.3. Fundamental Frequency (Fo) and the Contour 128

6.3.4. Spectral Analysis 131
6.3.5. The Onset Frequency of the First Formant 134
6.4. Discussion 137
6.5. Summary 140

CHAPTER 7 STOPS IN HINDI 141
7.1. Introduction 141
7.2. Experimental Procedure 143
7.2.1. Subjects 143
7.2.2. Linguistic Materials 143
7.2.3. Acoustic Analysis 144
7.3. Results 144
7.3.1. Acoustic Portions in Waveforms 144
7.3.2. Voice Onset Time (VOT) 149
7.3.3. Fundamental Frequency (Fo) and its Contour 150
7.3.4. Spectral Analysis 154
7.3.5. The Onset Frequency of the First Formant 157
7.4. Discussion and Summary 161

CHAPTER 8 CROSS-LANGUAGE COMPARISONS 164
8.1. Introduction 164
8.2. Classification of Stops in Each Language 164
8.3. Voice Onset Time (VOT) 165
8.4. Fundamental Frequency (Fo) and its Contour 173
8.5. Spectral Analysis 177
8.6. The Onset Frequency of the First Formant 180
8.7. Discussion 182
8.7.1. Phonetic Aspects between Supralaryngeal and Glottal Events 187
8.7.2. Voicing Categories of Stops 189
8.8. Summary 191

CHAPTER 9 CROSS-LANGUAGE PHONETICS: VOICING CONTRASTS OF STOPS 193
9.1. Introduction 193
9.2. Relation between Phonetic Variation and Phonological Contrasts 193
9.2.1. A Proposal 194
9.2.2. Phonological Features 196
9.2.3. The Phonetic Bases of Phonological Features 199
9.2.4. Implementation by Language-Particular Constraints 201
9.3. Some Generalized Aspects in Phonetic Patterns 203
9.3.1. Phonetic Categories and VOT Ranges 204
9.3.2. Aspiration and VOT Variation 205
9.3.3. VOT on Velar Stops 205
9.3.4. Fo and Initial Stop Voicing 206
9.3.5. Intensity in Spectral Analysis 208
9.3.6. The Onset Frequency of the First Formant 208
9.4. Summary 209

CHAPTER 10 CONCLUSION 210

REFERENCES 215

LIST OF TABLES

2.1 Mean VOT values, s.d. and range of Japanese stops 26

2.2 Mean VOT values for each subject 26

2.3 Mean VOT values as a function of the following vowel 28

2.4a Average Fo value at vowel onset for female group 29

2.4b Average Fo value at vowel onset for male group 29

2.5 Fo values at vowel onset as a function of the following vowel 31

2.6 Mean F1 onset frequency of Japanese stops 37

3.1 List of words in Mandarin Chinese 57

3.2 Mean VOT value and the range in Mandarin Chinese 59

3.3 Mean VOT value as a function of the following vowel 59

3.4 Mean VOT value as a function of tones 60

3.5 Mean Fo at vowel onset 61

3.6 Mean Fo as a function of following vowel 61

3.7 Mean F1 onset frequency in Mandarin Chinese 68

4.1 List of Korean words 77

4.2 Mean VOT value of Korean stops 78

4.3a Mean Fo value at vowel onset 79

4.3b Mean Fo value at steady-state of vowel 79

4.4 Mean onset F1 frequency in Korean stops 86

5.1 List of Burmese words 97

5.2 Mean VOT value and range for each stop 99

5.3 Mean VOT values of voiceless unaspirated stops as a function of tone types 100

5.4 Mean Fo value at vowel onset 100

5.5 Mean F1 onset frequency in Burmese stops 107

5.6 Mean duration of nasal murmur portions in Burmese 114

5.7 Mean Fo value at vowel onset 114

6.1 List of Thai words 123

6.2 Mean VOT values of Thai stops 126

6.3 Mean VOT values as a function of the following vowel 127

6.4 Mean VOT values as a function of tones 127

6.5 Mean Fo value at vowel onset 128

6.6 List of words examined for Fo contours 129

6.7 Mean onset F1 frequency in Thai stops 134

7.1 A list of Hindi words 144

7.2 Mean duration of breathy portion 148

7.3 Mean VOT values for Hindi stops 149

7.4a Mean Fo at vowel onset 151

7.4b Mean Fo at breathy portion 151

7.4c Mean Fo at vowel steady state portion 151

7.5 Mean onset F1 frequency in Hindi stops 157

8.1 Mean VOT and range in six languages 167

8.2 One-way ANOVA results: Language effect on VOT 172

8.3 Mean Fo for major categories at vowel onset 174

8.4 Mean F1 onset frequency of voicing categories in six languages 181

8.5 The use of acoustic features for differentiating voicing categories in six languages 183

9.1 Articulatory and acoustic correlates of phonological features 201

LIST OF FIGURES

2.1 Waveform of [ka] and [ga] in Japanese 25

2.2 Normalized fundamental frequency patterns of vowels following voiced and voiceless stops in Japanese 33

2.3(a-c) Spectra of CV syllables in Japanese 35

2.4(a,b) Examples of smoothed power spectra in Japanese stops 36

2.5a Formant patterns of Japanese stops [ka] and [ga] 38

2.5b Examples of wide-band spectrograms of [ka] and [ga] 39

2.6 Recordings of airflow in Japanese stops 46

2.7(a-c) EPG charts showing tongue-palate contacts 51

3.1 Example of waveform of two types of stops in Mandarin Chinese 58

3.2 Fo curves of voiceless unaspirated and aspirated stops 63

3.3(a,b) Examples of power spectra sampled at the time window after release of stops in Mandarin Chinese 66

3.4 Examples of smoothed power spectra of two types of stops in Mandarin Chinese 67

3.5 Examples of formant patterns of Mandarin Chinese stops 70

4.1 Normalized Fo contour of three types of Korean stops 80

4.2 Examples of smoothed power spectra of Korean stops 83

4.3 Power spectra sampled at the time window after release 84

4.4 Waveforms and intensity patterns of Korean stops 85

4.5 Inkwriter traces of airflow for Korean stops 88

4.6 Waveform and Fo contour of [p*ul] 94

5.1 Averaged Fo contours of a three-way of Burmese stops 103

5.2 Examples of smoothed spectra of Burmese stops 105

5.3 Power spectra of Burmese stops sampled at the time window after release 106

5.4 Waveforms and intensity patterns of [hna] and [na] 112

5.5 Wide and narrow-band spectrograms of [hna] and [na] 113

5.6 Smoothed spectra of voiced-voiceless nasals in Burmese 116

6.1 Examples of waveforms of three-way Thai stops 125

6.2 Averaged Fo contours of a three-way Thai stops 130

6.3 Examples of smoothed spectra of three-way Thai stops 132

6.4 Examples of short-time power spectra of three-way Thai stops 133

6.5 Examples of formant patterns of Thai stops 136

7.1 Examples of waveforms of four-way Hindi stops 147

7.2 Averaged duration of acoustic portions of Hindi stops 148

7.3 Fo curves of four-way distinction of Hindi stops 153

7.4a Smoothed power spectra of four-way Hindi stops 155

7.4b Examples of power spectra of four-way Hindi stops 156

7.5 Examples of formant patterns of Hindi stops 160

8.1 Voice onset time distribution based on the mean VOT values 170

8.2 Schematized Fo curves in Japanese, Korean and Hindi 176

8.3 Examples of power spectra of Japanese and Hindi stops 179

Chapter 1

Introduction

1.1. Introductory Remarks

Why do languages differ in the voicing categories of stops and how do they differ from each other ? These questions have always been in my mind in my investigation of the phonation types of stops. English and Japanese, like many other languages, have a basic distinction between voiced and voiceless stops, while Hindi, like other Indo-Aryan languages, has four types of voicing categories for the same point of articulation. Each language has its own particular inventory of consonants and syllable structure, and has its own characteristics in its sound system. A major task for linguists and phoneticians has been to examine what such characteristics are and how language specific properties of the same or similar sounds differ within and across languages. The present study focuses on the cross-language voicing contrasts of stops and attempts to examine the cross-language characteristics of voicing categories of stops in six Asian languages.

Voicing contrasts of stops have been one of the major topics in phonetics and phonology, and have been extensively studied for many languages. Through a number of experimental studies, it is now fairly well understood what the articulatory and acoustic bases are for the voicing contrasts of major languages, though the specific aspects of voicing in individual languages remain to be clarified. There are several types of voicing contrasts of stops in the languages of the world. One of the basic voicing distinctions is that between voiced and voiceless and it is found in a number of languages. It is significant that almost all of the world's languages, except a small percentage, have this kind of voicing distinction .[1] There are also other kinds of voicing distinctions which involve such voice qualities as creaky or breathy voice. These phonation types are also found in many languages, though they are less common than the voicing contrast between voiced and voiceless categories. Thus, the voicing contrasts of stops of languages in the world vary in several ways. One of the recent trends in speech research has been to characterize the details of these linguistically significant units by various objective measures in speech technology, and to exploit the results to explain the phonological processes and other aspects of the languages, as exemplified in experimental phonology by Ohala and Jaeger (1986). The introduction of experimental results into phonology serves to clarify such phonologically-relevant issues as phonological and phonetic "definitions" of voicing, the naturalness of voicing, allophonic variation of voicing, and distinctive features.

It is also known that the "same" voicing categories are differently realized across languages. There are many examples which show cross-language differences in the voicing contrasts of stops. For instance, the voiced stops /b, d, g/ in English are phonetically realized as voiceless unaspirated stops [b̥, d̥, g̊] in initial position and these realized forms are different from the realizations of voiced stops in French. Flege and Eefting (1988) examined how English and Spanish speakers realized word-initial /t, d/ in their languages. They mentioned that the two languages use different phonetic categories to implement the contrast between /t/ and /d/, and the two language speakers showed a distinct difference in the onset of voicing for alveolar stops. /t/ and /d/ in Spanish are realized as [t] and [d] with VOT values of approximately 20 and -80 ms, respectively. On the other hand, /t/ in English is realized as [tʰ] with VOT values of approximately 80 ms, and /d/ is realized as either [d] or [t] with the values of approximately -80 to 20 ms (Flege and Eefting 1988:730). Furthermore, it is common to find that foreign language learners may have to adjust their voicing

---

[1] It is known that Hawaian, a native language in Hawaii, does not have a phonological contrast between voiced and voiceless stops. According to Maddieson (1984:26), 15.8 % of UPSID languages have one stop series. He mentions ,"All languages have at least one series of stops, but two is the most common, just over 50 % having that number." (p.26).

characteristics of stops in learning foreign languages, when the voicing of stops in mother and target languages are realized differently; for example, it is known that Japanese learners of English have to adjust the voicing of their stops to improve their English proficiency. These indicate that the "same" or similar voicing categories in languages are manifested differently across languages in some cases and there exists cross-language variation in the phonetic realization of voicing categories of stops. This implies that languages may have language-specific articulatory and acoustic targets for speech sounds which are represented by the same phonetic symbols. On the other hand, while admitting the existence of such differences among the "same" voicing categories, we can think that there should be some similarities in the voicing categories which are common to many languages and may be constrained by universal aspects of language. We can expect that variation does not take place without limit and is within a range which is functionally controllable. We can also consider that properties of speech sounds are constrained by physiological and perceptual factors. Therefore, the problems here are what specific aspects of voicing characteristics vary across languages and to what extent they vary. Cross-language examination of voicing contrasts will shed light on which phonetic aspects are fundamentally significant in characterising stop voicing, and the studies supported by objective investigation will contribute to the clarification of the production mechanisms and the characteristics for voicing distinction.

1.2. The Scope of the Present Study

The present study is mainly concerned with a cross-language analysis of voicing contrasts of stops in six Asian languages; Japanese, Mandarin Chinese, Korean, Burmese, Thai, and Hindi. These languages are selected since they provide a good example of voicing contrasts of stops and the contrasts involve various vocal fold adjustments. These languages are divided into three groups depending on the phonological contrasts of voicing in their languages. Japanese and Mandarin Chinese are two-category languages, while Korean, Burmese, and Thai are three-category languages. Hindi, like other Indo-Aryan languages, is a four-category language. The study itself is descriptive and synchronic, and the data presented in the following chapters are mainly acoustic in nature.[2]

The main issues of the investigation in the present study are as follows:

(1) What are the phonetic characteristics of voicing categories of stops in these languages? Is it possible to make generalizations about the characteristics of the major voicing categories?

(2) How do the language-specific properties of voicing categories differ from one another across languages? For example, are voiced stops /b, d, g/ in Japanese similar to or different from those in Burmese or Hindi in their phonetic realization ? If they are different, how are they different from each other and how is the variation represented ?

(3) How are the phonetically different manifestations of the "same" voicing categories incorporated into the phonetic theory? Is a single phonetic feature [voice] sufficient to capture the generalizations on the constraints within and across languages? If not, what other features are needed ?

1.3. Review of Literature on Cross-Language Studies

The study of voicing contrasts of stops has attracted the attention of many investigators for many years, and there has been a great deal of effort to examine phonetic characteristics of stops in a number of languages. In past decades, phoneticians and phonologists have made considerable efforts to discover what the language-specific properties of the stops are, and how they are incorporated into the theoretical frameworks of phonetics and phonology. The study has been made from three major areas of research: acoustic study, physiological study using electromyography (EMG) and/or fiberoptic observations, and

---

[2] For some languages such as Japanese and Korean, aerodynamic and/or electropalatographic data were added to examine properties such as airflow rate and tongue contact areas.

auditory study, and the topic of voicing contrasts of stops still continues to be a subject in these areas.

There have also been some efforts to characterize the cross-language variation of the same or similar stops in different languages. One of the most significant contributions in the cross-language study of voicing contrasts of stops is Lisker and Abramson (1964). They examined the voicing categories of eleven languages to see how voice onset time (VOT) serves to separate the stop categories. They classified these languages into three category groups, depending on phonological contrasts the languages use: (1) two-category languages (American English, Cantonese, Dutch, Hungarian, Puerto Rican Spanish, and Tamil); (2) three-category languages (Korean, Eastern Armenian, and Thai); (3) four-category languages (Hindi and Marathi). They measured the VOT values in the initial stops of these languages, and found that "the stop categories fall generally into three ranges - one from about -125 to -75 msec, one from zero to +25 msec, and a third from about +60 to +100 msec"(Lisker and Abramson, 1964:403), and the categories are distributed in the ranges centering at -100, +10 and +75 msec. These three categories are termed voicing lead, short-lag and long-lag. Furthermore, they mentioned that VOT does not always function appropriately in separating the voicing categories from each other in some languages, and pointed out that VOT is not sufficient for distinguishing stops in Korean, Hindi, and Marathi; in Korean it fails to separate tense, glottalized stops from lax, slightly aspirated stops; in Hindi it also fails to separate voiced unaspirated stops from the voiced aspirated stops which are commonly known as breathy voiced stops. These observations on Korean and Hindi stops are also made by Han and Weitzman (1970) and Benguerel and Bhatia (1980).

Lisker and Abramson's study is certainly important in establishing VOT as a measure of voicing contrasts of initial stops, and contributed to the clarification of the timing dimension between the onset of voicing and articulatory release in various languages. Other relevant studies (Flege, 1979; Keating et al. 1983) on VOT of stops in various languages support their view on three basic ranges of VOT. It should be noted, however, that their study mainly dealt with the timing dimension of voicing and has not dealt with other phonetic aspects which are characteristic of voicing contrasts of stops.

More recently, Keating et al. (1983) carried out an investigation of allophonic variation for voiced and voiceless stops in 51 languages and examined how the voicing categories vary contextually in these languages. They pointed out that all the languages in the investigation have at least some kind of voiceless unaspirated stops in all contexts and there exists a trend in languages "to maintain aspirated /p, t, k/ before more stressed vowels, and to voice medial /b, d, g/" (Keating et al. 1983:283). They also pointed out that there are differences between two similar languages (English and Swedish) with respect to closure duration and suggested that these variations may be accounted for by universal principles as found in an aerodynamic model.

Furthermore, UCLA Working Papers in Phonetics (1987) is a good collection of studies on phonation types in various languages which were carried out at the UCLA Phonetics Laboratory. In this issue, Dixit (1987) examined the mechanisms of glottal adjustments of Hindi stops by using photo-electric glottographic techniques, and compared the results with those in other languages. He specifically examined the timing of the glottal gesture in relation to the articulatory closure and opening and the degree of glottal opening for four types of voicing of stops in Hindi: voiced unaspirated plosives, unvoiced unaspirated plosives, voiced aspirated plosives, and unvoiced aspirated plosives. Based on the comparison of the results with others, he made a detailed description of laryngeal mechanisms for four types of stops, and indicated that the observations made for Hindi stops are valid for the similarly labelled stops in other languages.

Meanwhile, there have been quite a few studies on voicing contrasts of stops of individual languages. Some of them are as follows: Danish (Hutters, B., 1985), English (Lisker and Abramson, 1967; Weismer, G., 1979); Finnish (Suomi, K., 1980); French (Caramazza, A. and

Yeni-Komshian, G., 1974); German (Haag, W. K., 1979); Nepali (Poon, P. G. and Mateer, C. A. (1985); Polish (Keating, P., 1984a); and Spanish (Williams, L., 1977). These studies have dealt with voicing contrasts of stops from several angles of investigation. Among these studies, Hutters (1985) examines vocal fold adjustments in Danish aspirated and unaspirated stops using electromyography (EMG) and photo-electric glottography techniques, and mentions that two types of stop are produced by different types of glottal gesture rather than by different timing in glottal and supraglottal events. Suomi (1980) investigates the stops in both English and Finnish and compares the VOT values of Finnish stops with those of English stops. Haag (1979) examines VOT in German stops and its relationship to place of articulation, and shows that there is a correlation between place of articulation and the magnitude of VOT. Furthermore, Williams (1977) investigates initial-stops in several Spanish dialects and the cues distinguishing the voicing categories in a perception study, and provides support for VOT as a measure for describing differences across dialectal groups in Spanish. Studies of stop consonants in the six Asian languages in the present study vary between languages in scope and direction. Each chapter presents a review of studies of stop consonants in each language. There has been considerable work on Korean and Hindi stops, but there have been no systematic studies on other languages, in particular, little has been made on Mandarin Chinese and Burmese stops.

1.4. Current View on Voicing of Stops

In the articulation of speech sounds it is generally understood that voiced speech sounds are produced with the vibration of the vocal folds and voiceless sounds are produced with the vocal folds being apart. The vibration of the vocal folds requires delicate adjustments of laryngeal and aerodynamic mechanisms in the vocal organs. Studies on the mechanism of the vocal folds are quite numerous and extensive, and have been done from anatomical and aerodynamic points of view. Therefore, the understanding of their details requires comprehensive knowledge in these fields.

It is understood that when the vocal folds come together, they are set to vibrate when the aerodynamic conditions are met for vibration. It has been examined for a long time how the vocal folds are set to vibrate and how the respiratory system is coordinated with the laryngeal one. Among several theories on the vibration of vocal-folds, the currently accepted theory is the so-called myoelastic aerodynamic one which is considered to have combined the two hypotheses of the myoelastic theory and the Bernoulli Effect.[3] Furthermore, Hirano and Kakita (1985) clarified the details of the mechanical properties of the vocal folds and found that they are not a vibrator of uniform structure but a layer - structured vibrator. Although recent studies of the mechanisms of the larynx have advanced knowledge on the vibration of the vocal-folds, the full description of their mechanism still remains to be clarified.

As indicated above, the main difference between voiced and voiceless categories lies in the area of vocal fold adjustments. The difference can be briefly exemplified in the production of /p, b/ in Japanese. In producing initial /b/, the articulatory closure is made at the lips, and while the closure is maintained, the vocal folds are set to vibrate and the folds continue vibrating during the closure. The oral pressure during the closure increases and the transglottal pressure difference decreases. After a very brief period, the closure is released and a brief transient burst occurs, and air flows in the tract, exciting the resonance of the vocal tract. Likewise, in producing initial /p/, the articulatory configuration is almost the same as that for /b/, but the vocal folds are kept apart. It is considered that the subglottal pressure is the same as for /b/, but oral pressure during the closure rises more quickly than for /b/.[4] After a brief period, the closure is released and at the time of release there is a strong noise from the turbulence of the air-flow. When the vocal folds are set to close, vibration of the vocal folds occurs for the following vowels. What is important in the

---

[3] See Borden and Harris (1980:74-89).

[4] Pickett (1980:133) mentions as follows: "The subglottal pressure curve too is the same in general. However, in the mouth pressure there is an important difference during the occlusion : the pressure rises slowly for [b] but rapidly for [p]. This is because of the difference in vocal fold position during the occlusion."

production of voicing contrasts of stops, therefore, is the state of the vocal folds, air stream in the sub- and supra-glottal cavities, articulatory timing of closure and opening in relation to laryngeal gestures. Considerable efforts have been made to describe the physiological and physical aspects of the larynx. In the current theoretical framework, it is known that the main difference between voiced and voiceless stops is in the state of vocal folds adjustments, but there are also some differences in other aspects of articulatory mechanisms. Some research (Trullinger and Emanuel, 1983) suggests that voiceless stops are produced with higher airflow and a slightly longer period of contact in the oral cavity than voiced stops. Furthermore, the position of the larynx is slightly lower for voiced stops than for voiceless ones. All these changes in sub- and supralaryngeal cavities are major causes of acoustic differences and can be reflected in several acoustic dimensions.

1.5. Acoustic Properties of Voicing Contrasts of Stops

As mentioned above, the production of voiced and voiceless stops involves delicate adjustments in the state of the vocal folds, air stream, and articulatory timing in relation to glottal closure and opening. The differences in these aspects are variously manifested in several acoustic properties, and their changes lead to a complicated acoustic realization. There have been a number of studies on acoustic properties of voicing contrasts of stops in many languages, especially in English, and it is well known that the phonetic properties cueing the voicing contrasts vary according to the phonetic context in which the stop occurs. The acoustic properties are characterized in temporal and spectral terms. For the word-initial stops, the following properties are known to be relevant:

1. Voice Onset Time (VOT)
2. Fundamental Frequency (Fo) at Vowel Onset and the Fo contour
3. Release Burst Intensity
4. Formant One (F1) Onset Frequency and Transition Period

VOT, the Fo contour and the F1 transition period are temporally changing properties, while release burst intensity and F1 onset frequency are considered to be spectrally changing ones. VOT is defined as the time interval between the onset of voicing and articulatory release of the stop (Lisker and Abramson, 1964). It has been extensively studied and is known to be an efficient cue to the voicing contrast of stops. VOT is acoustically manifested as the time interval from the beginning of release burst to the onset of periodicity of voicing. As the definition of VOT implies, VOT is a single timing dimension of the onset of voicing to articulatory stop release, and, as mentioned above, can be divided into three domains for voicing categories: voicing lead, simultaneous or short lag, and long lag. These domains roughly correspond to the phonetic and phonological categories such as voiced stops, voiceless unaspirated stops, and aspirated stops. A problem in measuring the duration defined as VOT is that there is a difficulty in defining the onset of periodicity; for example, there is a difference between the beginning of the first formant and other higher formants in some cases. This problem was raised by Fischer-Jørgensen and Hutters (1981), and they pointed out that the beginning of voicing is sometimes ambiguous; it may be marked by the onset of periodicity of the F1 or the onset of the higher formants. Both authors show that periodic excitation is evident in the lower harmonics before it goes to higher ones. The implications of these findings are taken into consideration in identifying the voicing cues in speech perception.

The second acoustic property is Fo at vowel onset and the Fo contour. It is known that voicing contrasts of initial stops are closely related to pitch perturbations of the following vowels. These trends are well documented in the study of tonogenesis in South-East Asian languages. As Hombert et al. (1979) pointed out, voiceless stops are associated with a higher pitch range, and voiced stops with a lower pitch range. The difference between the two pitch ranges is considered to be one of the properties of the voiced - voiceless distinction. It is also known that there is a difference in the Fo contour from the onset of the vowel to the steady-state portions of the vowel. There are two types of hypothesis as to the phonetic reasons for the characteristic Fo perturbations; namely aerodynamic and vocal-cord tension hypotheses. Hombert et al. (1979) state as follows:

> "The aerodynamic hypothesis runs as follows: During a voiced stop, oral pressure gradually builds up, thus decreasing the pressure drop across the vocal cords---which in turn decreases the Fo. Upon the release of the stop, the pressure drop returns to normal, producing an initially low and rising Fo contour after voiced stops. In the case of voiceless stops (particularly aspirated ones), the airflow past the vocal cords is supposedly very high upon release, creating a higher-than-normal Bernoulli force---which will draw the vocal cords together more rapidly, and thus increase the rate of their vibration at vowel onset. As the airflow returns to normal, the Fo will too. Thus, after voiceless stops, the Fo contour will be initially high and falling." (Hombert et al.,1979:42)

> "The basic assumption of the vocal-cord tension hypothesis is that, in the course of making the voiced vs. voiceless distinction on stops, vocal-cord tension is changed so as to affect the Fo of adjacent vowels. Halle & Stevens 1971 suggest that these intrinsic variations are the result of horizontal vocal-cord tension: the vocal cords are presumably slack in order to facilitate voicing during voiced stops, and stiff in order to inhibit voicing during voiceless stops; and these vocal-cords states spread to adjacent vowels, affecting their Fo."(Hombert et al. 1979:42)

Hombert et al. (1979) discuss these hypotheses in detail and mention that these are inadequate for explaining some aspects of pitch perturbation; i.e. they are not capable of explaining the fact that postvocalic consonants do not exert the same effect on the Fo curve as prevocalic ones. Based on some experimental data (Ewan & Krones, 1974), they propose that larynx vertical height is more responsible for the correlation between Fo and voicing contrasts of stops. They mention:

> "...the vertical position of the larynx differs for voiced and voiceless stops. ...Their findings of higher larynx position for voiceless as opposed to voiced stops - coupled with the well-documented fact that in normal speech, other things being equal, Fo is positively correlated with larynx elevation - is compatible with the 'vertical tension' hypothesis.
>
> Additional evidence in favor of this hypothesis is the fact that, in general, the difference in larynx elevation between the two stop types is greatest at the end of the consonant closure, and this difference persists well into the following vowel." (Hombert et al., 1979:43-44)

The third acoustic property is release burst intensity. It is considered that there is a difference in the articulatory force between voiced and voiceless stops, that is, the rate of airflow. In the articulation of voiceless stops, the pressure in the oral cavity is considered to be very high, a strong airflow goes out when the closure is released, and the high pressure and turbulence of the airflow at the release results in higher intensity in the spectrum analysis. On the other hand, in the articulation of voiced stops, the pressure in the oral cavity is not as high as in voiceless stops, because the vocal folds are closed and are vibrating, thus less strong airflow goes out upon the release of the closure. This results in a lower intensity level in the spectrum analysis. Therefore, the difference in the supralaryngeal pressure is considered to be the reason for the difference in intensity level at the time of the release burst.

Lastly, it is understood that there is a difference in the onset frequency of the first formant; it tends to be generally higher in voiceless stops than in voiced ones (Summers, 1988). For the voiceless stops, since the release of the stop closure causes the burst and the turbulent airflow, the F1 transition does not appear for the initial brief period until the vocal folds are excited for vibration of the following vowel. For the voiced stops, on the other hand, the release portion shows a rather gradual transition to the first formant excitation, since the vocal folds are closed and set to vibrate for the following vowel. It does not show strong frication or aspiration.

1.6. Perceptual Cues for Initial-Stop Voicing

There have been a number of experiments in speech perception to examine perceptual cues for the identification and discrimination of initial-stop voicing in several languages, mainly in English. As shown in the previous section, there are several temporal and spectral cues for distinguishing voiced stops from voiceless ones. Specifically, the release burst, F1 onset characteristics (F1 formant transition and intensity) and degree of aspiration provide information to listeners on initial-stop voicing. It is known, however, that there may be no one-to-one relation between these acoustic cues and the voicing features. In the preparation of synthetic stimuli, each cue can be emphasized and can be used as the main cue to voicing. Among these cues, the one which has been most systematically studied is so-called F1 cutback in which the F1 transition portions are systematically removed, and it is pointed out that the stimuli whose transition portions are removed are heard as voiceless, while those containing the F1 transition by means of noise excitation are heard as voiced (Liberman et al., 1958). The systematic cutback of the first formant relative to the higher formants is considered to correspond to the VOT dimension, and the synthetic stimuli in which the F1 transition is appropriately manipulated have been widely used to examine the perceptual relevance of major categories of stop consonants. Lisker and Abramson (1970) examined the identification modes of English, Spanish, and Thai speakers by using synthetic speech continuum with a systematically varying VOT dimension. Further, Shimizu (1977) examined the identification mode of initial-stops by Japanese speakers using the similar synthetic continuum prepared by a parallel resonance synthesizer. The phonetic boundary in the perception of synthetic stimuli is usually defined as that point on the stimulus scale which would receive 50 % responses from either category involved, and the identification values for voiced and voiceless stops in English, Spanish, Japanese and Thai are shown as follows:

Boundary Value for Initial-Stop Voicing for English
Spanish, Japanese and Thai Speakers in Perception (ms)

	English	Spanish
/b - p/	+20 - +30	+10 - +20
/d - t/	+30 - +40	+20 - +30
/g - k/	+30 - +40	+20 - +30

	Japanese
/b - p/	+15 - +20
/d - t/	+20 - +30
/k - g/	+25 - +30

	Thai
/b - p/	-20
/p - p^h/	+40
/d - t/	-10
/t - t^h/	+40 - +50
/k - k^h/	+40 - +50

The above figures indicate the range in which phonetic boundaries lie in the identification curves of voiced and voiceless stops. The perceptual value for distinction between voiced and voiceless stops by Japanese is similar to that by Spanish. It is noticeable that Japanese and Spanish speakers have similar characteristics in the perception of voicing features. From the above data, it appears that if languages have a binary distinction of voicing features, the perceptual boundaries will lie in the VOT values from +20 to +40 ms.

In experiments on speech perception, it was also found (Lisker and Abramson, 1970) that the perception of either category, i.e., voiced or voiceless category, would go into abrupt shifts from one category to another in identification and discrimination, and this mode of perception is called categorical perception along a continuum of the systematically manipulated F1 in relation to the higher formants. The reason for this categorical mode of perception has been discussed in detail in Studdert-Kennedy (1974) and Pickett (1980), and it is known that listeners are poor in discriminating the stimuli which are considered to be within the same phonemic category, whereas they are good in discriminating the stimuli which are considered to belong to the different phonemic categories. Categorical perception shows that listeners divide a physical continuum into sharply defined categories and assign names to them, and reflects an importance of phonetic categories of speech sounds.

1.7. Physiological Aspects of the Voicing Contrasts of Stops

There have been a number of physiological studies using fiberoptic filming, photo-electric glottography, and electromyography (EMG) on the voiced - voiceless distinction of obstruents, and these have contributed to the understanding of the underlying mechanisms of the larynx and the supralaryngeal mechanisms. Although the present study is mainly concerned with the acoustic properties of stops, it will be important to examine what has been done in the fields of physiology research.

Physiological research has been carried out to examine the glottal stricture, timing coordination of several articulatory systems, muscle activities, and air-stream mechanisms in languages such as Hindi, Japanese, Korean, Swedish, and English, and has provided us with a number of interesting findings on laryngeal and supralaryngeal activities. The basic physiological mechanisms for the voiced - voiceless distinction, as mentioned earlier, relate to the state of the glottis; it is open for voiceless and closed for voiced stops, and the vocal folds are positioned with various degrees of separation for the various phonation types. As to the opening of the glottis, it is found that for voiceless stops it is greater in the initial position than in the medial one for the same consonant (Sawashima and Niimi, 1974). Furthermore, it has been observed through the examination of such languages as Hindi (Kagaya and Hirose, 1975), Korean (Kim, 1965; Kagaya, 1974) and Mandarin Chinese (Iwata and Hirose, 1976) that voiceless aspirated stops require a greater opening of the glottis than voiceless unaspirated stops, and the consonantal release almost coincides with the moment when the glottis is maximally opened.

EMG studies have shown that two intrinsic muscles, the posterior cricoarytenoid (PCA) and interarytenoid (INT), are most relevant for the distinction between voiced and voiceless obstruents, and these two muscles work in a reciprocal way. Although it is reported that there are differences between subjects in the activities of these muscles, INT activity increases for voiced stops (with the decrease of PCA activity), and it decreases for voiceless stops (with the increase of PCA activity). A detailed analysis has been made of the voiced - voiceless distinction in Japanese stops. Furthermore, it has been one of the issues in physiology research whether there is physiological evidence for the laryngeal features proposed by Halle and Stevens (1971), and it seems that physiological findings on tenseness of intrinsic muscles are at the present stage still not very consistent.

In connection with EMG studies, Lŏfqvist and Yoshioka (1981) carried out combined experiments of EMG, fiberscopic filming and aerodynamic records on voiceless sounds in several languages, and suggested that voiceless obstruents are produced by the coordination of several articulatory systems. They mentioned:

> "The laryngeal gestures are tightly coordinated with supralaryngeal events to meet the aerodynamic requirements for producing a signal with a specified acoustic structure."(Lŏífqvist and Yoshioka, 1981:29)

The timing of the articulatory events in the laryngeal and supralaryngeal systems is quite important to the understanding of the underlying mechanisms of production of obstruents, and in particular variations of timing are closely related to the production of voicing contrasts of stops.

1.8. Structure of the Present Study

The present study is to examine the cross-language variation of voicing contrasts of stops in several Asian languages and attempts to describe the characteristics of the variation. It also attempts to explore a way to explain the variation and characteristics systematically. The study is organized as follows: Chapters 2 to 7 present the analysis of individual languages in some detail, some aspects of stops in each language, the results of the acoustic analysis, and an examination of the results. Chapter 8 examines the cross-language comparison in acoustic dimensions such as VOT, Fo and its contour, spectral characteristics, and the onset F1 frequency. In particular, a one-way ANOVA was applied to VOT values to examine language effect, and the implication for classification of phonetic categories is also discussed. Chapter 9 discusses some theoretical aspects of cross-language phonetics and proposes a model as to how surface phonetic variations across languages can be explained in the relation between phonetics and phonology.

1.9. General Experimental Procedure

This section presents some general information relevant to the present experiment to give an overview of the experimental procedure.

1.9.1. Subjects

Subjects in the present experiments are native speakers of each language under investigation and are speakers of the "standard" dialect in each language. Most of them are professional - postgraduate students in Linguistics and Applied Linguistics, visiting scholars to the Department of Linguistics, some of them are lecturers teaching their native language or phonetics. The number of subjects in each language is as follows:

Language	Subject
Japanese	6(3 female, 3 male speakers)
Mandarin Chinese	3(3 male speakers)
Korean	3(3 male speakers)
Burmese	2(1 female and 1 male speakers)
Thai	2(2 female speakers)
Hindi	3(3 male speakers)

1.9.2. Languages investigated and their voicing categories

The languages selected for investigation are all Asian languages: Japanese, Mandarin Chinese, Korean, Burmese, Thai, and Hindi. As mentioned above, these languages are classified into three types, depending on the phonological contrast of voicing they use: two-category languages (Japanese, Mandarin Chinese), three-category languages (Korean, Burmese, and Thai), and a four-category language (Hindi). Although they are all Asian languages, they are not genetically or etymologically related to each other, and their voicing categories can be shown as follows:

Japanese	voiced stops /b, d, g/ voiceless stops /p, t, k/
Mandarin Chinese	voiceless unaspirated stops /p, t, k/ voiceless aspirated stops /pʰ, tʰ, kʰ/
Korean	voiceless tense stops /p*, t*, k*/ voiceless lax stops /p, t, k/ voiceless aspirated stops /pʰ, tʰ, kʰ/
Burmese	voiced stops /b, d, g/ voiceless unaspirated stops /p, t, k/ voiceless aspirated stops /pʰ, tʰ, kʰ/
Thai	voiced stops /b, d/ voiceless unaspirated stops /p, t, k/ voiceless aspirated stops /pʰ, tʰ, kʰ/
Hindi	voiced stops /b, d̪, g/ breathy voiced stops /b̤ʱ, d̪̤ʱ, g̤ʱ/ voiceless unaspirated stops /p, t̪, k/ voiceless aspirated stops /pʰ, t̪ʰ, kʰ/

Word lists were prepared containing minimal or near minimal contrasts of an initial consonant in these languages, and the recording was made in the sound-proofed recording studio, Phonetics Laboratory, Department of Linguistics, University of Edinburgh.

1.9.3. Acoustic Analysis

Acoustic analysis of the recorded materials was made through the ILS software package and KAY Sonagraph 7800 in the Phonetics Laboratory, University of Edinburgh. Test materials were, after 5 kHz low-pass-filtering, digitized at a sampling rate of 10 kHz. The digitized materials were stored for reviewing and listening in a sampled file, and were analyzed at 50 points per frame by linear prediction and Fo extraction. At each point, Fo frequencies, formant frequencies, bandwidths, and amplitude for the lower three formants were obtained. The measurement of such durations as onset of voicing and consonant closure was made by manually positioning two cursors in the display of the waveform in the sampled file. Furthermore, energy spectra were sampled in the short period (25 - 30 ms) immediately following the consonantal release to examine the spectral characteristics of the voiced - voiceless distinction. For some of the utterances, wide-band and narrow-band spectrograms were prepared for confirming and identifying some of the ILS outputs.

Chapter 2

Stops in Japanese

2.1. Introduction

Japanese has two types of voicing contrast for stops: voiced stops /b, d, g/ and voiceless stops /p, t, k/, and is a two-category language in the classification of voicing contrast of consonants. The difference between the two categories is one of voicing, not aspiration or breathiness. Japanese stops are said to be "typically" voiced and voiceless, as described by Martin (1954). It is known that voiceless stops are usually slightly aspirated in initial position. These stops occur in the initial and medial positions in the word and are contextually influenced in various phonetic environments. The "syllabic" structure in Japanese mainly consists of CV, V, or CGV (C=consonant, V= vowel, G=glide), and consonant clusters are not allowed other than geminates in the medial position.[1] Furthermore, as a phonotactic constraint, stops /p, b, k, g/ contrast in word-initial position before five vowels /i, e, a, o, u/, but alveolar stops /t, d/ contrast only before three vowels /a, e, o/. It is known that /t, d/ before /i, u/ are subject to affrication, and are phonetically realized as [tʃ, dʒ, ts, dz]. High vowels /i, u/ tend to influence the phonetic realization of preceding consonant and result in palatalization and labialization of the consonants, and these affricated consonants of /t, d/ before /i, u/ are considered to be conditioned allophones of /t, d/ (Wenck, 1966:17). There are several phonological processes acting on stops in Japanese. One of the well-known processes is so-called "rendaku", and this is a common process by which the initial voiceless consonant of the second word is turned into the voiced counterpart across the word or compound boundary. The other one is a nasalization of /g/; i.e., /g/ is often nasalized to [ŋ] in the intervocalic position and the change is a phonetically conditioned one, though it is subject

---

[1] The definition of syllable in Japanese is slightly different from the one generally understood in English. The syllable is often considered in connection with the mora, the prosodic unit in Japanese.

to dialectal variation. These processes have been major topics in Japanese phonology for a long time, and have been extensively studied by many researchers.

The acoustic properties of Japanese stops have been studied to some extent. Han (1961) examined the voicing contrast by spectrographic analysis and presented some measurements of the onset frequencies of the formants in CV sequences in a limited scale. Itoh et al.(1979) examined the voice onset characteristics of stops uttered by normal and apraxic subjects, and presented the results on VOT values of /te, de, ke, ge/ in initial position. More recently, Shimizu (1979) examined the voicing characteristics as a part of a study to examine the relationship between speech production and perception, and presented the results on the measurements of VOT for six stop consonants followed by the vowel /a/. On the other hand, physiological aspects of voicing contrasts of stop consonants in Japanese and other languages have been considerably studied. Sawashima, Hirose and their co-workers have extensively examined the laryngeal features by the use of electromyography (EMG) and electro-fiberoptic techniques and have clarified the physiological details of laryngeal adjustment for voicing of consonants (Hirose et al. 1978; Yoshioka et al. 1982, Hirose et al. 1985; Yoshioka et al. 1986). Through their studies, they found that there is a difference in the physiological condition between cessation of voicing at the time of onset and initiation of voicing at the release, and the voicing distinction of stops is substantiated by adduction vs. abduction of the vocal folds for the period of articulatory closure.

Although these studies have contributed to clarifying the properties of consonants, it can be said that there has been no systematic study of the acoustic characteristics of consonants in Japanese and the data on acoustic properties are rather scarce. Some of the current problems which these studies address are how these acoustic properties vary in different phonetic environments, how much native speakers vary in their production of the same voicing type, and how these properties can be characterized by quantitative measures. The present chapter presents the results of a detailed acoustic analysis and examines the acoustic properties distinguishing voiced-voiceless stops in Japanese.

2.2. Experimental Procedure

2.2.1 Subjects

The subjects of the present study consist of six native speakers of Japanese (three female and three male speakers). All of them are postgraduate students of the University of Edinburgh, and their ages range from 26 to 35 years. All the subjects are considered to have a good command of English. Two of the subjects show some features of a regional accent in Japanese, but they are negligible and all of the subjects are considered to be speakers of standard Japanese.

2.2.2. Linguistic Material

The recording materials consist of 22 monosyllabic words embedded in the carrier sentence of "Kore wa ____ desu." (This is ____) and 16 onomatopoeic words which consist of four CV syllables. The materials can be shown as follows:[2]

---

[2] The phonetic symbol [r] here may not be appropriate to represent the Japanese flap, which is usually defined as a voiced alveolar tap.

Bilabial	bi	'beauty'	pi	---
	ba	'place'	pa	'apiece'
	bo	'mother'	po	---
Alveolar	da	'Particle'	ta	'rice field'
	de	'coming out'	te	'hand'
	do	'degree'	to	'door'
Velar	ga	'moth'	ka	'mosquito'
	gi	'justice'	ki	'tree'
	ge	'low grade'	ke	'hair'
	go	'five'	ko	'child'
	gu	'tool'	ku	'poem phrase'

Onomatopeic words

piripiri - biribiri
pukapuka - bukabuka
kisikisi - gisigisi
kutakuta - gutaguta
taratara - daradara
torotoro - dorodoro
karakara - garagara
kurikuri - guriguri

Thus the materials comprise voiceless stops and voiced stops in initial position and intervocalic position. These materials except /pi, po/ are meaningful in Japanese, and presented to the subjects written in Japanese characters. The recording of these materials was made in the sound-proof recording studio of the Phonetics Laboratory, University of Edinburgh. After some brief practice reading, each subject recorded the materials three times at a natural speed, having been instructed to place a short pause before each test word. Three tokens from each subject were prepared for analysis for each item of the test materials.

2.2.3. Acoustic Analysis

The procedure of acoustic analysis is exactly the same as described under general experimental procedure in Chapter 1.

2.3. Results

2.3.1. Voice Onset Time (VOT)

The measurement of VOT was made for the interval from the consonantal release (R) to the onset of voicing (OV) for voiceless stops and for the interval from onset of voicing (voicing lead, VL) to consonantal release (R) for voiced stops, as shown in Figure 2.1. Table 2.1 shows the mean VOT values of stop consonants for three places of articulation and Table 2.2 indicates the mean VOT values for each subject.

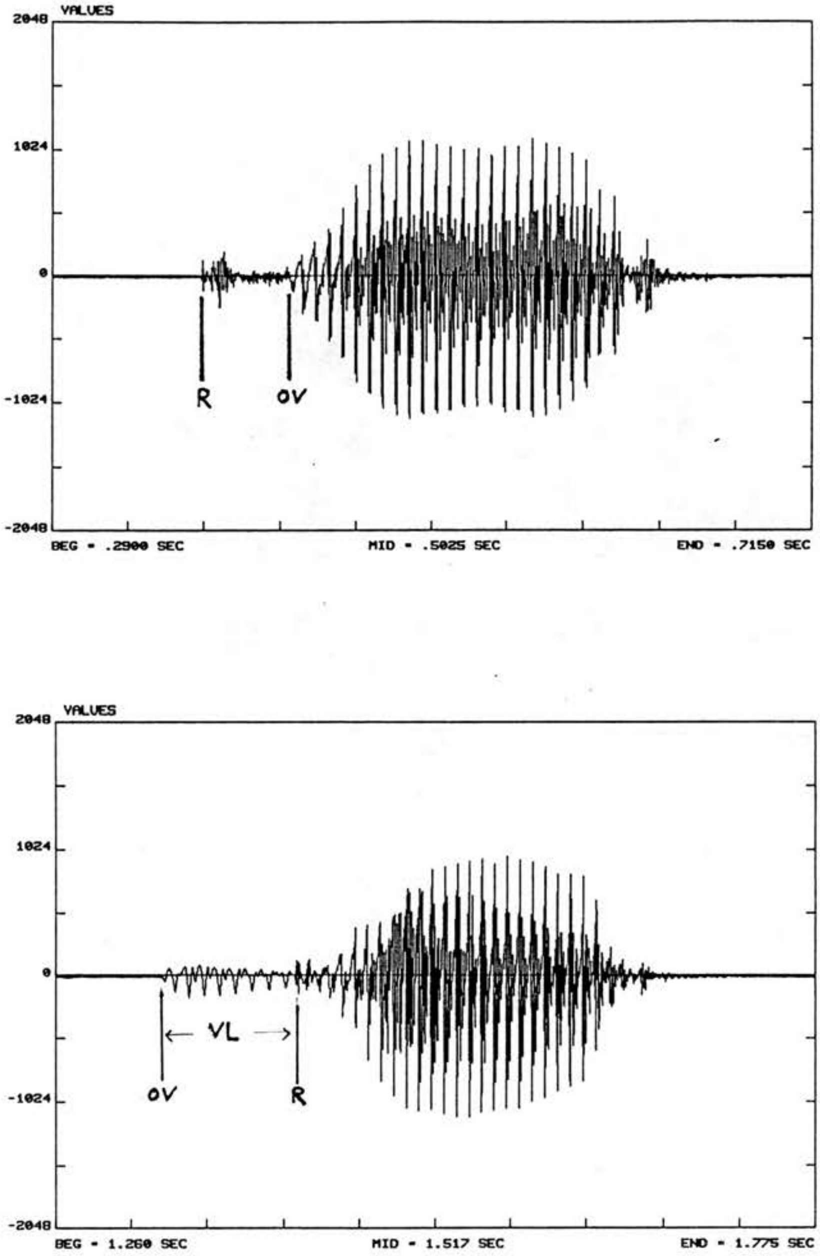


Figure 2.1 Waveform of [ka] (top) and [ga] (bottom)
R = Release, OV = onset of voicing,
VL = voicing lead

Table 2.1 Mean VOT values, s.d. and range of Japanese Stops for six subjects (ms)

	VOT	s.d.	Range		VOT	s.d.	Range
/b/	-89	28.5	-65 - -125	/p/	41	17.1	15 - 65
/d/	-75	32.7	-40 - -135	/t/	30	12.7	15 - 50
/g/	-75	27.0	-35 - -125	/k/	66	12.1	50 - 100

(N=72 for bilabial and alveolar, 120 for velar)

Table 2.2 Mean VOT values for each subject

		VOT (ms)		VOT (ms)
Subject A	/b/	-65	/p/	35
(F)	/d/	-45	/t/	23
	/g/	-101	/k/	71
Subject B	/b/	-88	/p/	48
(F)	/d/	-78	/t/	37
	/g/	-56	/k/	76
Subject C	/b/	-63	/p/	50
(F)	/d/	-50	/t/	22
	/g/	-36	/k/	58
Subject D	/b/	-70	/p/	35
(M)	/d/	-77	/t/	38
	/g/	-72	/k/	54
Subject E	/b/	-125	/p/	15
(M)	/d/	-140	/t/	13
	/g/	-104	/k/	56
Subject F	/b/	-123	/p/	65
(M)	/d/	- 68	/t/	48
	/g/	- 86	/k/	83

(N=6 for bilabial and alveolar, 10 for velar)

From the results, it can be seen that Japanese voiceless stops have a short to medium range of voicing lag, while voiced stops have a rather long voicing lead. These data show that there is a distinct difference in VOT between voiced and voiceless stops, and that there is no overlapping in VOT values between voiced and voiceless stops. The difference between the two categories is quite significant (two-tailed t-value=5.36, $p < .01$). The results on voiceless stops are in general agreement with a previous study (Itoh et al., 1979), but the ones on voiced stops differ in showing a considerable voicing lead before the release of oral closure. For the voiceless stops, it can be noted that velar stops show a relatively long voicing lag compared with bilabial and alveolar stops.

As to individual subjects in Table 2.2, there are some variations among them but the direction of the variations is generally consistent with the group data for voiceless stops. For subject A, the reverse to the group trend is shown in /b/ and /g/, and the voiced velar stop /g/ shows a long lead of VOT value. For subject B, the voiceless velar stop /k/ shows a long voicing lag in the data, and the values for voiced stops (absolute figures) become shorter, as the tongue moves from bilabial to velar. For subject C, the voiced stops show the shortest voicing lead among the subjects and the values (absolute figures) become shorter from bilabial to velar. For subject D, the voiced stops show almost the same voicing lead for each of the three types of voiced stops and the figures for voiced stops are more or less close to the means in Table 2.1. For subject E, the values of voiced stops show the longest VOT values among the subjects, while those of voiceless stops show very short voicing lag. For subject F, the voiceless stops show a rather long voicing lag and are fairly aspirated. In comparing the data of the male speakers with those of the female speakers, it can be observed that the male speakers tend to show a longer voicing lead than females.

As shown in these tables, there is a clear-cut distinction in VOT values for voiced vs. voiceless stops. It can be noted that voiceless bilabial and alveolar stops show a voicing lag and the voiceless velar stop shows a medium length of voicing lag. Since the delayed voicing value is often related to aspiration (Lisker and Abramson, 1964), it can be said that voiceless stops in Japanese are moderately aspirated.[3] In connection with this, it can be noted that the VOT value for the voiceless velar stop /k/ was greater than that of bilabial and alveolar stops. This is in accord with a previous study (Itoh et al., 1979). It is pointed out in Zue (1976) that as the point of articulation goes backward in the oral cavity, the VOT value

---

[3] VOT has been used to characterize the timing differences including what is commonly referred to as aspiration. However, it is important to remember that aspiration may also be characterized by turbulent airflow and aperiodic noise that accompanies it.

increases, but this was not observed between voiceless bilabial and alveolar stops.[4] The finding that /k/ has a longer value can be explained in terms of aerodynamic conditions in supra- and sub-glottal areas; as the tongue goes backward, the oral cavity behind it will be decreased, resulting in a higher pressure which would then cause the delay of voicing. For the voiced stops, there is considerable voicing lead before the release of consonant closure, and there doesn't seem to be any correlation between its value and the point of articulation.

In order to examine the effect of the following vowels, the data was analyzed as a function of the following vowel. Table 2.3 presents the results of VOT for each consonant followed by five vowels.

	/a/	/i/	/u/[5]	/e/	/o/
/p/	37	48	---	---	40
/t/	29	***	***	29	31
/k/	53	87	73	55	63
/b/	-93	-92	---	---	-83
/d/	-73	***	***	-95	-57
/g/	-63	-81	-70	-83	-80

Table 2.3 Mean VOT value as a function of the following vowel (N=12 for bilabial, alveolar and velar) (ms)
--- data was not obtained.
*** nonoccurring string

There seems to be some interaction between the vowel height and VOT values, though it is not so consistent. The table shows a tendency for VOT values of /ki/ and /ku/ to be greater than those of other syllables and it can be said that the VOT values tend to be longer when a high vowel follows, compared to non-high vowels. This may be attributed to a narrower oral

---

[4] According to Zue(1980), the increase of the VOT values after velar stops in initial position in English is explained as follows: "...the constriction for /g/ is formed by the tongue body, which is rather massive and cannot move away from the palate too rapidly following the release. It has also been observed that the motion of the tongue body after the release is such that a tapered, narrow opening is maintained for a longer period of time. Therefore, the constriction for /g/ opens slowly, allowing turbulence to be generated for a longer period of time."(p.54)

[5] /u/ is phonetically realized as unrounded [ɯ] in standard Japanese.

cavity created by raising the tongue or the effect of combining the articulatory configuration for the velar stop with high vowels.

2.3.2. Fundamental Frequency (Fo) and its Contour

As mentioned in Chapter 1, voicing contrasts are closely related to pitch perturbation. Voiceless stops tend to be associated with a high pitch range, and voiced stops with a low pitch range. Therefore, pitch range is considered to be one of the major properties for a voiced - voiceless distinction. Pitch perturbation is usually attributed to the degree of tenseness of the glottis and the rate of aerodynamic flow in the transglottal area. In the present analysis, measurement of Fo was made at the vowel onset following a stop release and between this and the steady state portion of the vowel 60 ms after the stop release. Since female and male subject groups show a different range of Fo perturbation, the data was shown in two separate groups, and the results are summarized in Tables 2.4 (a - b).

		Fo value		Fo value	Difference
Female group	/p/	253.9(8.5)	/b/	224.7(7.5)	29.2
	/t/	248.3(21.4)	/d/	212.7(5.5)	35.6
	/k/	245.5(9.7)	/g/	208.0(14.4)	37.5
	/p/	255.7(10.4)	/b/	231.3(12.5)	24.4
	/t/	252.6(3.1)	/d/	222.0(1.6)	30.6
	/k/	247.2(6.3)	/g/	219.7(11.8)	27.5

Table 2.4a Average Fo value at vowel onset (top) and steady-state (bottom) for female group (in Hz, N=18 for bilabial and alveolar, 30 for velar) (s.d. in parenthesis)

Male group	/p/	143.2(8.5)	/b/	130.9(17.6)	12.3
	/t/	140.2(13.4)	/d/	132.3(12.3)	7.9
	/k/	141.6(8.1)	/g/	123.0(17.2)	18.6
	/p/	139.2(10.0)	/b/	131.7(19.7)	7.6
	/t/	135.1(8.5)	/d/	130.3(19.1)	4.8
	/k/	136.6(9.2)	/g/	125.3(21.3)	11.3

Table 2.4b Average Fo value at vowel onset (top) and steady-state (bottom) for male group (in Hz, N=18 for bilabial and alveolar, 30 for velar) (s.d. in parenthesis)

Tables 2.4(a - b) show that there is a clear difference between voiced and voiceless stops in the Fo values at the onset of voicing, and there is also a varying degree of difference between the two groups of subjects. It is evident that the onset Fo values following voiceless stops are much higher than those of the voiced stops in both groups of subjects. In the case of female subjects, the mean difference between voiced and voiceless stops in onset Fo value is 34 Hz, and is statistically significant (two-tailed t-value=8.46, $p < .001$). In the case of male subjects, the same trend is observed, and there exists a difference between the two types of voicing at vowel onset as well, but the difference is not as great as that seen in the female group (two-tailed t-value=3.84, $p < .01$). It can also be seen that the difference between the two categories is still significant even at the steady-state portion (60 ms after vowel onset) of the vowel in both groups of subjects, though the magnitude of difference is less than that of the onset. These results are in general agreement with a previous study (Hombert, 1978). Furthermore, it can be observed that the values do not differ in any consistent way as the place of articulation of the stops changes, though there is a very weak trend that the voiced velar stop shows somewhat lower values than those of other places of articulation.

The effect of the vowels following the stop consonant on Fo can be presented as in Table 2.5. The figures indicate the onset Fo frequency values in Hz.

		/a/	/i/	/u/	/e/	/o/
Female group	/p/	248.0	261.0	---	---	252.0
	/t/	253.4	***	***	247.8	244.1
	/k/	243.3	246.4	248.4	246.0	244.3
	/b/	220.7	232.6	---	---	220.7
	/d/	211.0	***	***	217.3	209.6
	/g/	221.4	202.3	214.4	202.0	200.0
Male group	/p/	139.0	153.7	---	---	137.0
	/t/	138.3	***	***	139.0	143.0
	/k/	138.3	145.0	143.3	138.7	142.7
	/b/	132.0	138.3	---	---	122.3
	/d/	129.7	***	***	136.0	131.1
	/g/	125.0	122.2	122.0	127.3	117.6

Table 2.5 Fo values at vowel onset as a function of the following vowel for female (top) and male (bottom) speakers (in Hz, N=18 for bilabial and alveolar, 30 for velar)
--- data was not obtained.
*** nonoccurring string

From the above table, it can be observed that in voiceless consonants, there is a weak trend that onset Fo values followed by high vowels /i, u/ are higher than those followed by non-high vowels /a, e, o/ in both groups of subjects. In particular, the bilabial stop followed by /i/ shows the highest value in the data, but this trend does not hold in the voiced stops. Therefore, it is not apparent whether there is a systematic correlation between vowel height following stops and Fo perturbations.

Associated with the differences in the onset Fo values and steady-state Fo values, it is known that there is a difference in the Fo curves from the onset of vowel to the steady-state portion of vowels, as discussed by Hombert (1978). In the present study, the patterns of Fo curves were examined together with the measurements of Fo values at onset and steady-state portions. Figure 2.2 shows the normalized (averaged) frequency patterns from the onset of voicing to the steady-state portion (60 ms) of the vowel following the voiced and voiceless stops for the two groups of subjects. Upper lines represent the contours for female speakers and lower lines, those for male speakers, and the asterisk-marked line represents those for voiced stops, while the circle-marked line shows those for voiceless stops. The zero point indicates the onset of voicing and it should be noted that onset in the voiceless stops does not coincide with the onset in real time, i.e., as shown in VOT, the voicing starts rather later in real time in voiceless stops. Figure 2.2 clearly shows that there is a difference in average Fo onset value and the direction of change of Fo contours. Voiceless stops always start from a higher region than voiced ones, and the voiced stops show a slightly rising pattern in both groups of subjects. The voiceless stops, on the other hand, show a level (for female speaker) or a slight falling pattern (for male speakers).

Fo curves from the onset to the steady-state portion may be considered as another dimension for the voiced - voiceless distinction. However, Fo curves for the voiceless stops are not the ones predicted by previous studies (Hombert,1978; Laver, 1980) and show a rather different pattern from those of voiceless stops in English. According to Hombert (1978), it is reported that there is a difference in the direction of Fo curves between voiced - voiceless stops in the word-initial position; voiced stops show a rise from the onset of voicing, and voiceless stops shows a fall from the onset. As is well-known, there are several hypotheses about Fo perturbations;i.e., the aerodynamic hypothesis and the vocal fold tension hypothesis (Hombert, 1978). If the aerodynamic hypothesis is adopted here, the curves in Japanese voiceless stops may be attributed to the fact that there is not much difference in the transglottal air-flow rate at the glottis within the period from the onset of the vowel to the steady-state portion.

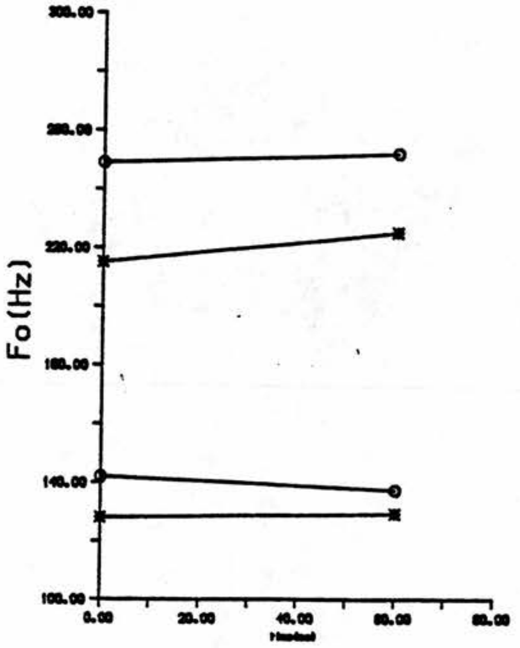
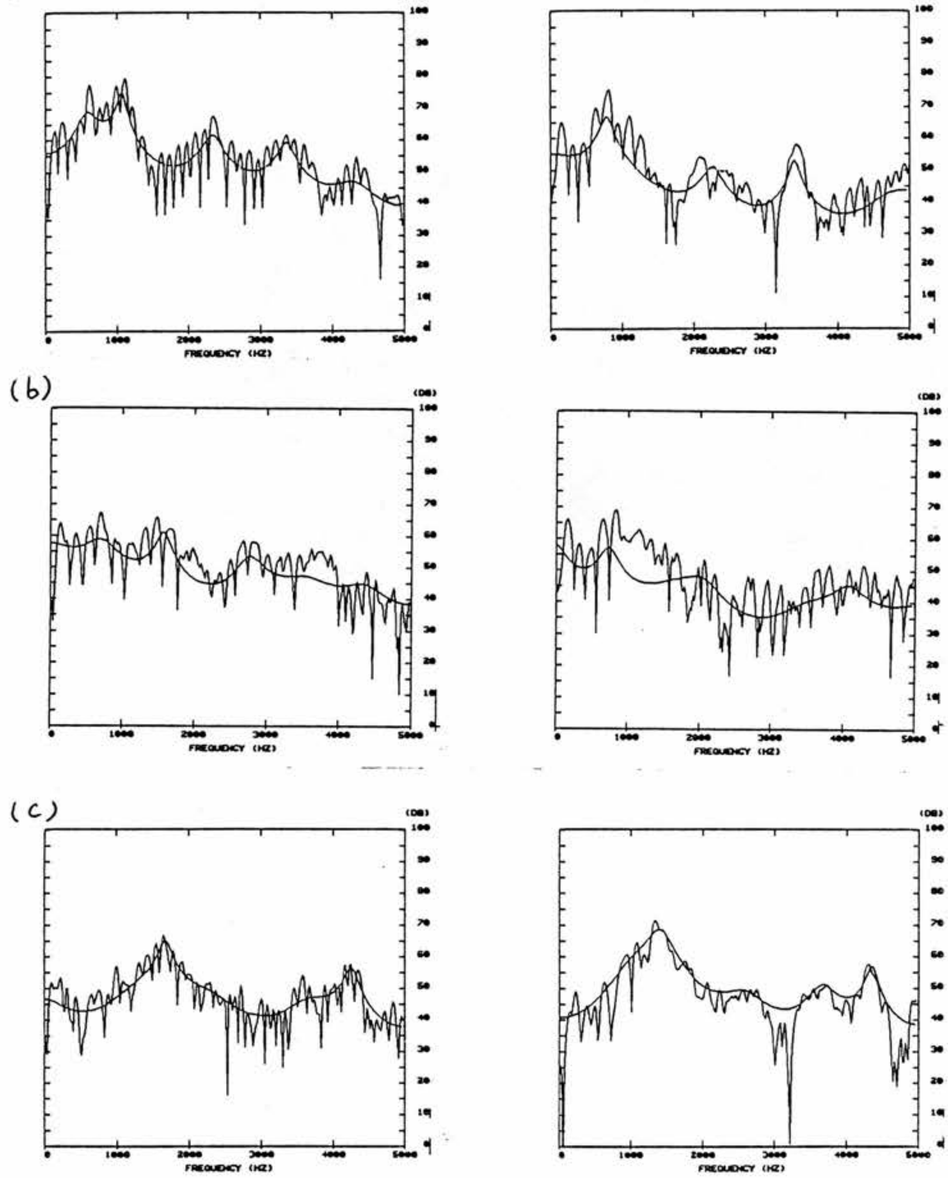


Figure 2.2 Normalized fundamental frequency patterns of vowels following voiced and voiceless stops in Japanese. Upper two lines for female speakers, and lower two lines for male speakers.
* voiced stops o voiceless stops

2.3.3. Spectral Analysis

It is generally considered that there is a difference in the air pressure in the oral cavity at the time of consonantal release between voiced and voiceless consonants, which are sometimes termed lenis and fortis, respectively. In the present experiment, a short-time spectral analysis was made in the initial portion beginning at the consonantal release. Figures 2.3 (a - c) show some of the spectra obtained by ILS linear prediction for the initial 25 msec time period. This period includes both the burst and some portion of the voicing onset. In the case of voiced stops, this period includes some portion of the vowel onset, whereas in the case of voiceless stops, this period includes mainly the burst and some aspiration and does not extend to vowel onset. This period is considered to contain some invariant cues for place of articulation (Blumstein and Stevens, 1980), but the present study mainly examined the spectral characteristics between voiced and voiceless stops. Figures 2.3(a - c) show power spectra for [ba - pa], [da - ta], and [ga - ka], and the solid line in each figure indicates the estimated filter function by LPC means - simply called smoothed spectra. It can be noted that there are many small peaks, more or less regularly distributed, in the voiced categories, compared to the voiceless ones. These peaks are considered to reflect the periodic nature of the rate of vibration of vocal folds. As to the overall intensity level shown in Figure 2.4 (a, b), it can be noticed that there is no consistent difference between the two voicing categories, though voiceless stops, as a weak trend, show a slightly higher intensity level than voiced ones. But in some cases, the intensity level of voiced stops is higher than that of voiceless ones. One thing to be noticed with these figures is that the intensity level of voiced stops in the lower frequency region (less than 700 Hz) is often greater than that of voiceless stops, in some cases, one small peak is observed in smoothed spectra. Although there are some differences in some other aspects such as the overall shape of the spectrum and slope of the smoothed spectra, these are considered to reflect both the glottal sources and the vocal tract filter function and are beyond the scope of the present chapter.



Figures 2.3 (a - c) Spectra of CV syllables in Japanese

(a) [ba - pa]
(b) [da - ta]
(c) [ga - ka]

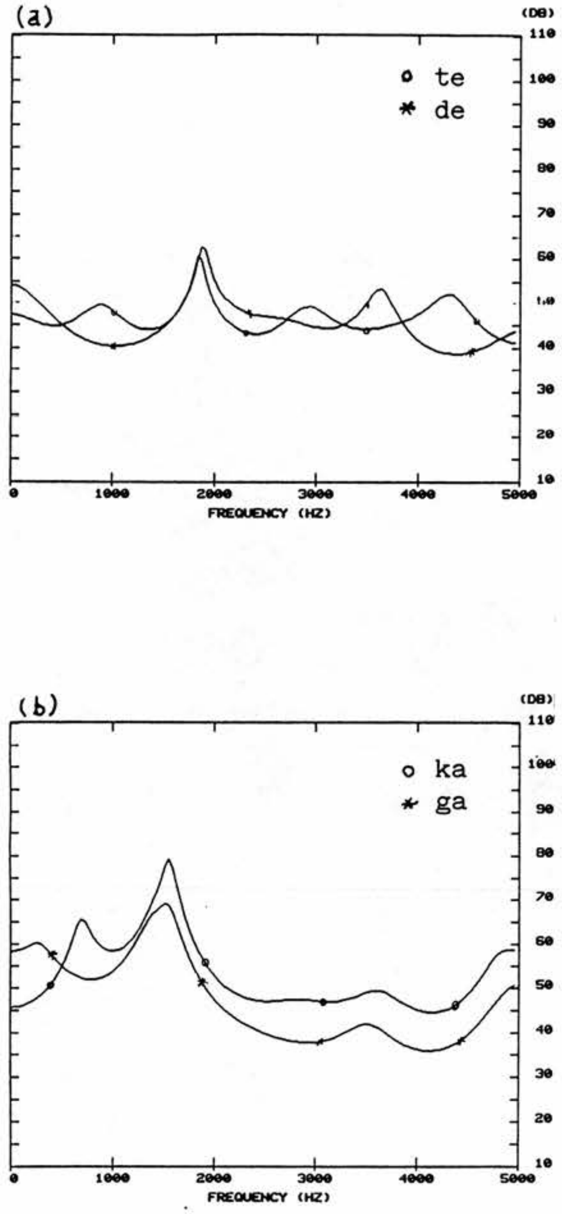


Figure 2.4 (a, b) Examples of smoothed power spectra in Japanese stops

2.3.4. The Onset Frequency of the First Formant

It is known in some studies (Lisker, 1980) that some characteristics of the first formant (F1) provide important cues for the distinction between voiced and voiceless stops. In particular, it is known that the onset frequency and the duration of the period from release to onset of the F1 cutback are important for the voicing distinction. In the production of voiced stops, it can be observed in the formant tracking that the burst portion is very brief, and F1 rises immediately after release, that is, as soon as the periodicity starts, the rising transition of F1 is observed. On the other hand, in the production of voiceless stops, the onset of periodicity does not appear for the initial brief period because of the release burst and the turbulent airflow, and occurs at some later time when the vocal folds are set in vibration for the following vowel, resulting in higher onset frequencies than those of vowels following voiced stops. The measurement of F1 onset frequency was made for each utterance in the Japanese data, and the mean values for each utterance are shown in Table 2.6. Figures 2.5 (a, b) show the formant patterns of Japanese stops [ka] and [ga].

Table 2.6 Mean F1 Onset Frequency of Japanese Stops (in Hz)

pa	883	ba	582
pi	230	bi	242
po	537	bo	469
X	550.0	X	430.9
ta	462	da	574
te	352	de	304
to	431	do	325
X	441.4	X	400.9
ka	744	ga	349
ki	232	gi	226
ku	245	gu	275
ke	350	ge	302
ko	544	go	327
X	391.2	X	295.7

(N=36 for bilabial and alveolar, 60 for velar)

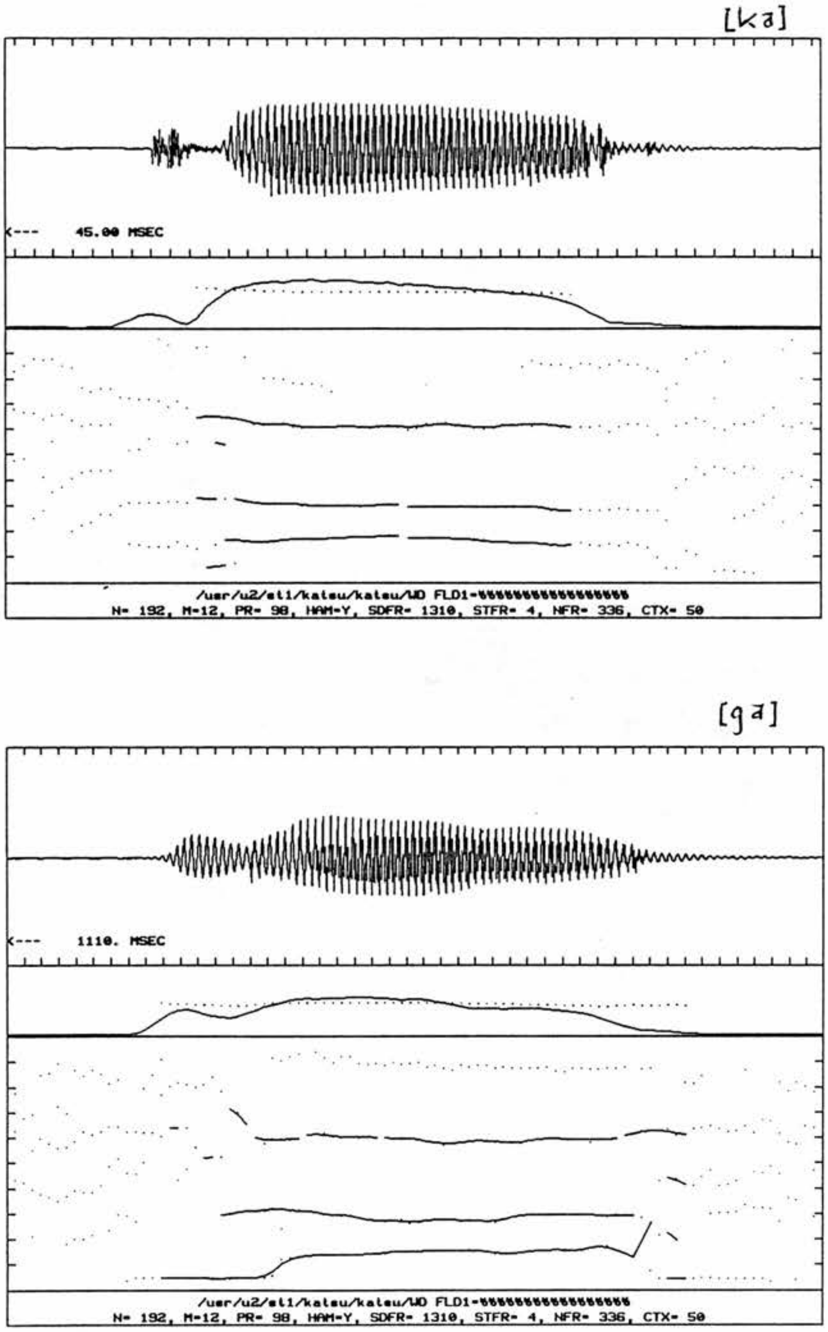


Figure 2.5a Formant patterns of Japanese stops [ka] and [ga]

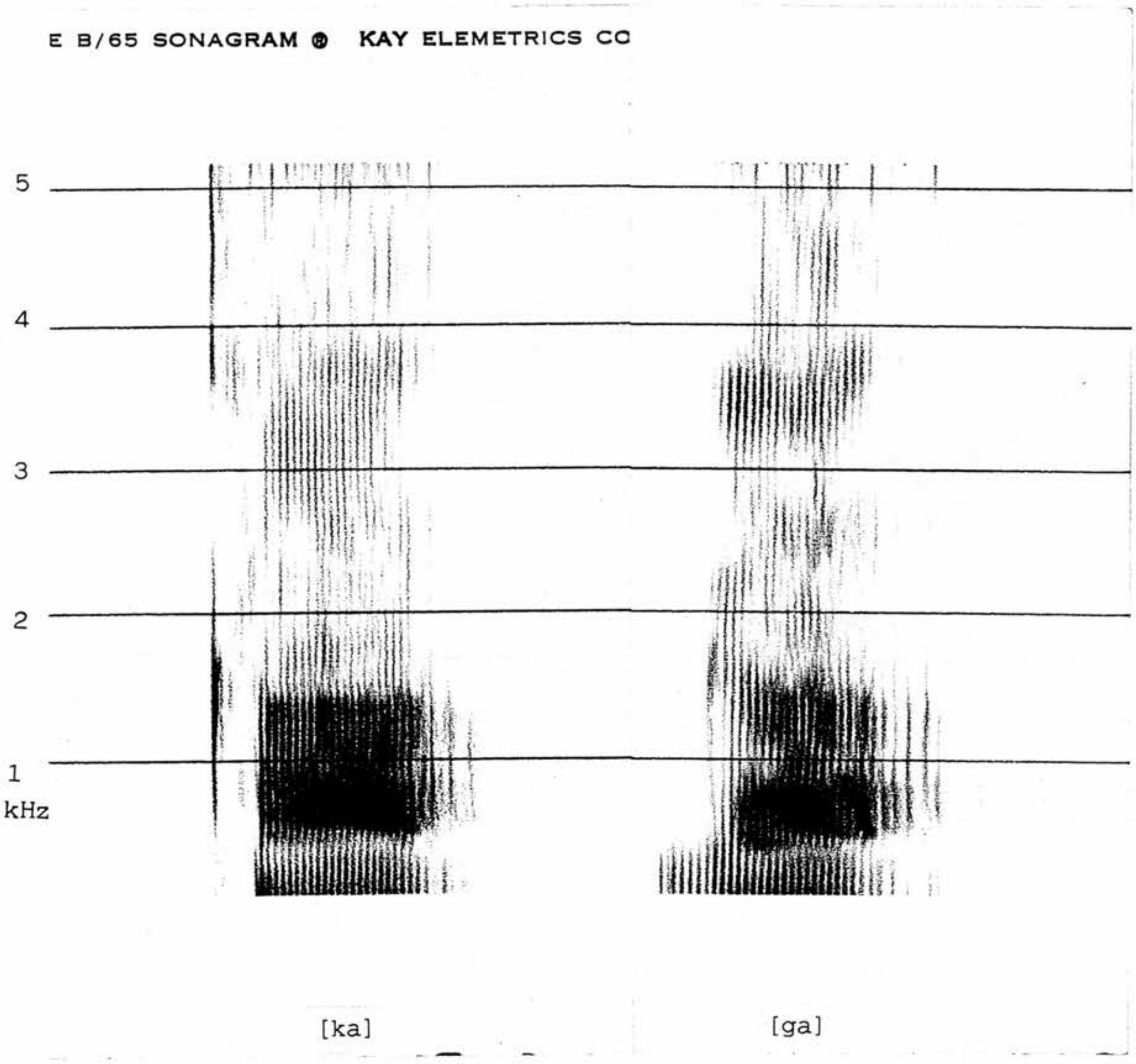


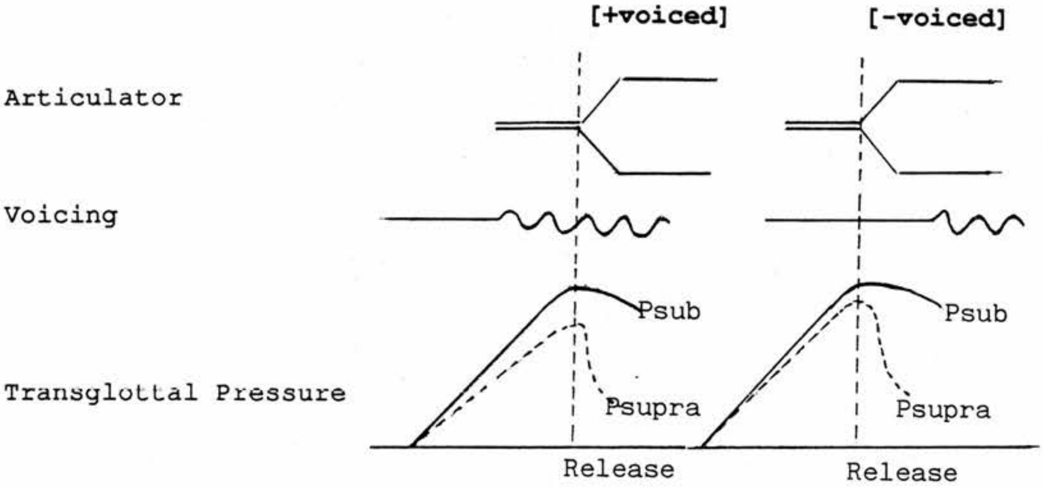
Figure 2.5b Examples of wide-band spectrograms of [ka] and [ga]

As seen in Table 2.6, the grouped mean for bilabial, alveolar and velar stops is higher in voiceless stops than in voiced stops, and it can be considered that the onset frequency of F1 cues for distinguishing voiced stops from voiceless ones. These trends agree with the results reported by Lisker (1975, 1980). A statistical test for the comparisons between voiced - voiceless stops, however, did not show that the difference was significant. Two-tailed t-values for the pairs of bilabial, alveolar and velar stops are 1.96, 0.75 and 1.74, respectively. Furthermore, it is observed that there is a variation in the effect of the following vowels; there is very little difference in F1 onset frequency between voiced and voiceless stops when a high vowel /i/ follows. This means that the F1 onset frequency difference is more apparent for utterances containing non-high vowels than for those containing high vowels.

2.4. Discussion on Acoustic Analysis

The present chapter shows that the voicing contrast of a word-initial stop in Japanese is signalled by the differences in several acoustic dimensions; VOT, Fo values and curves, spectral characteristics, and F1 onset frequency. Specifically, voiceless consonants show a certain range of voicing lag (30 - 70 ms), a higher Fo range, a level or slightly falling Fo curve and higher F1 onset frequency, while voiced consonants show 75 - 90 ms of voicing lead, a lower Fo range, and a rising Fo curve, higher spectral energy in the lower frequency region, and lower F1 onset frequency. The overall spectral energy, however, does not appear to be a consistent acoustic dimension for the distinction.

Based on the present data, the following schematized articulatory conditions can be considered:



As regards articulatory conditions, glottal postures for stops and air pressures above and below glottis are expected to vary with time. For voiceless stops, the glottis is open, the extent of the opening varying with different types of phonemes and the phonetic environments, and the degree of opening is greater in word-initial position than in word-medial position, as pointed out by Sawashima (1980), and, as VOT shows a voicing lag of 30 - 70 ms after the closure is released, there will be almost no difference in transglottal pressure at release; that is, oral closure builds up the pressure in the oral cavity, resulting in almost no airflow at the glottis. Then after the closure is released, built-up pressure rapidly decreases and as the transglottal flow starts, voicing starts. The Fo curve shows a high value at the time of release because of the apparent higher air-flow rate, but as the air flow gradually decreases, it has a level or falling curve. For voiced stops, since VOT shows 75 - 90 ms of voicing lead, there should be a transglottal air-flow and a slight enlargement in the oral cavity during the closure to allow airflow and to maintain voicing. Although it is not clear whether this enlargement is due to larynx lowering or pharynx enlargement or both, voicing is maintained as sufficient air flows through the glottis. Then after the closure is released, air rapidly flows out to the atmosphere, causing Fo to rise in a period of 50 - 100 ms.

As to the VOT, voiceless stops in Japanese are moderately aspirated, as shown in Table 2.1, and voiced stops show 75 - 90 ms of voicing lead. Comparing this to other previous studies, Itoh et al. (1979) reported the following VOT values.

/te/	20.54	/de/	-63.46
/ke/	43.74	/ge/	-40.70

(Itoh et al. 1979:127)

Compared to this data, the results of /te/ and /ke/ in the present study show a rather longer VOT for the above monosyllables, but the group trend is generally in accord with the above data. The finding that our results showed a rather long voicing lag may be partly due to the subjects' linguistic backgrounds. All the subjects are considered to have a relatively good command of English, and their phonetic realization of timing might be affected by acquiring English as a foreign language. [6]

As to the effect of the following vowel, it is reported in Klatt (1975) that there is an interaction between VOT and vowel height, and Weismer (1979) also reports that VOT is longer if the following vowel is tense, rather than lax. The present results show that VOT values were longer when the following vowel was a high vowel, compared to the other vowels /a, e, o/, but interaction with the vowel does not show itself in a systematic way. As is shown in Table 2.3, /ki/ and /ku/ showed a longer VOT, but this effect may be due to the combination with velar configurations. Among other syllables, /ko/ and /ke/ showed a rather long VOT, and /ba/, /bi/ and /de/ have a longer voicing lead (in terms of absolute

---

[6] It is known that there are some changes in the VOT values in one's speech in learning a foreign language, and one of the problems which remains to be explored is how learning a foreign language affects the VOT values in mother and target languages.

measurements). We do not have a satisfactory explanation for these syllables, except to say that onset of voicing is delayed because of the smaller size of the oral cavity when the voiceless velar configuration /k/ is formed. [7]

As is well attested in the literature, the voicing distinction exerts an effect on fundamental frequency and the curve within the short period following the consonantal release. The present results show that the onset Fo frequency is around 30 Hz (for female speakers) and 10 Hz (for male speakers) higher in the voiceless stops than in the voiced stops, and the magnitude of difference for female speakers is somewhat greater than that reported in Mohr (1971). The direction of the Fo curve from the onset to the steady-state is also associated with the distinction, and the present study showed that the curve for voiceless stops was a level or slightly falling pattern, while the one for voiced stops was a rising pattern. Although the pattern for voiced stops is in accord with the previous study (Hombert, 1978), that for voiceless stops is not consistent with the expected pattern.

As mentioned before, there are several hypotheses as to the cause of the different effects of Fo perturbation by voiceless and voiced stops: hypotheses based on aerodynamic conditions, vocal-fold tension, and larynx height. As reflected in these hypotheses, there are certainly several conditions which affect the Fo difference between voiced and voiceless stops. If we consider the aerodynamic conditions, the Fo patterns of voiceless stops can be explained as follows: at the time of release the air-flow rate is considered to be high, resulting in a high Fo onset, and the rate is maintained high within a short period, as the level pattern is observed from the release in Fo perturbations. However, this is a less convincing explanation of the present data, when we examine other data on Fo perturbations. It was observed in our present data that the onset Fo values do not differ as a function of a place of articulation, despite the fact that airflow rate and pressure are quite sensitive to the different articulatory configurations for /p, t, k/. If the aerodynamic conditions

---

[7] See Zue (1980:54).

are mainly responsible for Fo perturbations, then the different articulatory configurations would lead to different air-flow rates, and would result in different Fo onsets. The fact that stops /p, t, k/ show almost the same Fo values implies that the initial state of the vocal-folds is the same, and it is improbable that the aerodynamic conditions are the same for the different articulatory configurations. We may be better advised to look for explanations in the state of laryngeal tension and larynx height.

As to the examination of spectral characteristics, the present results show that there are some slight differences in the intensity level of the low frequency region between voiced and voiceless stops. Voiced stops show rather higher intensity in the lower frequency region, compared to voiceless stops. The difference appears to reflect that at the time of stop release, and more small peaks, somewhat regularly distributed, reflect the quasi-pulsing of the vocal folds in the initial state. However, there is no significant difference between the two voicing categories in the overall intensity level and the effect was not consistent.

Lastly, the measurement of F1 onset frequency shows that there is a difference between voiced and voiceless stops, and it was observed that the mean onset frequency of voiceless stops is higher than that of voiced stops. The result suggests that initial-stop voicing has an influence on the first formant transition characteristics of following vowels. This is mainly attributable to the difference in the onset of periodicity in relation to release in the two types of stops.

2.5. Oral Airflow

It is known that some articulatory differences between voiced and voiceless stops are reflected in their airflow characteristics. Airflow and air pressure are important aerodynamic parameters which are influenced by the state of glottis and the changes in the configuration of subglottal and supraglottal cavities. Thus it is considered that airflow characteristics are closely tied up with the voicing distinction of stops. In the present section, an attempt was made to examine the airflow characteristics for voicing contrasts of stops.

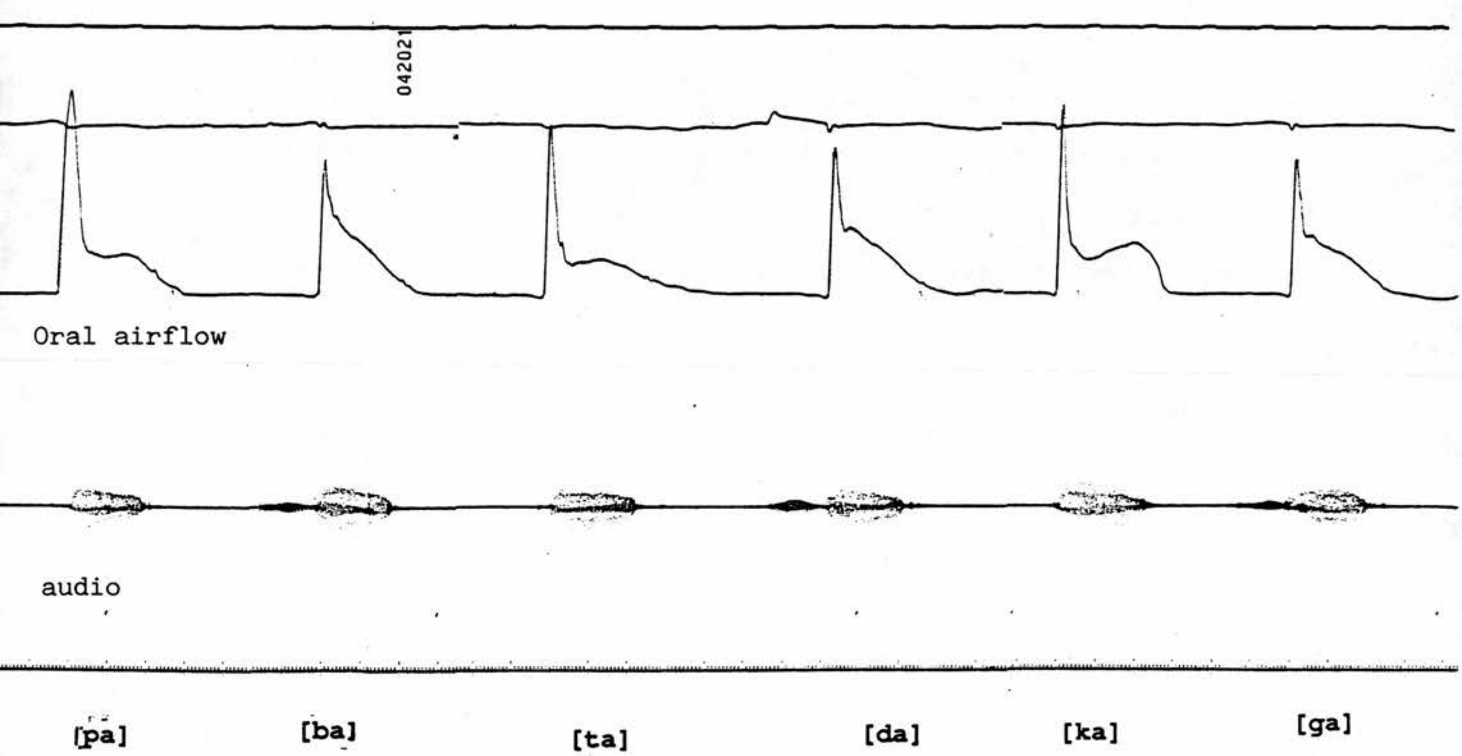


Figure 2.6 Recordings of airflow in Japanese stops

For collecting airflow data, the device prepared at the Phonetics Laboratory, Department of Linguistics, was used. Two pneumotach heads were fitted into an oxygen mask partitioned into two parts for oral and nasal airflow. The airflow passed through the pneumotach heads, and the result of the pressure difference was measured at the Gaeltec transducer, the sensitivity of which is ± 2 cm H_2O. The transducer was connected to the Gaeltec transducer amplifier, which was further connected to a mingograph which displayed the airflow visually on inkwriter paper. A time pulse generator is attached to the mingograph to give a timing pulse (10 ms and 100 ms).

The data is solely from the present author's own speech. The author is a native speaker of Japanese (male, 47 years old). Precautions were taken to ensure that the mask fitted tightly to the subject's mouth and nose. The author recorded oral (and nasal) airflow of the materials in section 2.2.2. in isolation and each word was uttered three times. Additionally, a larynx microphone was attached to a strap around the subject's neck to examine the audio-signal which reflects the activities of the vocal folds.

Some of the results are shown in Figure 2.6. As is seen in the figure, there is a clear difference in airflow between the two types of stops in Japanese. The peak flow level is consistently higher for the voiceless stops than for the voiced stops at the release, but the flow level after the release drops in the same way for both types of stops. Further, the duration of the peak flow is longer for voiceless stops than for voiced stops. It is not apparent whether there is a difference in airflow as the point of articulation moves from bilabial to velar.

2.6. Electropalatographic Analysis of Japanese stops

As mentioned in Chapter 1, the main difference between voiced and voiceless stops lies in vocal folds adjustments, their coordination with supralaryngeal activities, and aerodynamic conditions. There are other articulatory differences. For instance, there is a difference in tongue contact area between [t] and [d]; [t] extends farther back than [d], and further the duration of the tongue contact with the upper articulator is slightly longer for voiceless stops than for voiced stops (Pickett, 1980:137).

The present section reports the results of electropalatographic analysis on the tongue contact areas for voiced and voiceless stops in Japanese. There have been several studies on tongue contact areas of some specific sounds in Japanese by electropalatography (EPG), but the literature is quite limited of numbers in investigators and scope. Sudo et al. (1982) attempted to examine Japanese /r/ followed by five vowels. Ohnishi (1987) examined Japanese consonants in a rather limited way and reported that there is subject variation in the production of /r/. The present EPG experiment was carried out to examine whether there is any difference between voiced and voiceless stops in Japanese as a part of the investigation of voicing contrasts of stops.

The linguistic materials consist of the following syllables which were embedded in a carrier sentence "Kore wa ___ desu" (This is ___.). Three tokens were prepared for each syllable.

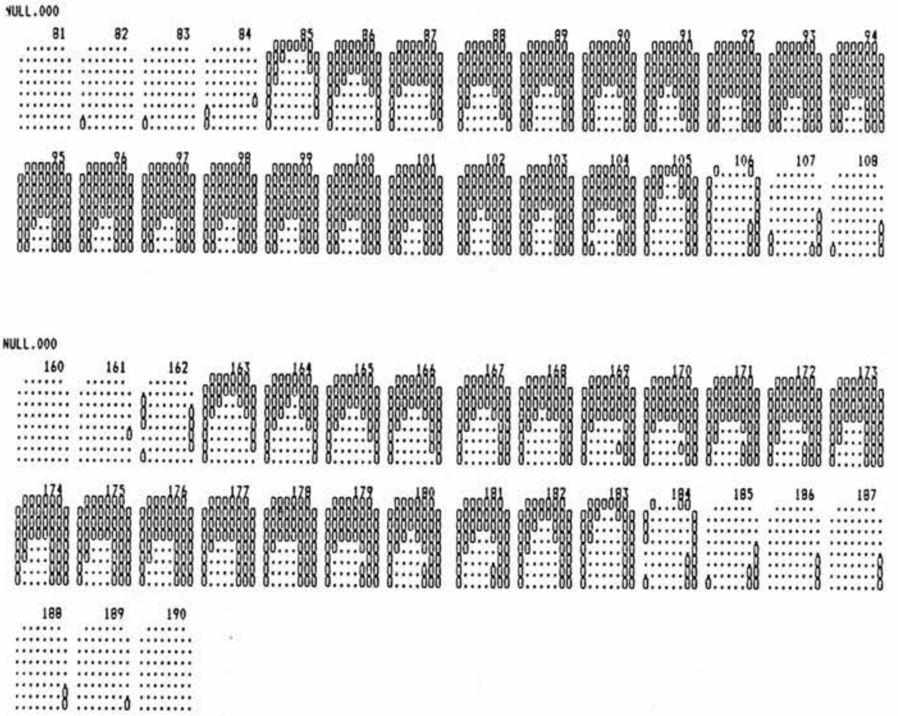
Word List for EPG Experiment

[pi] - [bi]
[ta] - [da]
[te] - [de]
[to] - [do]
[ki] - [gi]

The subject is a 47 year-old male speaker (the present author) and a speaker of standard Japanese. The system is the EPG2 manufactured in the Speech Research Laboratory, University of Reading. The palate is made at Broughton and Tyrrell, Newbury, and has sixty-two electrodes. Palato-lingual contact patterns were displayed on a computer monitor and could be visually examined. The contact patterns and analysis of data were printed out.

Figures 2.7 (a - c) show EPG charts of [t - d] followed by three vowels /a, e, o/, and, as is pointed out, /t, d/ in Japanese only contrast before these three vowels. Each diagram represents a contact area of the palate with the alveolar ridge at the top. The contact area, represented in circle, is read to be from left to right with 10 ms-interval between each diagram. In three utterances for each syllable, there were slight variations in tongue movement, contact area, and duration of contact, but it can be seen that the general trend is consistent.

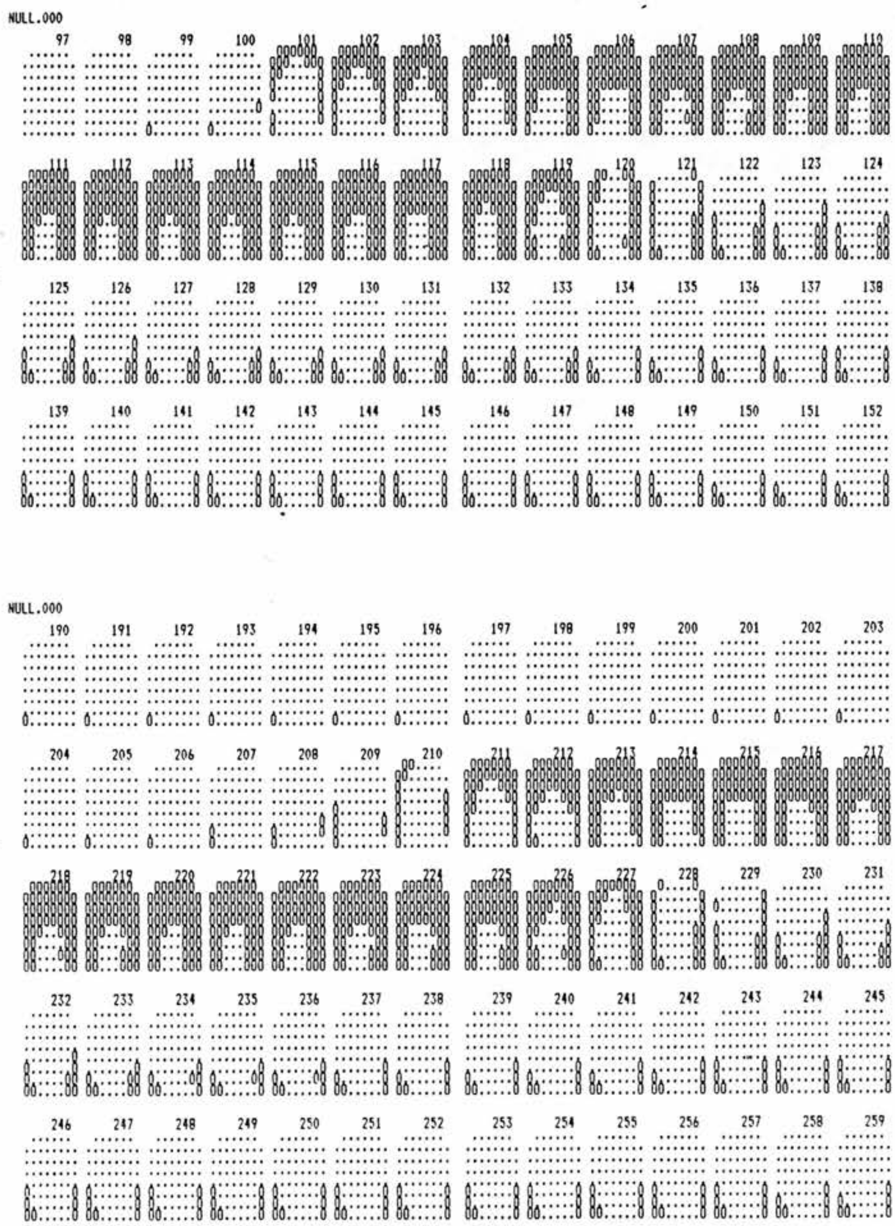
(a)


NULL.000
81 82 83 84 85 86 87 88 89 90 91 92 93 94
95 96 97 98 99 100 101 102 103 104 105 106 107 108
NULL.000
160 161 162 163 164 165 166 167 168 169 170 171 172 173
174 175 176 177 178 179 180 181 182 183 184 185 186 187
188 189 190

[ta]

[da]

(b)


NULL.000
97 98 99 100 101 102 103 104 105 106 107 108 109 110
111 112 113 114 115 116 117 118 119 120 121 122 123 124
125 126 127 128 129 130 131 132 133 134 135 136 137 138
139 140 141 142 143 144 145 146 147 148 149 150 151 152
NULL.000
190 191 192 193 194 195 196 197 198 199 200 201 202 203
204 205 206 207 208 209 210 211 212 213 214 215 216 217
218 219 220 221 222 223 224 225 226 227 228 229 230 231
232 233 234 235 236 237 238 239 240 241 242 243 244 245
246 247 248 249 250 251 252 253 254 255 256 257 258 259

[te]

[de]

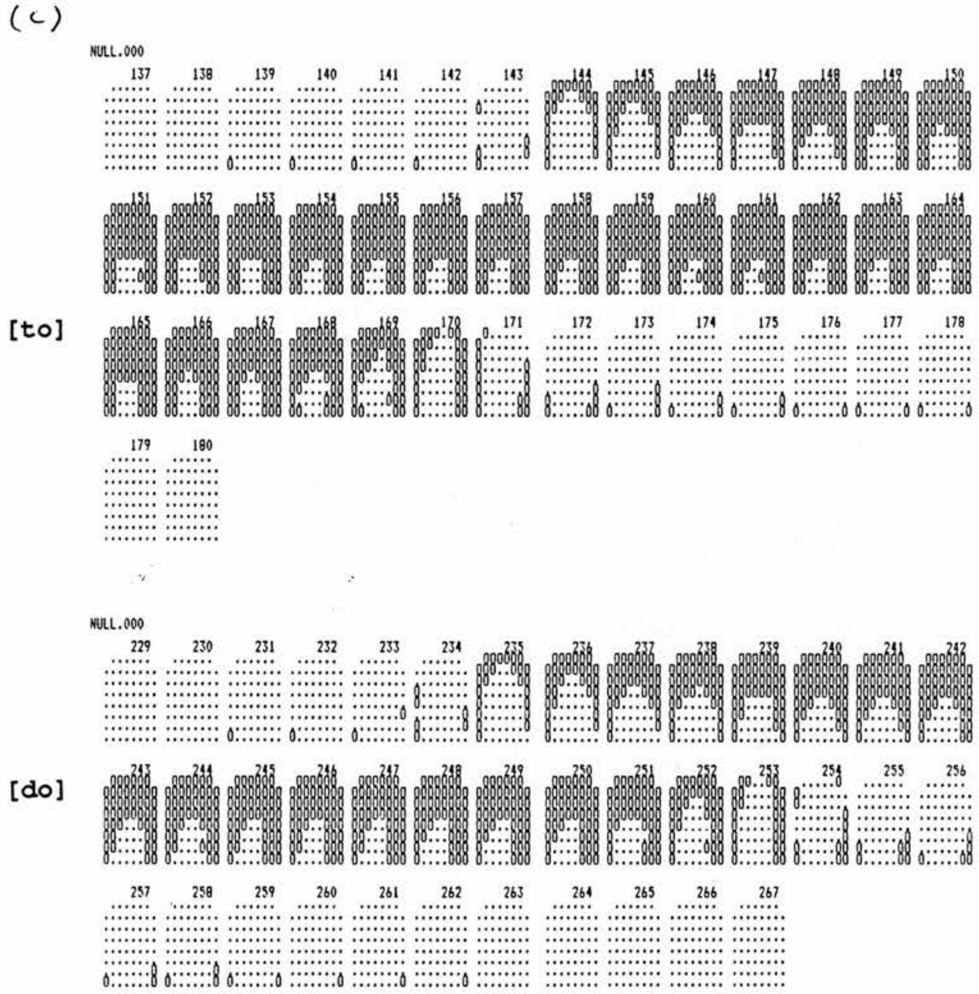


Figure 2.7 (a-c) EPG charts showing tongue-palate contacts in [ta - da], [te - de], and [to - do]

In comparing [t] and [d] followed by [a], the maximum contact area for [t] is seen from frame numbers 93 - 101, while the one for [d] is seem from frame 173 - 178. In both tokens, the contact proceeds from alveolar ridge to post palato-alveolar region. In examining the difference between [ta] and [da], [ta] shows a greater contact area than [da], and the duration of maximum contact is longer in [ta] than in [da], but it doesn't seem that there is a difference in the order of contact between them. Next, in examining the difference between [te] and [de], the maximum contact starts from frame numbers 112 - 115, whereas the one for [de] starts from 220 - 222. Although the difference is not great, there is a weak trend that [te] shows a rather greater contact area than [de]. Furthermore, in comparing the charts of [to] with [do], the difference is seen in the contact area; [to] shows a greater contact area than [do].

Examining these EPG charts for one speaker, we notice that there is a tendency for voiceless stops [ta, te, to] to show greater tongue contact areas than their voiced counterparts [da, de, do]. This observation is not in agreement with the previous study (Keller, 1971:34) in which she concludes that there is no consistent difference in the contact area of voiced and voiceless sounds in American English. Since the present EPG experiment is completely different from Keller's in its method and in the languages investigated, it may not be appropriate to compare these data. The difference between voiced and voiceless stops in the present study indicates that voiceless stops in Japanese are articulated with a greater muscular tension, and may be termed as a "strong" consonant compared to a "weak" voiced one.

2.7. Summary

The present chapter examined several acoustic dimensions which are relevant for the voicing contrast in Japanese stops, together with oral airflow and EPG data. The results of the experiment can be summarized as follows:

(1) The VOT values for voiceless stops showed a range (30 - 70 ms) of voicing lag, while those for voiced stops showed a voicing lead of 75 - 90 ms. The values are distinctively distributed in the two voiceless - voiced categories. Among the three places of articulation, the voiceless velar stop /k/ showed the longest voicing lag, and this can be attributed to the turbulent air-flow in the narrow opening of the velar configuration. Although there seems to be some correlation between VOT and the following vowel, the effect of the vowel on VOT is not apparent, though consonants followed by high vowels seem to have a longer VOT than those followed by other vowels.

(2) The fundamental frequency at release and the curves from the onset to the steady-state are systematically related to the voiced - voiceless distinction. The Fo values of vowel onset after voiceless stops are about 30 Hz (for female speakers) and 10 Hz (for male speakers) higher than after voiced stops. The voiced stops show a rising curve in the initial period of 60 ms after the release, while the voiceless ones show a level or slightly falling pattern.

(3) The present results do not show in a systematic way that there is a difference between the two voicing categories in the overall spectral energy, but the intensity level in the lower frequency region is higher for voiced stops than it is for voiceless ones.

(4) The onset F1 frequency is higher in voiceless stops than in voiced stops, and is a sufficient cue for distinction between voiced and voiceless stops in Japanese.

(5) The oral airflow data show that there is a consistent difference between voiced and voiceless stops; voiceless stops show a higher oral airflow than voiced ones.

(6) The EPG experiment shows that voiceless stops [ta, te, to] have a greater contact area and a longer duration of contact than their voiced counterparts [da, de, do].

Chapter 3

Stops in Mandarin Chinese

3.1. Introduction

This chapter deals with stop consonants in Mandarin Chinese. Mandarin Chinese has two types of stops: voiceless unaspirated /p, t, k/ and voiceless aspirated /p^h, t^h, k^h/, so the difference between them is the one of aspiration, not voicing. It is known that stops /p, t, k/ which occur in syllable initial position are not entirely unaspirated but are slightly aspirated in most cases. In the notational system of Mandarin Chinese, there are several ways to represent the stop consonants, and in some notations unaspirated stops are symbolized as voiced plosives, and aspirated stops as voiceless plosives (Huang, 1969). Whatever the differences in the notational systems are, the obstruents in Mandarin Chinese are voiceless and the degree of aspiration is relevant for the two types of stops. It is also pointed out that the two types of stops have a fortis - lenis distinction; aspirated stops are referred to as fortis, and unaspirated stops as lenis (Dow, 1972:24).

As is well known, Mandarin Chinese is a monosyllabic language, and has four contrastive tones: High-level(55), High-rising(35), Low-dipping(213), and High-falling(51). The syllable in Mandarin Chinese represents the morpheme, or two morphemes when /r/ is suffixed to another morpheme. Phonetic and phonological studies of Chinese are always associated with the syllable structure. According to the generally accepted approach in Chinese phonology, the syllable structure is divided into two portions: the initials and the finals, and is represented as (C1)V or (C1)(V1)V(C2) where C1 is any consonant except /ŋ/, V1 is either /i/,

/u/ or /y/ and C2 is either /n/ or /ŋ/.[1] Stops occur only in word-initial position. Further, in what is known as contextual variation, /p, t, k/ become voiced in intervocalic position, and /p^{h}, t^{h}, k^{h}/ are less aspirated in the same environment.

As to the literature review, Mandarin Chinese is well-documented, and there have been a number of studies on the phonological system of the language, especially syllable structure and phonemic analysis. But experimental studies on Mandarin phonetics are rather limited in scope, and most of them have been concerned with tone and the variations in relation to intonation. Although there have been several attempts to clarify the phonetic properties of consonants, there has been no systematic study of them. Among some of the experimental studies, Howie (1974) examined the domain of tone in Mandarin Chinese and found that the basic contours of tones are not the entire voiced part of the syllable but are somewhat confined to the syllabic vowel and any following voiced segment. Howie (1976) further examined the acoustic properties and perception of vowels and the four contrastive tones, and discussed the formant frequencies, their contextual variations and perception of the four distinctive tones. Ho (1976) investigated the variations of tones and found that the influence of sentence environments is greater than that of the syllabic vowels. Moreover, Connell et al. (1983) investigated how far the shape of a tone can be changed before it is identified as a different type and found that lexical tones can withstand a large degree of perturbation. Meanwhile, Ren (1986) investigated the acoustic structures of diphthongs and discussed the interface between linguistic transcriptions and the physical realization of diphthongs. As to the physiological aspects of Mandarin Chinese, Iwata and Hirose (1976) measured the glottal width of stops and affricates by using a fiberscope, and pointed out that the distinction between voiceless unaspirated and aspirated stops involves various manners of laryngeal activities. As seen in these studies, the experimental studies have been mainly done on tones, and those on consonants are rather scarce. This chapter aims to

---

[1] There are three types of syllables in Mandarin Chinese: (1) (C1)V, (2) (C1)(V1)V(C2) and (3) /r/ alone, or suffixed to syllables of Types 1 and 2, forming a morphemically complex monosyllable.

examine acoustic properties of stops and to examine the characteristics and the variations of two types of stop consonants.

3.2. Experimental Procedure

3.2.1. Subjects

The subjects of the present study are three native speakers (male) of Mandarin Chinese. All of them are from the People's Republic of China and are postgraduate students of the University of Edinburgh. Their ages range from 30 to 35. Two of them are from the Beijing area and the other speaker shows some regional characteristics in his speech, but they are considered to be negligible. Two of them are majoring in Applied Linguistics, and have relatively a good command of English. The other is majoring in Social Science.

3.2.2. Linguistic Materials

As mentioned, Mandarin Chinese is a tonal contrast language. The recording materials consist of minimal or near minimal pairs with the same tone, listed in Table 3.1. Each word in the list was read five times; three times in isolation and twice in the carrier sentence *zhè gè zì nìan* _______ "This word is pronounced as ______". The recording was made in the sound-proof recording studio in the Phonetics Laboratory, University of Edinburgh. Three tokens (the second and the third items in isolation and one token in the carrier sentence) were recorded. Each subject was instructed to place a short pause before each test item in the carrier sentence. It does not appear that there is any significant difference in results between items in isolation and those in the carrier sentence.

Table 3.1 List of words in Mandarin Chinese[2]

bao 55	'to wrap, a bag'	pao 55	'to toss'
dao 55	'knife'	tao 55	'wave'
gao 55	'high'	kao 55	'hip'
ba 51	'father'	pa 51	'to fear'
da 51	'big'	ta 51	'to step'
ga 51	'embarrassment'	ka 213	'card'
bi 55	'to compel'	pi 55	'to criticize'
di 55	'low'	ti 55	'ladder'
bu 55	---	pu 55	'to extinguish fire'
du 55	'to supervise'	tu 55	'bald'
guo 55	'pot'	kuo 51	'to expand'

3.2.3. Acoustic Analysis

The procedure of acoustic analysis is exactly the same mentioned in the general experimental procedure in Chapter 1.

3.3. Results

3.3.1. Voice Onset Time (VOT)

The measurement of VOT was made on the display screen by manually controlling two cursors, and Figure 3.1 shows waveforms of two types of stops and acoustic portions for VOT. The accuracy level in measurement was 5 ms. Table 3.2 presents the mean VOT value (ms) for each stop in the initial position.

---

[2] In the present chapter, symbols of Pinyin (the new Chinese phonetic alphabet) are used to represent words in Chinese. IPA phonetic symbols are used for transcription of the phonetic contrast of voicing, whenever necessary.

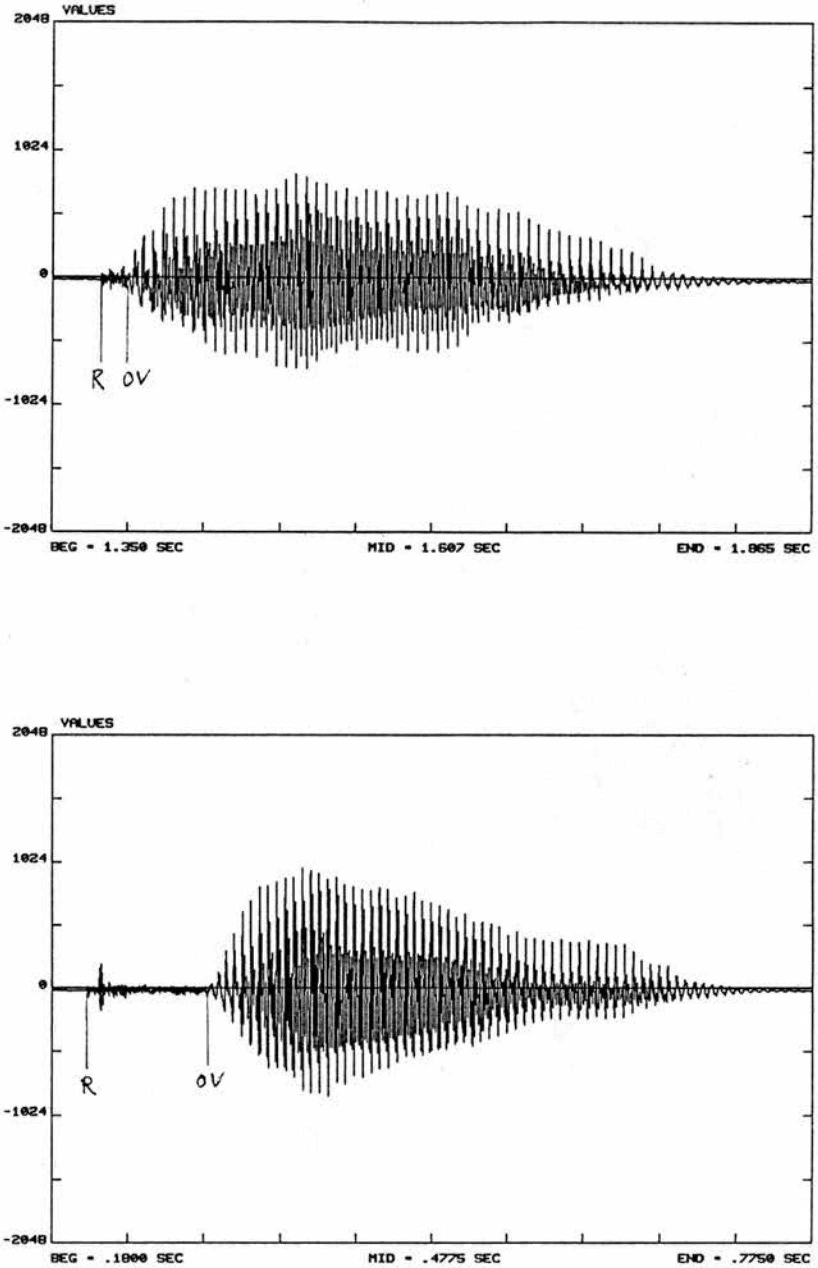


Figure 3.1 Example of waveforms of two types of stops in Mandarin Chinese. gao 55(top) and kao 55(bottom)

Table 3.2 Mean VOT value and the range in Mandarin Chinese (N= 18 for bilabial, alveolar and velar) (s.d. in parenthesis)

Stop	VOT	Range	Stop	VOT	Range
/p/	7 (2.3)	5 - 10	/pʰ/	96 (13.3)	80 - 115
/t/	12 (2.1)	10 - 15	/tʰ/	98 (16.1)	80 - 120
/k/	19 (3.8)	15 - 25	/kʰ/	112 (20.7)	90 - 130

Table 3.2 clearly shows the difference in VOT values for two types of stops at each place of articulation, and there is no overlapping in VOT between the two types of stops. Aspirated stops are heavily aspirated, and average about 6 to 13 times longer than unaspirated stops. The results distinctively show the general characteristics of voiceless unaspirated stops and voiceless aspirated stops.

It can also be seen in Table 3.2 that, as in other languages, the VOT values of velar stops tend to be longer than those of bilabial and alveolar stops (see p. 27).

In order to examine the effect of the following vowels, the data was analyzed as a function of the following vowel. Table 3.3 presents the mean VOT value followed by three vowels.

Table 3.3 Mean VOT value as a function of the following vowel(ms) (N=6 for bilabial, alveolar, and velar)

	/ɑ/	/i/	/u/
/p/	3	7	12
/t/	3	18	14
/k/	19	--	33
/pʰ/	105	81	101
/tʰ/	101	91	103
/kʰ/	112	--	115

Although it is reported (Klatt, 1975) that there is an interaction between vowel height and VOT in English, it does not appear that vowel difference affects the VOT in Chinese in a consistent way. As a weak trend, however, VOT is longer before /i/ and /u/ than before /ɑ/ in voiceless unaspirated stops.

Furthermore, an additional acoustic analysis was made to examine the tone effect on VOT. The words examined can be shown as follows:

bi 55	'to compel'	pi 55	'to criticize'
bi 35	'nose'	pi 35	'skin'
bi 215	'pen'	pi 213	'particle to count animals'
bi 51	'certainly'	pi 51	'lonely'

Table 3.4 presents the mean VOT values for these words for three subjects.

Table 3.4 Mean VOT value as a function of tones (ms)
(N=6 for each tone type)

bi 55	5 ms	pi 55	103 ms
bi 35	7	pi 35	93
bi 213	10	pi 213	123
bi 51	8	pi 51	97

From the above results, it does not appear that there is any correlation between the four tone types and VOT. As a very weak trend, however, VOT values in tone type 213 appear to be longer than other types of tones. As examined by Howie (1974), tone effect is usually considered to appear in the final portion of the syllable, and it seems to be reasonable that the difference of tones does not affect VOT, which is considered to be a timing event in the initial portion of syllable.

3.3.2. Fundamental Frequency (Fo) and its Contour

As we saw earlier, it is a commonly found phenomenon that voiceless and voiced obstruents have different effect on pitch perturbations of the following vowels. In the present study, the measurement was made at vowel onset for each word to examine whether there is any difference between the two types of voiceless stops. In order to minimize the tone effect, the measurement was restricted to the words with level tone 55. Table 3.5 presents the mean Fo at vowel onset for each stop.

Table 3.5 Mean Fo at vowel onset (in Hz)
(N=18 for bilabial, alveolar and velar)
(s.d. in parenthesis)

stop	Fo	stop	Fo	Dif.
/p/	153 (6.1)	/pʰ/	160 (8.8)	7
/t/	157 (6.8)	/tʰ/	160 (10.0)	3
/k/	141 (5.0)	/kʰ/	152 (5.7)	11

From Table 3.5, it can be seen that there is a difference between voiceless unaspirated stops and aspirated stops. The onset Fo is higher for aspirated stops than it is for unaspirated stops at each place of articulation. Although the difference is around 10 Hz, it is significant for velar stops (two-tailed t-value 4.22, $p < .01$) and for bilabial stops (two-tailed t-value 4.01, $p < .05$). The results agree with the generally assumed trend that voiceless aspirated stops show a higher Fo at the onset of the following vowel than voiceless unaspirated stops.

In order to examine the effect of the following vowel, the Fo results can be reanalyzed as a function of the vowel followed. Table 3.6 presents the averaged onset Fo in the vowel environment.

Table 3.6 Mean Fo as a function of following vowel (Hz)
(N=6 for bilabial, alveolar and velar)

	/ɑ/	/i/	/u/
/p/	142	155	162
/t/	143	161	168
/k/	141	---	154
/pʰ/	147	164	170
/tʰ/	144	165	173
/kʰ/	152	---	---

It can be seen in Table 3.6 that there is a difference in pitch perturbations between high vowels /i, u/ and non-high vowel /ɒ/. It is apparent that high vowels /i, u/ give a higher Fo than the non-high vowel. This is in accordance with the generally accepted view that the higher the vowel is, the higher the pitch is. There are currently two hypotheses to explain this effect; the coupling hypothesis and the "tongue pull" hypothesis. [3]

---

[3] See Ohala, J.J.(1978:29-30).

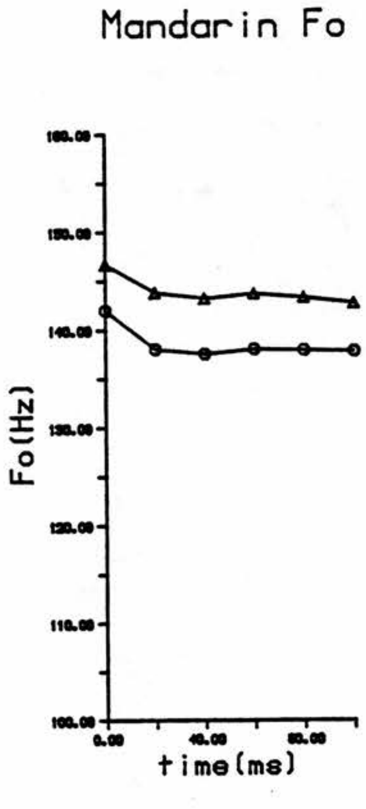


Figure 3.2 Fo curves of voiceless unaspirated and voiceless aspirated stops. The zero point shows the onset of voicing. ○ vl.unaspirated △ vl.aspirated

Furthermore, the Fo curves from the onset of vowel to the steady-state portion of vowels were examined to see if there is any difference between the two types of stops in the direction of the curves. In the present study, Fo curves were measured at onset and 20, 40, 60, 80 and 100 ms after onset for the words with level tone (55). Figure 3.2 shows the normalized (averaged) frequency pattern from the onset of voicing to the steady-state patterns. The upper line represents the pattern for voiceless aspirated stops and lower line that for voiceless unaspirated stops. It can be seen that the difference in Fo between the two types of stops in the onset portion is still found even 80 ms after the onset with almost the same magnitude. The two types of stops only differ in the relative value of Fo and aspirated stops give a higher Fo to the following vowel. The two types of stops show the same pattern of Fo curves and it can be seen that they show the pattern which is characteristic of voiceless stops; a slightly falling pattern immediately after the release. It can be presumed that there is a difference in the rate of airflow between the two types of stops in Mandarin Chinese. Glottal airflow might be higher at vowel onset after voiceless aspirated stops than after voiceless unaspirated stops. As the rate of airflow decreases after the release, the Fo will decrease too. Thus, after two types of voiceless stops, the Fo curves are considered to be initially high and then falling. So the two curves in Mandarin Chinese do not differ from each other in the direction of Fo change, but differ in the relative value.

3.3.3. Spectral Analysis

In order to examine spectral characteristics of the two types of stops, power spectra were sampled in the short period immediately following consonantal release for bao 55, pao 55, dao 55, tao 55, gao 55, and kao 55. A total of 36 spectra were sampled for three subjects. The 25 ms time window period includes both the burst and some portion of the voicing onset. In the case of voiceless unaspirated stops, on the other hand, the period includes both burst and some portion of the vowel onset. In the case of voiceless aspirated stops, it includes the burst and some portion of aspiration and does not extend to the vowel onset. Figure 3.3 (a - b) shows some examples of the spectra and the solid line represents the filter function in the vocal tract. The examination of these power spectra indicates that voiceless unaspirated stops show regularly distributed small peaks, which reflect the periodic nature of the rate of vibration of vocal folds, but voiceless aspirated stops show less regular distribution. Further, bandwidths are narrower for unaspirated stops than aspirated ones - as far as bandwidth can be assessed in these spectra.

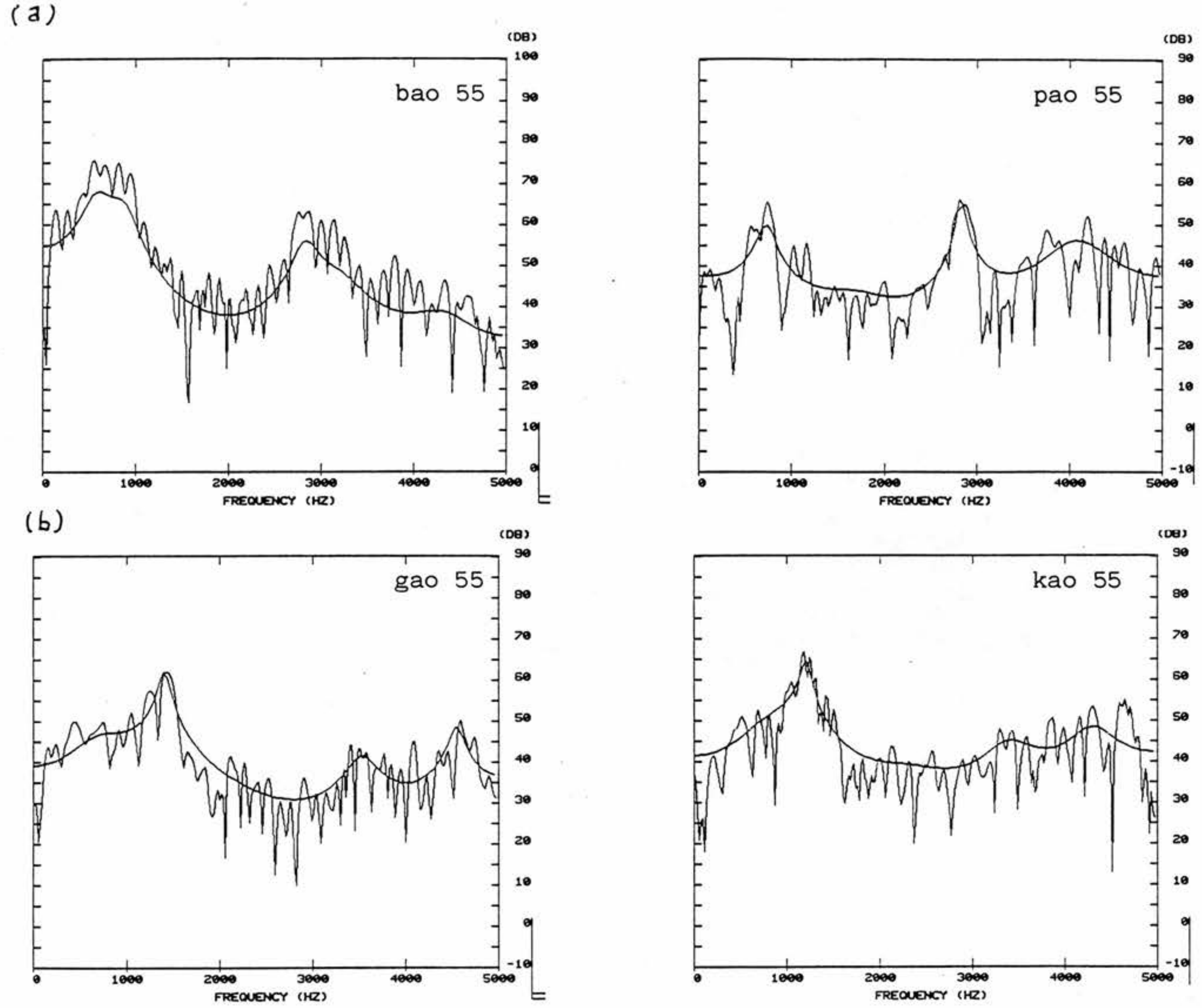
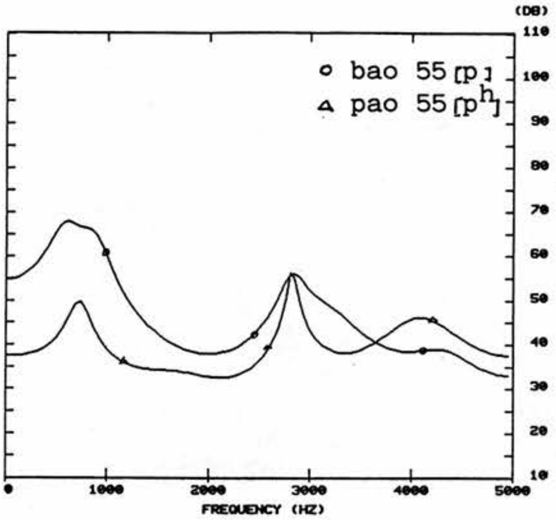


Figure 3.3(a-b) Examples of power spectra sampled at the time window (25 ms) after release of stops in Mandarin Chinese.
bao 55 - pao 55 (top) gao 55 - kao 55 (bottom)



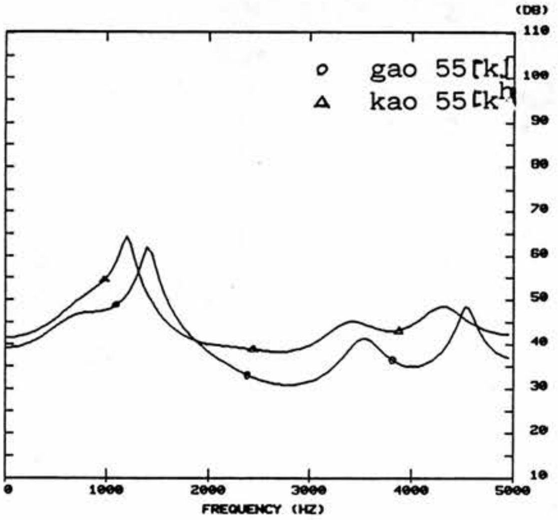


Figure 3.4 Examples of smoothed power spectra of two types of stops in Mandarin Chinese

As to the overall intensity level shown in Figure 3.4, it can be said that there is no consistent difference between the two types of stops. In some cases, however, voiceless unaspirated stops show a higher dB than aspirated ones, though this tendency may not be consistent. This means that the articulatory force, or specifically the force of air-flow is greater for unaspirated stops than for aspirated ones at the time of release.

3.3.4. The Onset Frequency of the First Formant

As mentioned in Chapter 1, there is a difference in the onset frequency of the first formant between voiced and voiceless stops. Stops in Mandarin Chinese are phonetically specified as voiceless stops and the difference between them is the degree of aspiration. The measurements of the onset F1 frequency were made to examine whether there is any difference in the initial transition of F1 between voiceless aspirated stops and voiceless unaspirated stops. Table 3.7 shows the mean onset frequency of F1 of the words listed in Table 3.1, and the vowels followed are /i, a, u/. Figure 3.5 shows some examples of formant patterns of Mandarin Chinese stops.

Table 3.7 Mean F1 Onset Frequency in Mandarin Chinese
(N=18 for bilabial, alveolar and velar)
(s.d. in parenthesis)

/p/	401.2 (56.0)	/p^h/	436.8 (50.1)
/t/	333.7 (44.6)	/t^h/	418.8 (27.8)
/k/	604.0 (229.2)	/k^h/	834.0 (125.6)

As seen above, there is a difference in the F1 onset frequency between voiceless unaspirated stops and aspirated stops, and the values for aspirated stops are higher than those of unaspirated ones. It can also be noted that velar stops show a considerable degree of variance, as shown in standard deviation. The difference in F1 onset frequency is considered to be due to the presence or absence of F1 initial-transition in the brief period after release. In the production of voiceless aspirated stops, as mentioned in Chapter 1, the onset of periodicity is delayed because of the large amount of turbulent airflow, and the F1 transition is not apparent for the initial brief period, and this may result in a higher onset F1 frequency.

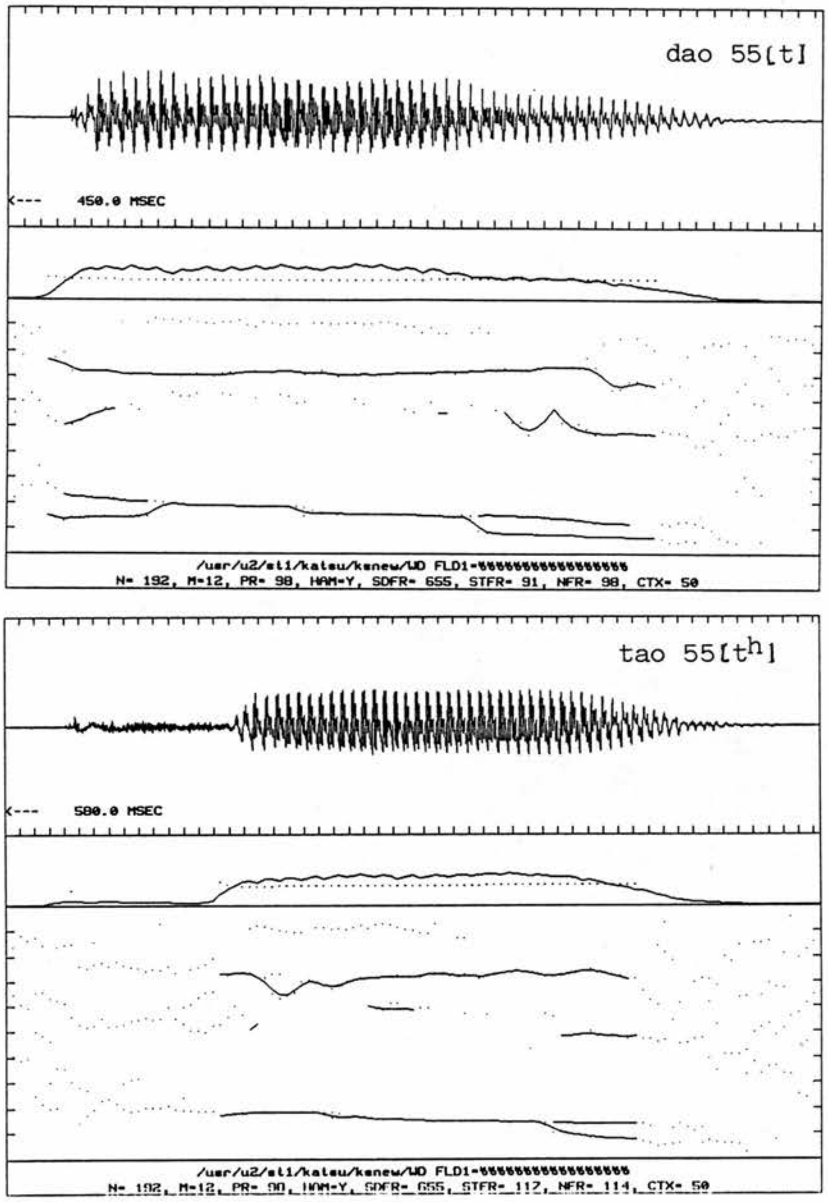


Figure 3.5 Examples of formant patterns of Mandarin Chinese stops

3.4. Discussion

The results of the present acoustic analysis indicate that the two types of stops in Mandarin Chinese can be distinguished by four acoustic dimensions: VOT, Fo, spectral characteristics and the onset F1 frequency, though the data of power spectra are less apparent.

The VOT results clearly indicate that there is a difference between the two types of stops; voiceless unaspirated stops show an average VOT of 11 ms and voiceless aspirated stops show an average VOT of 99 ms. The difference is statistically significant (two-tailed t-value 18.66, $p < .001$). This means that there exists a tendency to divide the timing continuum of laryngeal activities into two regions, and the results might give further evidence that VOT is an effective measure for distinguishing aspirated stops from unaspirated ones. Although there have been no data on VOT of stops in Mandarin Chinese, the results show the typical characteristics of voiceless unaspirated and voiceless aspirated stops.

According to physiological data by Iwata and Hirose (1976), it was mentioned that for voiceless unaspirated stops the vocal folds are nearly closed during the articulatory closure, but are not in vibration. This indicates that the glottis is in a position to vibrate immediately after the release, and it can be said that this state of the glottis corresponds to a slight delay of voicing onset. For voiceless aspirated stops, they found that the glottis is wide open and the timing for the maximum opening of the glottis is always after the release. This explains why onset of voicing is considerably delayed; that is, it takes some time to have a sufficient transglottal pressure difference and to initiate the vocal folds vibration.

It should also be noted that there is no correlation between vowel height and VOT. Although there is a report for English (Klatt, 1975) that VOT and vowel height are correlated to each other, vowel effect on VOT is not consistent, and the variations in laryngeal timing in Mandarin Chinese are not due to the following vowel.

Next, the Fo data at the onset reveals that there is a difference between the two types of stops; voiceless aspirated stops show a higher Fo than voiceless unaspirated stops. Although the difference at the onset is less than 10 Hz, it is statistically significant (two-tailed t-value = 4.45, p <.01). As to the Fo curves, it can be seen that there is a difference in relative value of Fo between the two types of stops, but there is no difference in the pattern of the curves; both types of stops show a fall at the onset, as one would expect when both are voiceless. The various hypotheses as to the cause of Fo perturbations of vowels immediately following the consonantal release were discussed earlier. In the aerodynamic hypothesis, the airflow is considered to be initially high for voiceless stops, resulting in the increased rate of vibration at the onset, and as the flow rate decreases in the course of time, the Fo will accordingly show a fall. The present data also shows that the two types of stops affect Fo even 80 ms after vowel onset. Hombert et al. (1979:42) point out that this may cast some doubt on the aerodynamic hypothesis, since the aerodynamic perturbation is not considered to last for such a long period. However, since there was no measurement of airflow rate in the present study, nothing can be said conclusively about it.

Furthermore, the examination of spectral characteristics also provides some cues to distinguish the two types of stops, though the effect may be less apparent. It can be observed that voiceless unaspirated stops show more regularly distributed small peaks which are considered to represent the periodicity of glottal pulsing, while in aspirated stops small peaks are less regularly distributed. As to the overall intensity level, it was observed that voiceless unaspirated stops show greater intensity than voiceless aspirated stops. It is indicated in the study of Korean (Kim, 1970) that aspiration requires a large amount of airflow, resulting in greater intensity level, but this did not hold in the case of stops in Mandarin Chinese. Although there is a distinction between fortis - lenis for the two types of stops; aspirated stops are fortis and unaspirated stops are lenis (Dow, 1972), this distinction was not observed in the intensity level. This implies that in the articulation of voiceless aspirated stops, the glottis may not be opened as wide as generally considered to allow a greater volume of airflow, and the laryngeal timing to the glottal opening is more relevant than the degree of glottal opening.

Lastly, it was observed that there is a difference between voiceless aspirated stops and voiceless unaspirated stops in F1 onset frequency. F1 onset frequency is higher in voiceless aspirated stops than in voiceless unaspirated stops. The differences of F1 onset frequency are attributable to the ones in the onset timing of periodicity. In case of voiceless unaspirated stops in Mandarin Chinese, F1 starts immediately after the release, and shows a rising transition. On the other hand, in the case of voiceless aspirated stops, F1 starts at some later time and may not show a rising transition.

3.5. Summary

On the basis of the above acoustic analysis, the results can be summarized as follows:

Voiceless unaspirated stops show a short voicing lag (an average VOT of 11 ms), while voiceless aspirated stops show a long voicing lag (an average VOT of 99 ms). The values are distinctively distributed in the two voiceless categories in the dimension of laryngeal timing. Among the three places of articulation, the VOTs associated with velar stops are always longer than those associated with bilabial or alveolar stops. Furthermore, the voiceless aspirated stops show a higher Fo than voiceless unaspirated stops and there is a difference of about 10 Hz between the two types of stops. Two types of stop show a falling pattern of Fo curves in the initial period after the consonantal release. As to spectral analysis, voiceless unaspirated stops show, as a weak trend, more regularly distributed energy peaks and rather higher intensity level than aspirated ones. It was also observed that initial-stop differences in aspiration influence F1 onset transition characteristics, and the onset frequency is higher in voiceless aspirated stops than in voiceless unaspirated stops.

Chapter 4

Stops in Korean

4.1. Introduction

Korean has a three-way contrast in its stop consonants, and they differ in manner and point of articulation. There has been a good deal of discussion on the classification and nature of its stops for many years; what are the articulatory and acoustic properties of the stops, what are the phonetic features which differentiate them, and how are these features incorporated in the phonological framework ? Although there have been some disagreements among phoneticians and linguists on their classification, they are usually classified as voiceless, tense, unaspirated stops (Type 1 stops), voiceless lax, weakly aspirated stops (Type 2 stops), and voiceless, strongly aspirated stops (Type 3 stops). [1] Some investigators (Iverson:1983) simply call them fortis, lax, and aspirated stops, respectively, while others (Han & Weitzman:1970) call them strong, weak, and aspirated. The disagreements on the classification and terms of the stops reflect some difficulty in characterizing their phonetic features on purely phonetic grounds.

The contrast in Korean stops has been of interest among investigators, since it involves several modes of vocal fold adjustments, and it serves to clarify the coordinated mechanisms of some phonation types of speech sounds. Studies on Korean stops have been made from physiological, acoustic, and aerodynamic points of view, and the experimental investigations began with Kim (1965) who focussed on the phonetic nature of three types of stops from these angles. Since then, several people have engaged in experimental analysis and through this analysis, the nature of the three types of stops has been understood to

---

[1] There is a disagreement over labelling of the tensing feature to Type 3 stops. Kim (1965) classifies them [tense], while Lisker and Abramson (1964) classify them as [lax].

some extent. First, physiologically, Kim (1965) carried out an electromyographic test of bilabial stops and reported greater muscle activity of the lips for Type 1 stops. Hirose et al. (1974) examined the activities of the intrinsic laryngeal muscles and found that these muscles serve to differentiate the three stops and Type 1 stops are specifically characterized by the activity of the thyroarytenoid muscle. Kagaya (1974) examined the glottal width and timing of the closure by using a fiberscope, and reported that there is a considerable difference among the three types of stops in glottal width; the smallest glottal width for Type 1 stops, the intermediate for Type 2 stops, the largest for Type 3 stops. He also found that there is a difference in the timing of the closure relative to oral release; the maximum opening of the glottis occurs at the time of oral release for Type 3 stops, while the glottis is approximated well before oral release for Type 1 stops. Next, acoustically, there have been some studies, and Han and Weitzman (1970) examined several properties such as voice onset time (VOT), fundamental frequency (Fo), and intensity for characterizing major types, and indicated that the timing of voice onset and intensity build-up are most relevant for differentiating the three stop types. Furthermore, aerodynamically, Dart (1987) measured the air pressure and oral flow of Type 1 and Type 2 stops, and found that Type 1 stops are produced with higher intra-oral pressure before release, but a lower flow after release. Hardcastle (1973) investigated some acoustic and aerodynamic properties of initial stops to examine a feature "tensity" and suggested that the feature "tensity", defined in terms of isometric muscular tension in the vocal folds and pharynx, can be effectively used to explain the properties of the stops. Through these studies, it can be summarized that the three types of stops in Korean differ in laryngeal muscle activities, glottal width, and timing of closing relative to release in articulatory terms, and these differences in articulation are reflected in those of acoustic features such as VOT, Fo, and intensity. The problem in these studies is the lack of examination on how these features are coordinated and interrelated with each other, and how the coordinated activities are manifested in a more detailed acoustic measurement. The present study was first undertaken as part of a research project to characterize the phonation types of some Asian languages, and to examine acoustic properties of the three-way contrast of Korean stops. It is also aimed at examining how acoustic properties are correlated with articulatory dimensions in the production of the stops.

4.2. Experimental Procedure

4.2.1. Subjects

The subjects in the present study are three native speakers of Korean, all male, and are speakers of the Seoul dialect. Two of them are postgraduate students of the University of Edinburgh, and the other one resides in Edinburgh. Their ages range from 29 to 40 years.

4.2.2. Linguistic Materials

As mentioned, Korean has three major types of stops, and, specifically nine stop phonemes, all voiceless, are classified into three points of articulation; bilabial, alveolar and velar. These stops occur in word-initial, word-medial, and word-final positions, and undergo allophonic variations in various phonetic circumstances. Although there are no voiced stops in their phonemic inventory, Type 2 stops have voiced counterparts in intervocalic position, except in a few cases.[2] Linguistic material for recording in the present study includes some minimal triplets, differing in the initial stop, some minimal pairs, and some words in various vowel environments. They can be shown as follows:

---

[2] According to Cho (1967), Type 2 stops become voiced in most of the intervocalic positions, except such a sequence as /o_a/ or /u_ə/.

p*ul	"horn"	pi	"rain"
pul	"fire"	p^{h}i	"blood"
p^{h}ul	"grass"	pam	"night"
t*am	"sweat"	p^{h}al	"arm"
tam	"wall"	p*a	"born"
t^{h}am	"envy"	tal	"moon"
kon	"zero"	t*al	"daughter"
k^{h}on	"bean"	to:l	"stone"
k*ul	"honey"	kat	"hat"
kul	"oyster"	kot	"place"
kilda	"be long"	t*uda	"float"

Table 4.1 List of Korean words

These words were written in Korean characters by one of the subjects, and were written in random order on the first sheet, and were arranged in triplets or in pairs on the second sheet. Each word in both sheets was read three times for recording in a natural way by three speakers, making the total tokens 198. The recording was made in the sound-proof recording room in the Phonetics Laboratory, University of Edinburgh.

4.2.3. Acoustic Analysis

The procedure of acoustic analysis is exactly the same as described in the general experimental procedure in Chapter 1.

4.3. Results

4.3.1. Voice Onset Time (VOT)

The measurement of VOT was made for the interval between the consonantal release and the onset of voicing for three types of stops. Table 4.2 indicates the mean VOT values (msec) for three types of Korean stops. [3]

Table 4.2 Mean VOT Value of Korean Stops (ms)
(N=24 for tense, 12 for lax and aspirated)
(s.d. in parenthesis)

Type 1	Type 2	Type 3
p* 10.3(13.5)	p 31.1(12.7)	p^h 86.1(7.9)
t* 11.3(7.5)	t 20.0(4.3)	t^h 85.0(17.3)
k* 23.3(10.4)	k 49.2(4.4)	k^h 100.3(10.1)

It can be found from Table 4.2 that VOT values increase in the order from Type 1 stops to Type 3 stops, and there is a clear-cut difference between Type 3 stops and other two types of stops. It is apparent that Type 3 stops are produced with considerable delay of voicing and are strongly aspirated. It can also be found that there is a difference between Type 1 and Type 2 stops, but the difference is not great. As the standard deviation shows, there is an overlap between these two types of stops in some cases. These results and tendencies are in general agreement with previous studies (Han & Weitzman, 1970; Hardcastle, 1973). Furthermore, it can be noted that velar stop consonants in each type show the longest delay of voicing among the three places of articulation. The tendency for velar stops to show a relatively longer voicing delay than bilabial and alveolar stops is also found in other languages such as English and Japanese (see p. 27). [4]

---

[3] In measuring the VOT value, it is usually understood that the onset of voicing is synchronous with the onset of Fo. In the present measurement, however, it was found that there is a difference between the onset of voicing usually defined as VOT and the onset of Fo in some cases. Figure 4.6 (inserted at the end of this chapter) illustrates this difference between them, and the interval between R and OV is the one usually defined as VOT, and the interval between R and OF may be called Fo onset. It is noted in Figure 4.6 that Fo starts 10 - 20 msec earlier than the generally defined VOT. In some cases such as aspirated velar stops the difference between VOT and Fo onset is not negligible, i.e., Fo starts about 30 msec earlier than the periodicity. It is not clear why the difference came out in some cases, but this might be due to some ILS sampling procedure.

[4] Hardcastle (1973) pointed out that there are two reasons for velar stops having longer VOT than bilabial and alveolar stops. See Hardcastle (1973) p.266.

4.3.2. Fundamental Frequency (Fo) and its Contour

The measurement of the fundamental frequency (Fo) values was made at vowel onset and the steady-state portions of the vowel. Tables 4.3 (a,b) indicate the mean value associated with each stop.

Table 4.3a Mean Fo Value at Vowel Onset(Hz)
(N=24 for tense, 12 for lax and aspirated)
(s.d. in parenthesis)

Type 1	Type 2	Type 3
p* 169.8(26.9)	p 138.7(34.0)	p^h 186.1(28.7)
t* 152.3(34.4)	t 137.0(30.1)	t^h 165.8(30.3)
k* 172.5(28.5)	k 146.3(30.4)	k^h 167.3(30.9)

Table 4.3b Mean Fo value at Steady-State of Vowel(Hz)
(N=24 for tense, 12 for lax and aspirated)
(s.d. in parenthesis)

Type 1	Type 2	Type 3
p* 163.5(36.7)	p 145.1(40.1)	p^h 169.4(35.7)
t* 149.8(34.5)	t 131.0(29.4)	t^h 158.0(32.4)
k* 169.8(36.0)	k 146.3(35.5)	k^h 160.3(30.3)

The above tables indicate that Fo values in vowel onset and steady-state portions following Types 1 and 3 stops are considerably higher than those of Type 2 stops. For example, the pooled mean of Type 1 and Type 3 stops is 20 - 50 Hz higher than those of Type 2 stops. It appears that there is a consistent difference between Type 2 stops and Types 1 and 3 stops, and it seems reasonable that Fo values in vowel onset and steady-state portions serve to differentiate Type 2 stops from other stop types. There are several reasons for a higher Fo, and, following the general views on Fo perturbations, they may be attributable to increased tension of laryngeal muscles and increased air-flow rate in the transglottal area.

Associated with the difference in the onset of Fo values and steady-state Fo values, it is generally known that there may be a difference in the Fo curve for the initial period after stop release. In the present study, Fo curves were examined for 36 tokens of triplet words, in which the initial stop differs, and they are schematically shown in Figure 4.1.

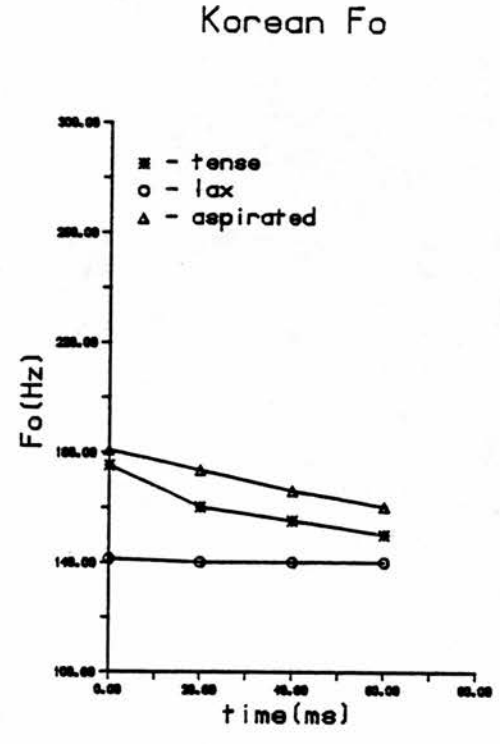


Figure 4.1 Normalized Fo contour of three types of Korean stops

The horizontal line represents time and the vertical line represents Fo frequency. From the examination of the curves, it can be said that Fo curves after Type 1 stops begin at a higher Fo region and abruptly fall in the initial period of 20 - 30 msec after release. The curves after Type 2 stops begin at a relatively low region and show a rather level pattern in the initial period. The ones after Type 3 stops begin at a higher region and show a steady lowering in the initial period after release. It can be said, therefore, that there is a difference in the Fo curves for major types of Korean stops, and the distinctive difference is that Type 1 stops show an abrupt fall after release and Type 3 stops show a steadily falling pattern. Although there are several factors which influence Fo curves, the relevant ones here seem to be inner tension of laryngeal muscles and the air-flow rate in the transglottal area. From the previous studies (Hirose et al.,1974), it is known that there is a sharp increase of tension in the intrinsic laryngeal muscles in the production of Type 1 stops, and the relaxation of tension may be responsible for the abrupt movement of the Fo curve. Hardcastle (1973:269) also pointed out that there is the relatively increased isometric tension in the folds for this type of stops. Furthermore, in the production of Type 3 stops, it is known that the glottis is maximally open, and that the airflow rate is considerably higher than for Types 1 and 2 in the period immediately after the release and then decreases as the air goes through the glottis. These aerodynamic factors may be related to the steady falling of Fo curve from the higher position. From this examination, there seem to be several different mechanisms responsible for the higher range of Fo in Type 1 and Type 3 consonants.

4.3.3. Spectral Analysis

In the present study, the short time spectral analysis was made by ILS linear prediction for the initial period of 30 msec after release. A total of 32 power spectra were sampled to examine the overall intensity and spectral characteristics. Figure 4.2 shows the smoothed power spectra of triplet words. It is known that the power spectra of the burst of the consonants might reveal invariant cues for the place of articulation, but no study of this has so far been made on Korean stops.[5] The initial period includes both the burst and some portion of the voicing for Type 1 and Type 2 stops, and for Type 3 stops includes mainly the burst and some portions of aspiration, but does not extend to vowel onset.

The examination of overall intensity in power spectra indicates that the intensity of Type 3 stops, as seen in [p^{h}ul], is considerably higher than that of other types and is in the range of 55 - 70 dB. The intensity for Type 1 stops is in most cases lower than in other types of stops, though the effect is not consistent.

Furthermore, Han and Weitzman (1970) showed that there is a difference in the time of build-up of intensity after the onset of voicing, and reported that there is a shorter time build-up for Type 1 stops than those of other types. However, as seen in Figure 4.4, the observations of intensity curves in the present study do not appear to show a consistent effect for intensity build-up, and there are some variations among subjects. As a weak trend, however, Type 3(aspirated) stops require a shorter time for build-up of glottal intensity than Type 1 and Type 2 stops, and show a rather sharp rise of intensity curve.

---

[5] Blumstein & Stevens (1979) pointed out that the spectrum sampled in a10 - 20 msec. time window in English contains invariant properties of place of articulation.

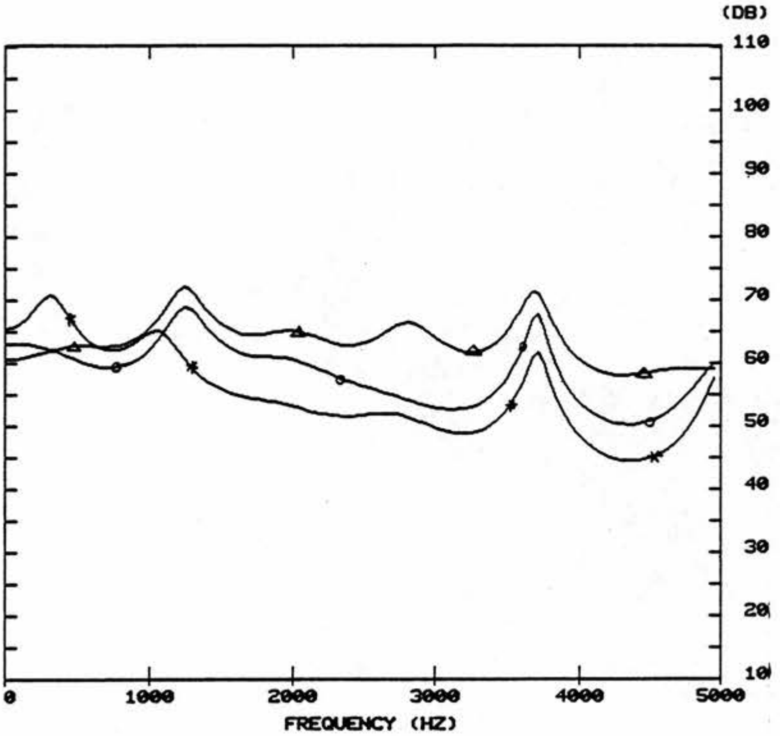


Figure 4.2 Example of smoothed power spectra sampled at the time window (30 ms) after release of Korean stops

* [p*ul] ∘ [pul] △ [p^hul]

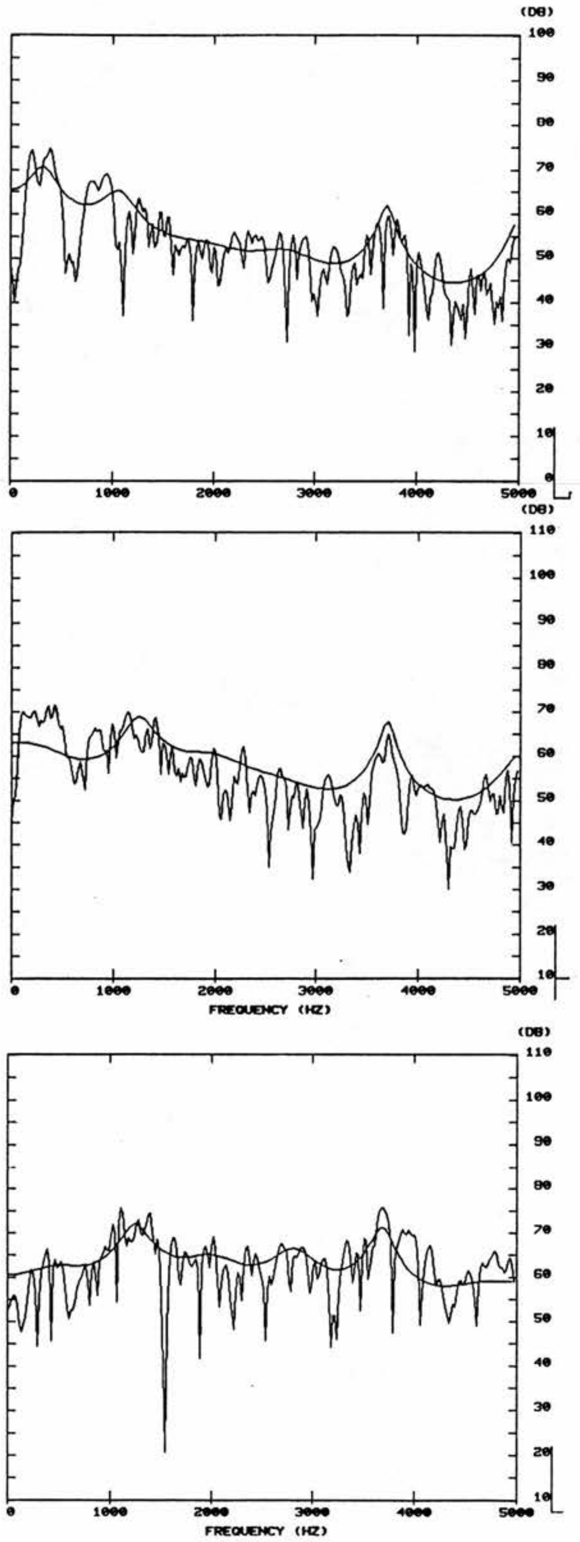


Figure 4.3 Power spectra sampled at the time window (30 ms) after release of Korean stops
[p*ul] (top) [pul] (centre) [p^{h}ul] (bottom)

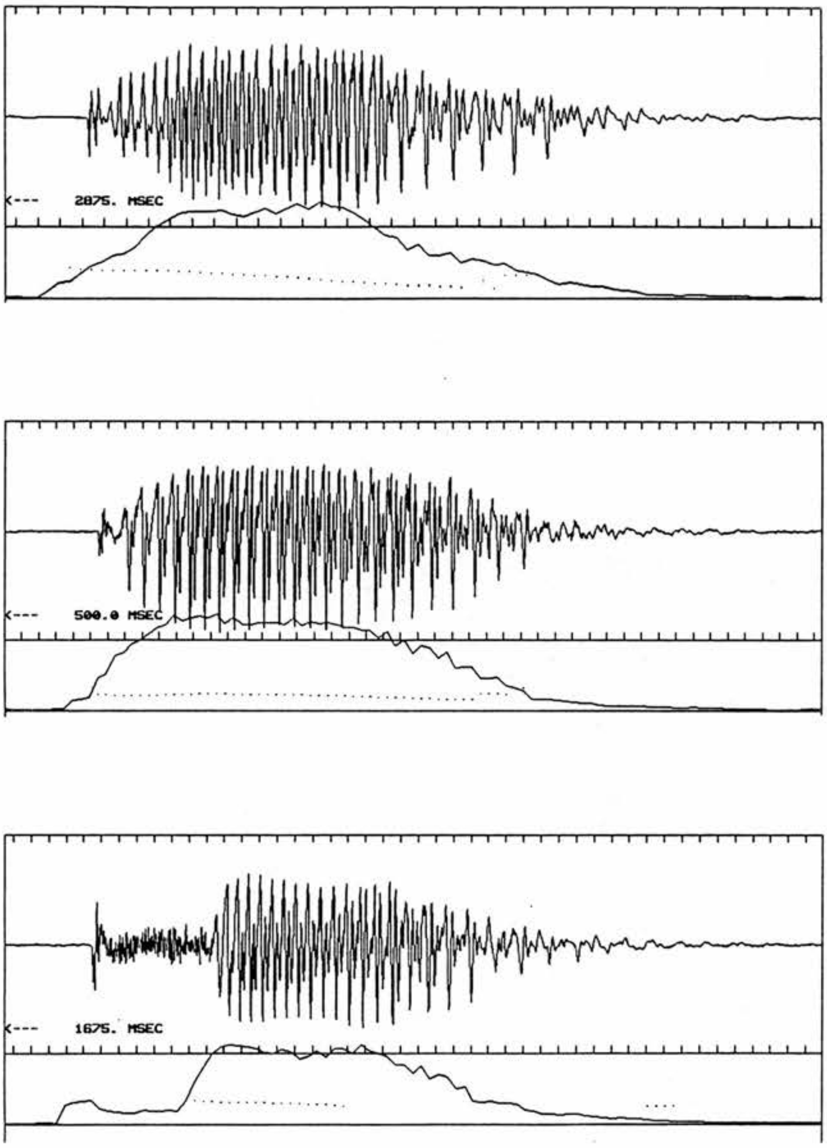


Figure 4.4 Waveforms and intensity patterns of [p*ul](top), [pul](centre), [p^hul](bottom)

4.3.4. The Onset Frequency of the First Formant

As mentioned in previous chapters, it is known that initial stop voicing has an influence on F1 onset characteristics of the following vowels; vowels following voiced stops contain rising F1 transition, while vowels following voiceless stops may not contain F1 transition. Korean stops are phonetically realized as voiceless ones in initial position. The measurements of F1 onset frequency were made to examine whether there is any difference in the frequency between three types of voiceless stops. Table 4.4 summarizes the mean F1 onset frequency for the words listed below for three subjects.

Table 4.4 Mean Onset F1 Frequency in Korean Stops (Hz)
(N=9 for each word)

Type 1 stops		Type 2 stops		Type 3 stops	
p*ul	431.3	pul	357.0	p^{h}ul	287.5
t*am	667.0	tam	824.0	t^{h}am	852.0
		kon	529.5	k^{h}on	603.3
k*ul	307.8	kul	294.5		
X	468.7		501.3		580.9

As can be seen above, F1 onset frequency differs in the amount of variation per consonant type and the following vowel. For example, the onset frequencies of the consonants followed by vowel /a/ show a higher value than the ones followed by other vowels such as /o, u/. It can also be noted that the mean value for the three types of consonants is higher in voiceless aspirated stops (Type 3) than in the other two types of stops. The difference between aspirated stops (Type 3) and voiceless lax stops (Type 2) is significant (two tailed t-value=2.82 $p < .01$), but the one between voiceless tense stops (Type 1) and voiceless lax stops (Type 2) is not(t-value=0.27, n.s.). As shown in VOT values, the onset periodicity starts earlier in Type 1 stops than in the other two types of stops, and this early start of periodicity results in a lower onset frequency of F1 in Type 1 stops.

4.4. Oral Airflow

Oral airflow is considered to be one of the parameters which are influenced by a change of glottal adjustments, and the state of airflow reflects some glottal states to some extent. In a previous study on Korean stops, Dart (1987) measured the oral air pressure and rate of "fortis" (Type 1, voiceless tense) and "lenis" (Type 2, voiceless lax) stops, and showed that fortis stops are produced with higher oral pressure before release, yet lower airflow after release than lenis stops. In the present experiment, we recorded oral and nasal airflow to examine whether there is any difference between the three types of Korean stops.

The experimental procedure for collecting airflow data is the same as described in section 2.5. The subject is a native speaker (male) of Korean who took part in the recording of linguistic materials for acoustic analysis described in 4.2.1. Linguistic materials for examining oral airflow were as follows:

Type 1		Type 2		Type 3	
p*ul	"horn"	pul	"fire"	p^hul	"grass"
t*am	"sweat"	tam	"wall"	t^ham	"envy"
		kon	"zero"	k^hon	"bean"

Figure 4.5 shows oral (and nasal) airflow data for three stop types of Korean words. Horizontal lines represent time dimension, and vertical line represents the airflow. As is seen in Figure 4.5, there is some variation among the same utterances. But there is a tendency in the airflow for the three types of stops. Type 3 stops show a considerably higher rate of airflow than other two types of stops, and Type 2 stops (lenis) show a slightly higher rate of flow than Type 1 stops (fortis). The difference in the airflow is due to the one of glottal state for three types of stops, and the highest rate of airflow for Type 3 stops is considered to be due to the maximum opening of glottis, whereas the lowest rate for Type 1 stops is due to the least opening.

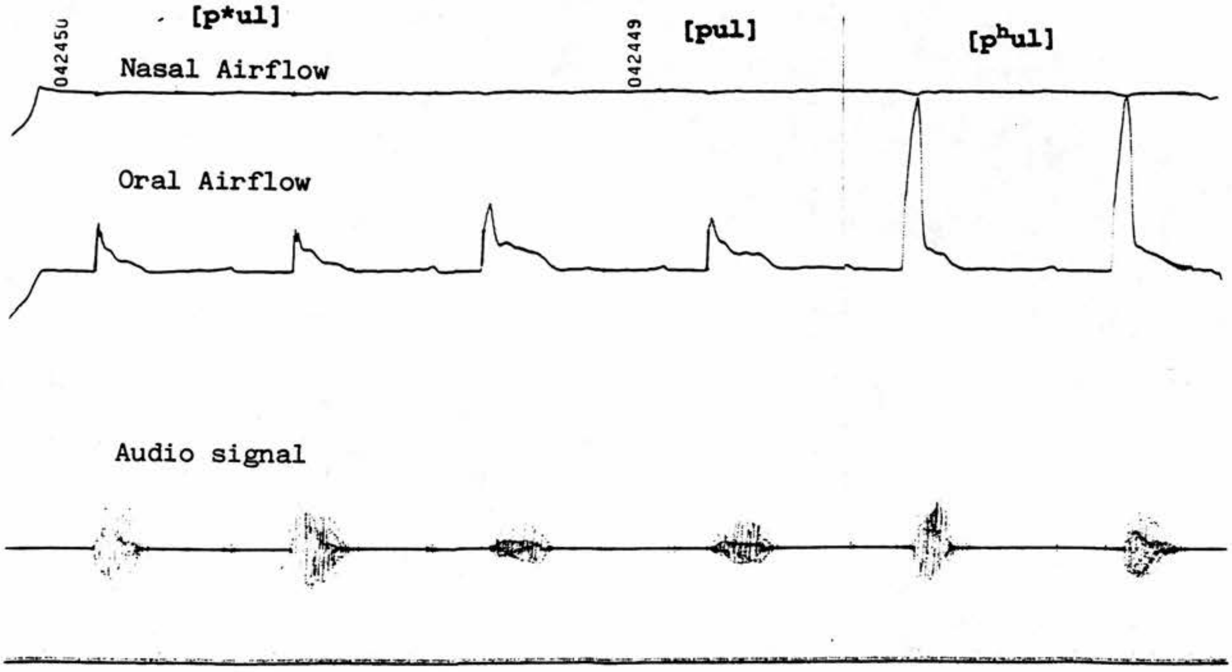


Figure 4.5 Inkwriter traces of airflow for Korean stops
[p*ul] (left) [pul] (centre) [p^hul] (right)

4.5. Discussion

The results of the present study show that several acoustic dimensions must be used to characterize the differences of a three-way contrast of Korean stops and that a single acoustic dimension is not sufficient to differentiate the three types of consonants. Based on the acoustic analysis of VOT, Fo and the curve, power spectra and F1 onset frequency, the three types of Korean stops are acoustically characterized as follows:

Type 1 (voiceless, tense, unaspirated stop)

Type 1 stops show the shortest value of VOT among three types and show a relatively high Fo at the vowel onset and steady-state. The Fo curve starts at a high region in the initial period following the release, and falls abruptly in the period of 20 to 30 msec. after release. In power spectra, the intensity is rather weaker than those of the other two types of stops. They show the lowest F1 onset frequency between the three types of stops.

Type 2 (voiceless, lax, weakly aspirated stop)

Type 2 stops show the intermediate value of VOT between Type 1 and Type 3 stops, and show overlapping with the values of Type 1 stops. The Fo onset value is lower than those of other types and the curve starts at a lower region and shows a level pattern. The intensity of power spectra is rather weak, but is slightly higher than Type 1 stops. F1 onset frequency is higher than that of Type 1 stops, but lower than that of Type 3 stops.

Type 3 (voiceless, strongly aspirated stop)

Type 3 stops show a characteristically longer value of VOT than those of Type 1 and Type 2 stops. The Fo value shows a higher value and the curve starts at a higher frequency region and steadily falls in the period of about 100 msec. after release. The intensity of the power spectra is greater for Type 3 stops than it is for Type 1 and Type 2 stops. F1 onset frequency is higher than for the other two types of stops.

The three-way contrast in Korean stops requires several acoustic dimensions to differentiate one from the others, and it will be significant to consider what articulatory dimensions are associated with the observed differences in acoustic dimensions for the distinction. In order to do so, it is necessary to examine previous findings on the physiological and aerodynamic properties of Korean stops, and the studies which are relevant here are Kagaya (1970), Hirose et al. (1974) and Dart (1987). Their major findings and the results of the present study can be shown as follows:

	Type 1	Type 2	Type 3
Physiological factor			
Glottal width*	small	intermediate	maximum
Timing of glottal closing*	Before release	Almost synchronous to release	After release
EMG**	Sharp increase of vocalis	Less suppression of adductor muscles	Suppression of adductor muscles
Aerodynamic factor			
Airflow***	Lower flow rate after release than Type 2		
Airflow****	Lowest	Medium	Highest
Intraoral Pressure***	Higher than Type 2		
Acoustic factor****			
VOT	10 - 20 msec	30 - 50 msec	85 - 100 msec
Fo	Higher value	Lower value	Higher value
Fo contour	Abrupt falling	Level	Gradual falling
Intensity	Low dB	Medium dB	High dB
F1 onset	Lowest among three types	Middle	Highest among three types

* From Kagaya (1970)
** From Hirose et al. (1974)
*** From Dart (1987)
**** From the present study

Based on the physiological and aerodynamic findings in the previous studies, it can be considered that the observed differences in VOT are mainly attributable to the differences in glottal width and the timing of glottal adjustments; for Type 1 stops, the glottal width is smaller than it is for other types, and the glottis is approximated before release, resulting in the least delay of onset, while for Type 3 stops the glottis is maximally open at the time of release, and it takes some time to configurate the vocal folds for vibration, resulting in the considerable delay of onset.

The higher Fo at the vowel onset and steady-state portions in Type 1 and Type 3 stops may be attributable to aerodynamic mechanisms in the transglottal and supralaryngeal areas. Dart's study suggests that there is a marked increase in intra-oral pressure before release for Type 1 stops, and this increase may be relevant to the higher Fo of Type 1 stops. It can also be considered that vocal fold tension is relevant to the sustained period of higher Fo after release and the abrupt drop in the curve (Hardcastle, 1973; Hirose et al., 1974). For Type 3 stops, it is presumed that the glottis is maximally open at the time of release, and the higher Fo after release may be caused by the larger amount and rate of airflow. The curves show a gradual decline from release, as the difference in transglottal pressure decreases.

The examination of acoustic features of Korean stops reveals that VOT, Fo, intensity and F1 onset frequency are necessary to characterize the contrast, and, in particular, the distinctions between Type 1 and Type 2 stops are hard to pin down. Although Type 1 stops are often said to be laryngealized, glottalized, or tense, and there have been some reports about tensity, their exact nature, particularly their physiological and acoustic correlates, has not yet been clarified. If, as Hirose et al. (1974) claim, there is an increase of inner tension of intrinsic laryngeal muscles, specifically thyroarytenoid, then the acoustic correlates may be defined as a higher Fo at vowel onset and an abrupt change in Fo curves. Furthermore, if VOT is used as the main parameter to determine voicing characteristics of stops, then Type 1 stops are in the category of "voiced stops". It can be considered that if a language has three categories of voicing of stops and all stops are phonologically specified as [-voice] as in Korean, then one of the types possesses the VOT value which is close to the "voiced category" in generally-accepted laryngeal timing. In order to keep the distinctiveness of voicelessness, other features may be more relevant, and in Korean such a distinction involving tensing of laryngeal muscles is maintained between Type 1 and Type 2 stops, though further physiological evidence is needed. Although it is known that Type 2 stops become voiced in intervocalic position, Type 1 stops are also used in loan words in Korean from English to represent voiced sounds in English such as [k*ɔlf] > golf, [t*ɑns] > dance, etc.[6] It can be considered that the replacement with Type 1 stops of English voiced stops is due to the phonetic similarity between Type 1 stops and voiced stops in English, in particular the proximity of VOT values between the two languages.

4.6. Summary

To summarize, the three types of Korean stops are acoustically characterized by several dimensions such as VOT, Fo and the curve, power spectra, and F1 onset frequency. Type 1 stops, conventionally termed fortis or laryngealized stops, are defined as having the shortest VOT (10 - 20 msec), higher Fo at the vowel onset, abrupt falling in the Fo curve, and low F1 onset frequency. Type 2 stops, conventionally termed lax stops, are defined as having intermediate VOT (30 - 50 msec), lower Fo and level curve, and intermediate intensity in power spectra. Type 3 stops, aspirated stops, are defined as having the longest VOT (85 - 100 msec), higher Fo and successive falling in the contour, higher intensity in power spectra, and higher F1 onset frequency than others. The differences in acoustic properties are considered to result from coordinated activities or adjustments of the intrinsic laryngeal muscles, glottal width, and aerodynamic mechanisms. In terms of VOT, Type 1 stops are completely within the domain of the "voiced" category, and some additional parameters such as tensing or laryngealization keep it as voiceless.

---

[6] These examples were mentioned in Kim (1965), and Kim also pointed out that Type 2 stops might be used in some cases to replace the voiced stops in English in loan word formation.

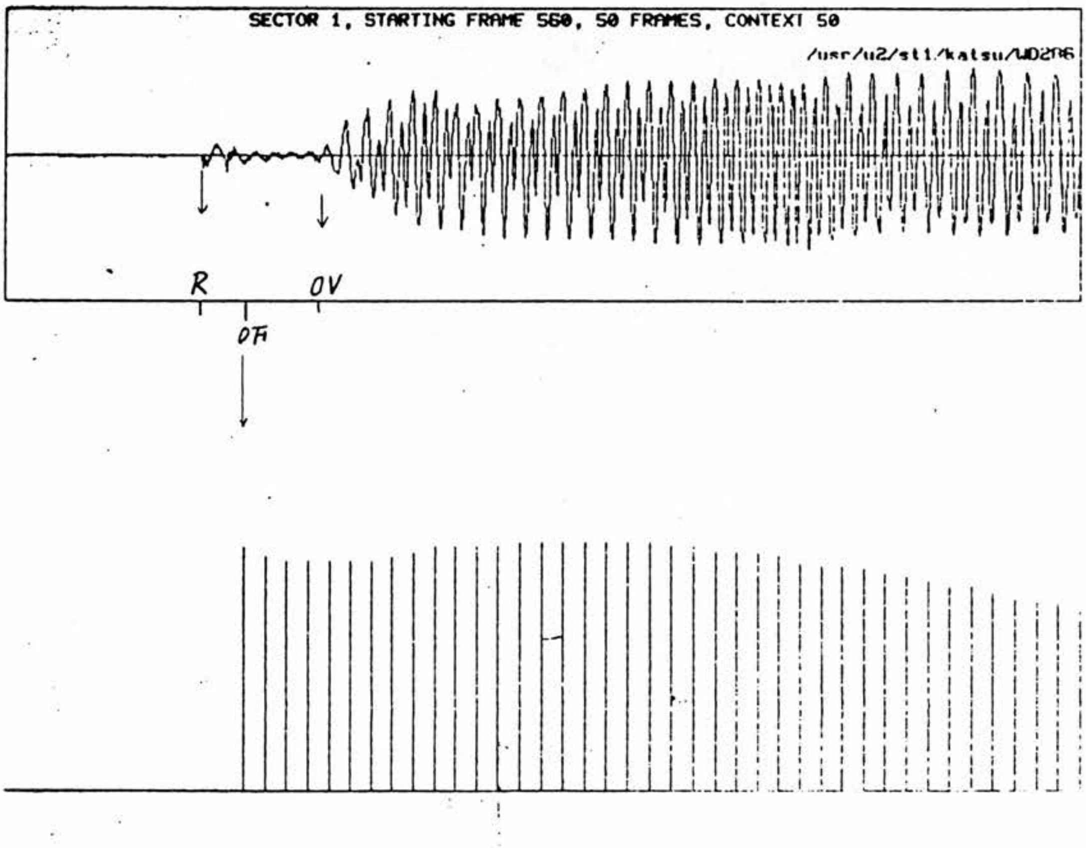


Figure 4.6 Waveform (top) and Fo contour (bottom) of [p*ul] in Korean (R - release, OF - onset of Fo, OV - onset of voicing)

Chapter 5

Stops and Nasals in Burmese

5.1. Introduction

The present chapter deals with stops and nasals in Burmese. Burmese is known to have a three-category voicing contrast for stop consonants; voiced stops /b, d, g/, voiceless unaspirated stops /p, t, k/, voiceless aspirated stops /p^h, t^h, k^h/. It is also well known that it has a phonemic contrast between voiced and voiceless nasals as well as laterals. Burmese and its related languages have been of interest to some phoneticians and linguists, since they make tonal and nontonal contrasts across their languages, and their tone systems offer an opportunity for historical and comparative studies, i.e.; the tone systems are considered to provide interesting facts for establishing proto-forms in Tibet-Burman language family. Although there have been studies of the phonetic properties of Burmese, most of them have dealt with the tones and their contrasts. Matisoff (1973) examined the tone development or tonogenesis in these languages and reported that some consonant types have specific effects on tones and their development. Tun (1982) made an acoustic analysis of tones using the spectrograph and found that fundamental frequency (Fo) and total length of tones are important in differentiating tones. Furthermore, Maddieson (1984) measured the fundamental frequency at vowel onset following voiced and voiceless sonorants and examined the effect and implications of the voicing contrast for Fo perturbations and the splitting of tones. These investigators have attempted to examine the phonetic and acoustic features of tones and their implications for tonogenesis, and little has been done on the phonetic and acoustic characteristics of Burmese segments. The present chapter is mainly concerned with the acoustic features of stops and nasals in Burmese and specifically aims at examining acoustic properties which differentiate the voicing categories of stops and nasals.

5.2. Stops in Burmese

5.2.1. Subjects

The subjects in the experiment are two native speakers of Burmese, one male and one female, and both are from Rangoon. The male speaker (Subject 1) is an English teacher and has been in Edinburgh for about six months to attend the EFL courses at College of Education. The female speaker (Subject 2) is a graduate of University of Edinburgh and majored in Applied Linguistics. She has been in Edinburgh for a long period. Their ages range from 44 to 47 years.

5.2.2. Linguistic Materials

As mentioned before, Burmese has a three-category voicing distinction in initial stops, and both manner and place of articulation serve to separate nine stop phonemes from each other; namely a voiced series, a voiceless unaspirated series, and a voiceless aspirated series for bilabial, alveolar, and velar. It also has tonal contrasts, as other Tibet-Burman languages do. But as the literature shows, several investigators classify them differently; McDavid (1945) classifies the tones into five types, while Stewart (1955), Cornyn and Roop (1968), and Okell (1969) classify them into four. This difference in classification stems from the fact that Burmese tones are not restricted only to pitch changes, but also involve some additional features of voice quality such as creaky voice. The following words illustrate a three-way category of voicing contrast and some tone types.[1]

---

[1] Burmese has four tonal contrasts, and they are represented here as tone 1 (-), tone 2 (:), tone 3 (.), and tone 4 (') following the notation by Cornyn and Roop (1968). They characterize each tone as follows: tone 1 is low, level and long, tone 2 high, long and falling, tone 3 high, short, and falling, and tone 4 high, extremely short. (p.xv).

Table 5.1 List of Burmese words

	Voiced stops		Voiceless unaspirated		Voiceless aspirated	
Labial	bi:	'comb'	pin-	'tree'		
	bo:	'football'	pou'	'rotten'	p^{h}ou'	'roast'
	be:	'duck'	pei-	'dirty'	p^{h}e'	'embrace'
Alveolar	do:	'anger'	te.	'straight'	t^{h}a:	'to place'
	da:	'knife'	taun:	'ask for'	t^{h}aun	'pound'
Velar	gu-	'cave'	ku:	'cross'	k^{h}a:	'bitter'
			kou:	'nine'	k^{h}ou:	'steal'

Words in minimal pairs or near triplets:

ba-	'what'	pa-	'accompany'	p^{h}a-	'basket'
da:	'knife'	ta:	'prevent'	t^{h}a:	'place'
gaiŋ	'sect'	kaiŋ	'bend'	k^{h}aiŋ	'order'
bu:	'container'	pu:	'united'	p^{h}u:	'view'
du:	'knee'	tu:	'dig'	t^{h}u:	'to answer'
gu-	'cave'	ku-	'to assist'	k^{h}u-	'caterpillar'
bo:	'football'	po:	'silk'	p^{h}ou'	'roast'
dou'	'stick'	toun'	'blunt'	t^{h}ou'	'bundle'
goun-	'head'	ko-	'body'	k^{h}ou-	'to steal'

Four tones:

1. Level	ti-	'earthworm'	pou-	'extra'
2. Falling	ti:	'to play the piano'	pou:	'insect'
3. Abrupt Falling	ti.	'to make it level'	pou.	'to send'
4. Glottal Ending	ti'	'one'	pou'	'rotten'

The words in Table 5.1 were checked by one of the subjects and were written in Burmese scripts. They were recorded five times as written here, three times in isolation and two times in a carrier sentence: "du. ti yaa: kyen ___" ("The second time is ___"). Each subject was instructed to place a short pause before each test item in the carrier sentence. The recording was made in the sound-proof recording room in the Phonetics Laboratory, University of Edinburgh. Two tokens (the second item in isolation and one token in the carrier sentence) were prepared for acoustic analysis. It does not appear that there is any significant difference in results between items in isolation and those in the carrier sentence.

5.2.3. Acoustic Analysis

The procedure of acoustic analysis is exactly the same as described in general procedure in Chapter 1.

5.2.4. Results

Acoustic analysis was made of voice onset time (VOT), fundamental frequency (Fo), power spectra, and the onset frequency of the first formant.

5.2.4.1. Voice Onset Time (VOT)

The measurement of VOT was made for the interval between the onset of voicing and release of oral closure. Table 5.2 presents the mean and range of VOT associated with each of the stop types for each subject.

Table 5.2 Mean VOT value and range for each stop (ms)
(N=6 for bilabial, alveolar and velar)
(s.d. in parenthesis)

Subject 1(male)											
/b/	-104	(12.4)	-90/-120	/p/	3	(4.2)	0/10	/pʰ/	48	(13.0)	35/70
/d/	-106	(12.5)	-95/-125	/t/	15	(6.3)	10/25	/tʰ/	67	(9.6)	55/85
/g/	no data			/k/	29	(2.5)	25/30	/kʰ/	76	(16.2)	55/95
Subject 2(female)											
/b/	11	(5.3)	5/20	/p/	11	(4.6)	5/20	/pʰ/	108	(18.1)	80/100
/d/	14	(9.2)	0/25	/t/	17	(2.6)	15/20	/tʰ/	128	(20.9)	100/155
/g/	30	(7.7)	25/45	/k/	33	(12.9)	15/55	/kʰ/	143	(12.5)	125/160

It can be seen from Table 5.2 that the two subjects differed considerably with respect to the voiced category. For subject 1, the voiced stops are uttered with voicing lead, while for subject 2 the voicing lead was missing throughout her utterances. For subject 1, there is a clear-cut difference in VOT for each voicing category. The voiced stops, though no data was obtained for /g/, show a longer voicing lead, voiceless unaspirated stops show a simultaneous or slightly delayed voicing, and voiceless aspirated stops show a long delay of voicing. For subject 2, there is a considerable overlap between voiced stops and voiceless unaspirated stops, though there exists a difference between voiceless aspirated stops and the other two types of stops. Subject 2 did not make any difference between the two categories with respect to VOT in most of her utterances. Although the absence of voicing lead is not uncommon in the realizations of voiced stops in other languages, it is an interesting result that the two subjects showed different realizations in the production of voiced stops and the result raises the question of whether voicing lead is a relevant and crucial feature for identifying voiced category.[2]

The VOT values in relation with the difference of tone types were measured for two sets of words, mentioned in Table 5.1. The mean VOT values of each word are presented in Table 5.3.

---

[2] Subject 2 has been in English-speaking community for a long time, and her realization of laryngeal timing might be affected by acquiring English.

Table 5.3 Mean VOT values of voiceless unaspirated stops as a function of tone types (ms) (N=6 for each tone type)

Level	ti- and pou-	19 ms
Falling	ti: and pou:	15
Abrupt Falling	ti. and pou.	15
Glottal Ending	ti' and pou'	13

It can be seen in Table 5.3 that VOT value is somewhat long in the level tone and it is short in the glottal ending, but the difference is not so great. Although there is a weak difference in VOT between the tone types, it does not appear that the difference of tone types affect the VOT values.

5.2.4.2. Fundamental Frequency (Fo) and its Contour

The measurements of fundamental frequency values were made at the vowel onset following a stop consonant. The words examined were the ones of nearly triplets and minimal pairs with the same tone type to minimize the effect of tone differences. Table 5.4 shows the average values of Fo associated with each stop consonant. Since male and female subjects show a different range of Fo perturbation, the data is shown for each subject.

Table 5.4 Mean Fo Value at Vowel Onset (Hz) (N=6 for bilabial, alveolar and velar) (s.d. in parenthesis)

Subject 1 (male)					
/b/	168 (7.2)	/p/	176 (5.0)	/pʰ/	185 (6.6)
/d/	165 (7.1)	/t/	195 (17.1)	/tʰ/	190 (15.5)
/g/	no data	/k/	189 (4.0)	/kʰ/	191 (8.3)
Subject 2 (female)					
/b/	176 (10.9)	/p/	185 (20.8)	/pʰ/	208 (16.3)
/d/	176 (7.6)	/t/	209 (13.8)	/tʰ/	204 (11.6)
/g/	171 (7.9)	/k/	190 (1.7)	/kʰ/	202 (14.1)

Although it is pointed out that there is a tendency to minimize the effect of a preceding consonant on Fo in a tone language, there appears to be a difference between voiced and voiceless stops.[3]

In examining the words in near triplets such as [bo:], [po:] and [p^hou'], it can be noticed that voiceless stops have a relatively higher onset Fo value than voiced stops: the Fo values following the voiceless series are 30 - 40 Hz higher than those following the voiced series. In the words of minimal pairs of voiceless aspirated and unaspirated stops in which the initial stop differs, the Fo value is higher for voiceless aspirated stops than it is for voiceless unaspirated stops and the difference is about 9 Hz (two tailed t-value=1.63, n.s.) in subject 2, but for subject 1, the difference between them is not evident.

Associated with the differences in the onset Fo values, the Fo curves were examined to see whether there is a difference between the three voicing categories of stops in the direction of the Fo curve. The words examined are limited to the ones having a tone of Type 2 (long and falling) in order to minimize the tone effect, and are shown below:

da:	"knife"	ta:	"to prevent"	t^ha:	"place"
bu:	"container"	pu:	"united"	p^hu:	"view"
du:	"knee"	tu:	"to dig"	t^hu:	"to answer"

The measurements of Fo were made at 20 ms intervals from vowel onset to 100 ms for the analysis data of Subject 2, and the average Fo values were calculated at each interval point and Figure 5.1 shows the averaged curves on the following vowel. It can be seen in Figure 5.1 that there is a difference in the direction of Fo curves between voiced stops and voiceless stops. Voiced stops start at a lower Fo region and show a slightly rising pattern.

---

[3] Hombert (1978) refers to the findings on Thai tones and points out that this tendency (to minimize the effect of preceding consonant on Fo) is "to render the different tones maximally perceptually distinct" (p.83).

Voiceless aspirated stops show an initially high and a slight falling pattern, and voiceless unaspirated stops also show an initially high and a tendency towards a slight falling or level pattern. The direction of these curves of voiced and voiceless stops are in agreement with those reported in Hombert et al. (1979). The point we should note is that the two types of voiceless stops (unaspirated and aspirated) did not show any marked differences, and this implies that the initial glottal state and aerodynamic conditions are similar in these two types of stops.

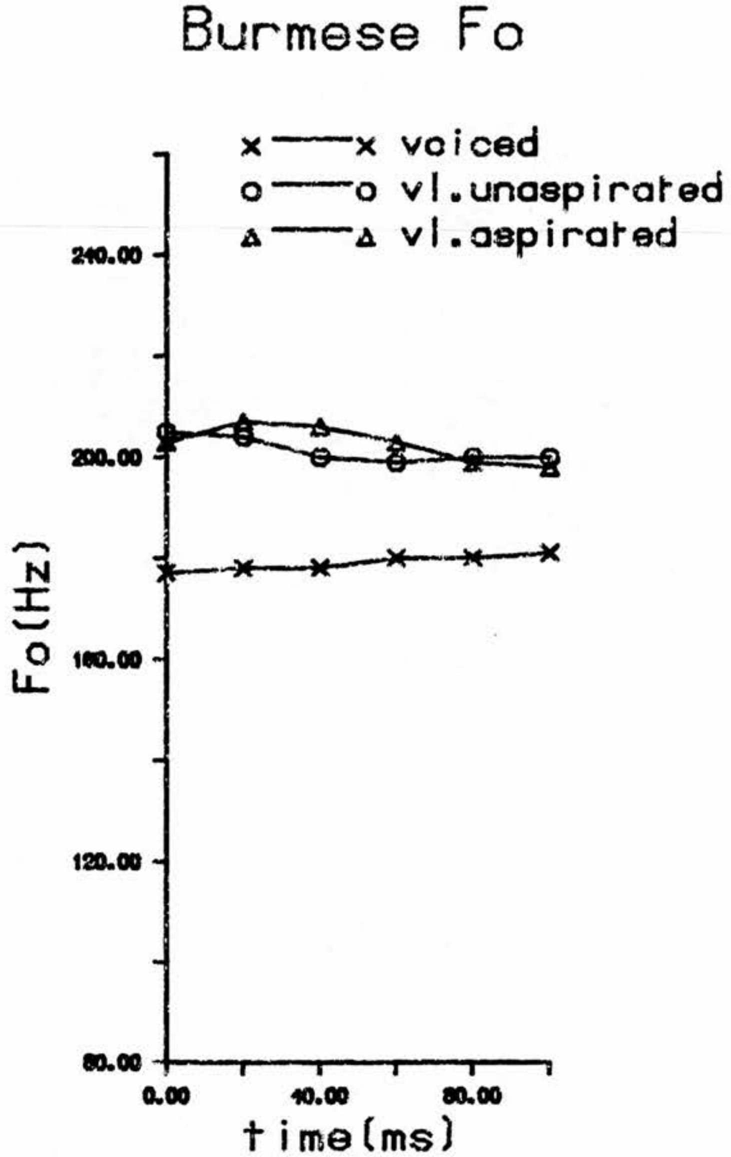


Figure 5.1 Averaged Fo contours of a three-way of Burmese stops

5.2.4.3. Spectral Analysis

In order to examine the overall intensity and spectral characteristics of voicing contrast, the short-time spectra were sampled at the consonantal release of the stops. The words examined are listed in the following table and three tokens for each word were sampled.

ba-	'what'	pa-	'accompany'	p^ha-	'basket'
da:	'knife'	ta:	'prevent'	t^ha:	'place'
gaiŋ-	'sect'	kaiŋ-	'bend'	k^haiŋ-	'order'
bo:	'football'	pou'	'rotten'	p^hou'	'roast'
da:	'knife'	taun:	'ask for'	t^haun:	'pound'

Some of the examples of power spectra are shown in Figure 5.2. These are smoothed spectra using a 25 msec time window beginning at the release. Since burst and onset of voicing varies across consonants, the short-time window includes both burst and some voicing portions for voiced stops and voiceless unaspirated stops, while for voiceless aspirated stops the burst and some portions of aspiration were examined.

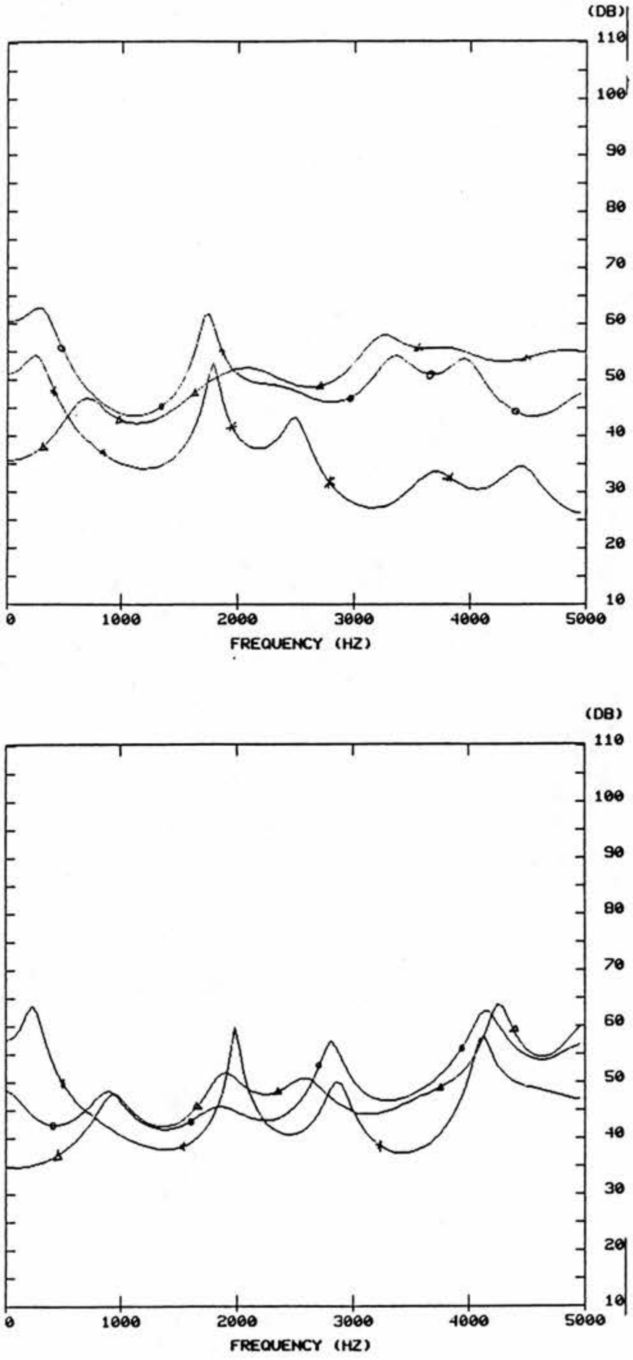


Figure 5.2 Examples of smoothed spectra of Burmese stops
(top) Subject 1 * [da:] ∘ [taun:] ▲ [t^{h}aun:]
(bottom) Subject 2 * [da:] ∘ [ta:] ▲ [t^{h}a:]

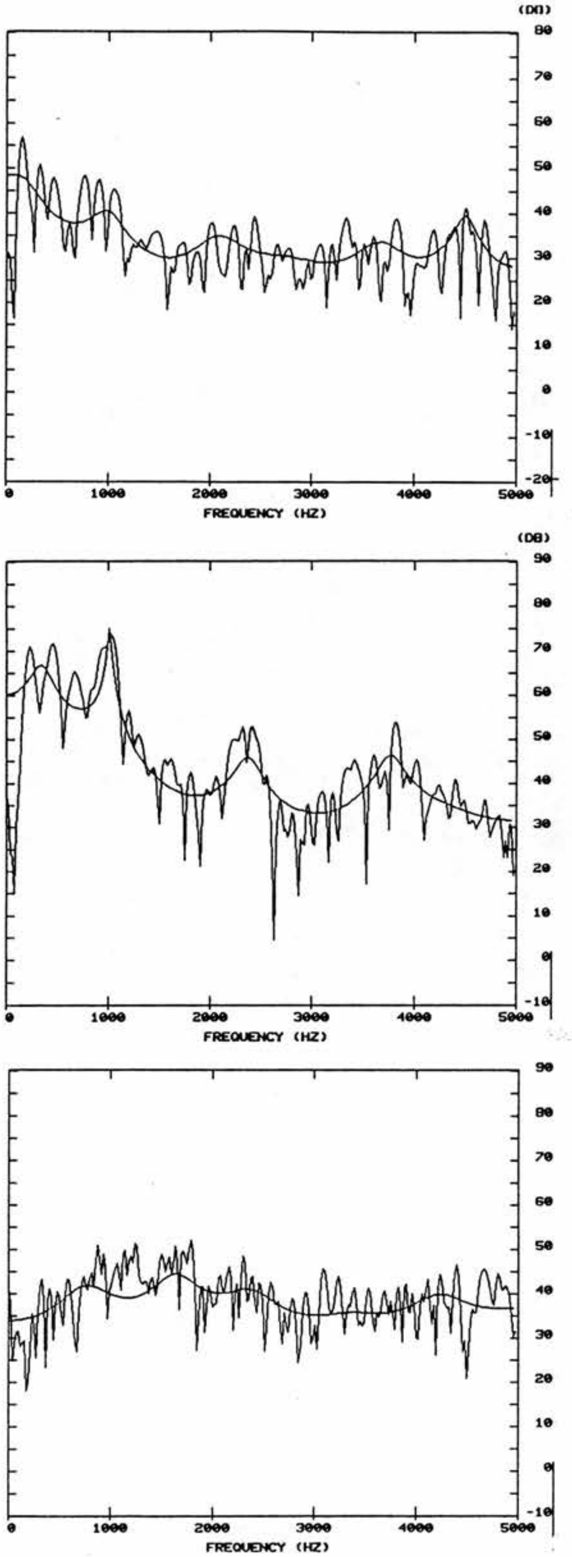


Figure 5.3 Power spectra of Burmese stops sampled at the time window(30 ms) after release [bo:](top) [pou'](middle) [p^hou'](bottom)

As shown in Figure 5.2, there is a weak trend for the overall intensity for the voiceless type consonants to be greater than that of voiced ones; for example, for some alveolar stops there is a difference of 10 - 20 dB between voiced and voiceless stops. For voiceless stops, there is no major difference in the intensity level between aspirated and unaspirated stops, and aspirated stops show a rather level pattern in the smoothed spectra. These imply that the aspirated stops in Burmese are not produced with a higher articulatory force; specifically rate of airflow. Furthermore, as shown in Figure 5.3, there is a difference in the regularity of spectrum energy between voiced and voiceless stops. Although the degree of regularity is difficult to measure or quantify, the voiced stops and voiceless unaspirated stops have apparently more regularity in energy distribution. There is also a difference in the bandwidths of the harmonics, as far as they are measured from power spectral analysis. The voiceless aspirated stops show a narrower bandwidth than voiced stops and voiceless unaspirated stops. The energy peaks are considered to provide information on place of articulation, and there are peaks in the vicinities of around 1, 2.5, and 3.5 kHz for bilabial stops, and 1.7 - 2 and 3 - 3.5 kHz for alveolar stops.

5.2.4.4. The Onset Frequency of the First Formant

The measurements of F1 onset frequency were made to examine whether there is any difference in the onset frequency between voicing categories of stops in Burmese. Table 5.5 summarizes the mean F1 onset frequencies of the triplet or near triplets words with the same vowel.

Table 5.5 Mean F1 Onset Frequency in Burmese Stops (Hz)
(based on Subject 2)

ba-	830.5	pa-	788.0	p^ha-	935.0
da-	228.5	ta-	264.5	t^ha-	936.0
be:	291.5	pei-	348.5	p^he'	579.0
du:	232.7	tu:	298.5	t^hu:	252.0
gu-	148.0	ku-	232.0	k^hu-	200.0

(N=6 for each word)

As shown in Table 5.5, there is some degree of variance in F1 onset frequency, depending on the difference of following vowels. It is apparent, however, that there is a clear difference between voiceless aspirated stops and other types of stops, and F1 onset frequencies are higher in voiceless aspirated stops than in voiced stops and voiceless unaspirated stops, except the triplets followed by vowel /u/. It also can be noted that there is no noticeable difference in F1 onset frequency between the triplets having vowel /u/. For utterances having low vowel /a/, F1 frequencies are higher than those having high vowel /u/. In other words, the difference in initial stop voicing is more apparent in utterances having higher F1 frequencies (non-high vowels) than those having lower F1 frequencies (high vowels).

5.2.5. Discussion

The results of the acoustic analysis indicate that the three major categories of voicing of Burmese stops differ from each other in acoustic dimensions such as VOT, Fo, power spectra and F1 onset frequency, and there are some variations between two subjects in these realizations of the categories.

VOT apparently serves to separate all three voicing categories from each other in subject 1, but not in subject 2. In subject 2, voicing lead is lacking in most of the voiced utterances and voiced stops show similar VOT values to those of voiceless unaspirated stops. This may raise the question of the significance of voicing lead as a relevant cue for voicing distinction. Furthermore, it can be noticed that there is a difference in the VOT values of voiceless aspirated stops between the two subjects: subject 1 shows a medium length of voicing lag, while subject 2 shows a considerably long lag and has a wide range of variability in the value. This indicates that in subject 2 aspirated stops are articulated with considerable force like those generally found in Hindi or Korean, but this is not the case for subject 1. What is commonly found in the two subjects is that VOT increases as the place of articulation moves from bilabial to velar position, and this is a tendency which is commonly found in other languages. As is suggested in previous chapters, this may be attributable to smaller volume of the supralaryngeal cavity for velars; that is, if other factors were kept constant, the time to attain the transglottal pressure difference sufficient for starting voicing would be longer than other types of stops.

The measurement of fundamental frequency at the vowel onset showed a difference between voiced and voiceless stops; the Fo values following voiceless stops are relatively higher than those following voiced stops. But there was no marked difference between voiceless unaspirated and aspirated stops, though aspirated stops tended to show a higher value than unaspirated ones. To explain the lowering of Fo for voiced stops, if we follow the assumption proposed by Hombert et al. (1979), we may speculate that the lowering of the larynx causes a lowering of Fo for sustaining voicing during the oral closure. It is rather interesting to note that subject 1 did not show a consistent difference between voiceless aspirated and unaspirated stops. Although aspiration is considered to accompany a high rate of airflow and causes the raising of Fo, this is not the case in subject 1.[4]

The spectral analysis showed that there is a difference in the overall intensity between voiced and voiceless stops; the intensity for voiceless stops is relatively higher than it is for voiced stops, but there is no noticeable difference between voiceless aspirated and unaspirated stops. The observations of these spectra show that there are several differences between the three types of stops in the energy distribution and bandwidths, but these differences are rather hard to quantify in the examination.

The results on F1 onset frequencies suggest that differences of initial-stop voicing are cued by the one of the onset frequencies. Vowels following voiceless aspirated stops show

---

[4] Hombert (1978) mentions as follows: "... voiceless aspirated stops give rise to a higher Fo at the onset of the following vowel. This claim is supported by data from languages such as Korean, Thai, and Danish. However, conflicting data are found from the same as well as other languages" (p.87).

higher F1 onset frequency than vowels following voiced stops or voiceless unaspirated stops. F1 onset frequency differences, however, do not appear to play a role in cueing initial-stop voicing in utterances containing high vowels /i, u/.

5.3. Nasals in Burmese

5.3.1. Nasals

As mentioned before, Burmese has a phonemic contrast between voiceless and voiced nasals and has six nasal phonemes /m, hm, n, hn, ŋ, hŋ/ where voiceless nasals /hm, hn, hŋ/ are phonetically represented as [m̥, n̥, ŋ̊]. According to Okell (1969), the presence of the h-sound relate to some grammatical functions, and an h-initial mark indicates "transitive, active, or causative" correlates of the verb with a plain initial.[5] Although the voicing contrast of nasals is phonetically an interesting phenomenon, there have not been many studies on their phonetic and acoustic properties. The study which is relevant to the present one is Maddieson (1984), and he examined the effect of voiced and voiceless sonorants on fundamental frequency, but did not go into details of the nasals themselves. As to articulatory characteristics of voiceless nasals, Ladefoged (1971) mentions that "voicing began shortly before the closure was released and these sounds should be called partially voiced as opposed to fully voiced." The experiment here was undertaken to examine the acoustic properties which differentiate two types of nasals.

5.3.2. Linguistic Materials and Acoustic Analysis

The language materials consist of three minimal pairs in which an initial nasal contrasts with or without /h-/, as listed below. The two native speakers of Burmese mentioned above read the word list three times in isolation, together with other materials of stops. The recorded

---

[5] See Okell (1969) "...These pairs of verbs (one with an aspirate initial consonant and the other with the corresponding plain one) are called 'h/non-h' pairs. The relationship between the verbs in each pair is that the verb with an aspirate initial is the 'transitive', 'active', or 'causative' correlate of the verb with a plain initial;..." (p.42).

materials were analyzed using the ILS software package and Sonagraph 7800, and the analytical procedure is exactly the same as mentioned in Chapter 1.

hma'	'from'	-	ma'	'Miss'
hna-	'nostril'	-	na-	'pain'
hŋa:	'rent'	-	ŋa:	'fish'

5.3.3. Results

Figures 5.4 and 5.5 show some of the waveforms and wide and narrow-band spectrograms of the two types of nasals. The figures show several important differences between voiceless and voiced nasals. The first thing we notice is the difference in the duration of nasal murmur portions. The nasal murmur portions can be identified by the presence of a low F1 formant and a change of intensity pattern in Figure 5.4. The measurement of the murmur portions in both types of nasals was made in the display of waveforms and Table 5.6 presents the mean values for each word.

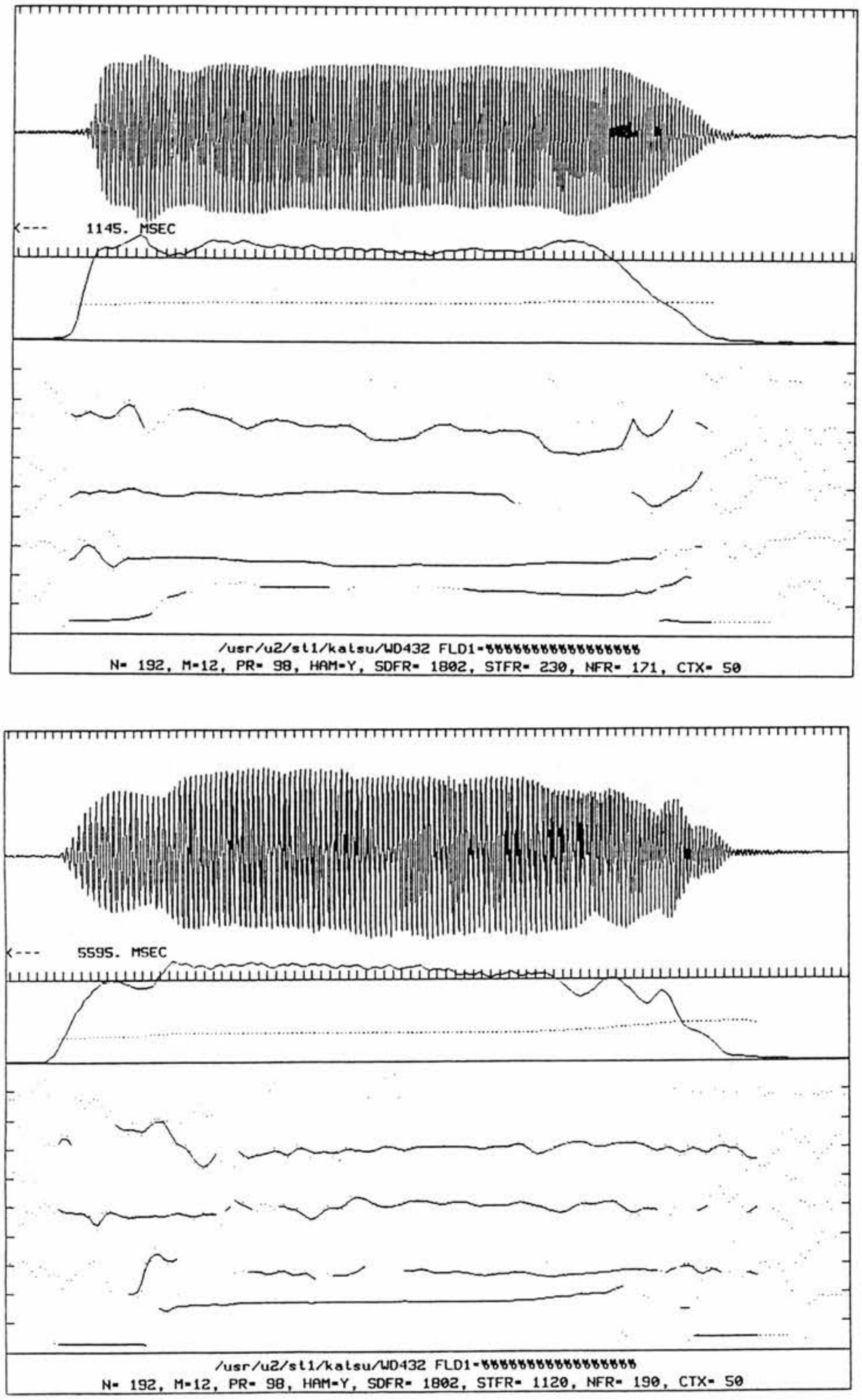


Figure 5.4 Waveforms and intensity patterns of [hna˧] (top) and [na˧] (bottom)

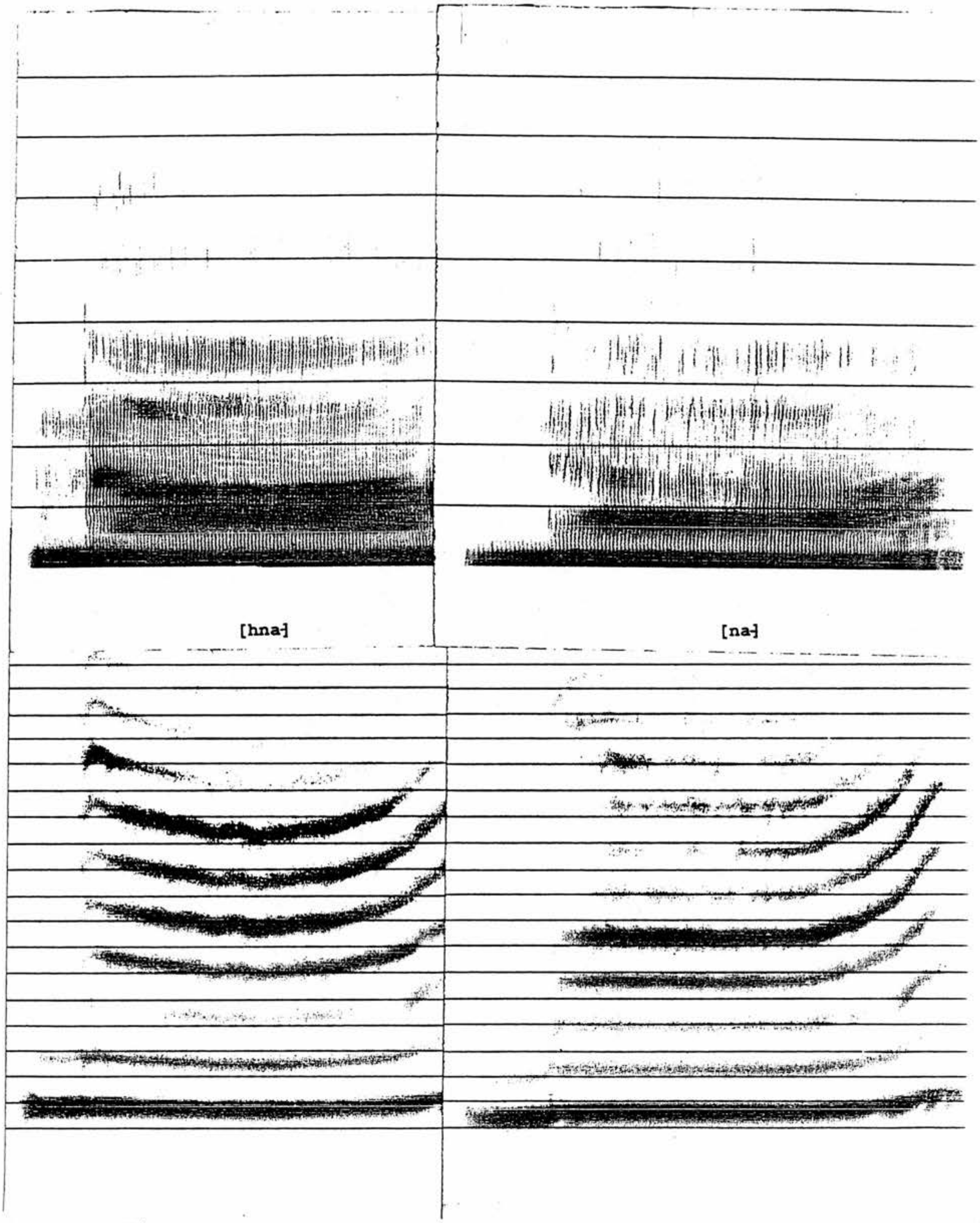


Figure 5.5 Wide and narrow-band spectrograms of [hna˧] and [na˧] in Burmese

Table 5.6 Mean duration of Nasal murmur portions in Burmese (ms)
(N=6 for bilabial, alveolar and velar)

Bilabial	hma'	69 ms	ma'	145 ms
Alveolar	hna-	75	na-	133
Velar	hŋa:	68	ŋa:	114

From the above, the duration of nasal murmur portions is considerably shorter in voiceless nasals than in voiced nasals. We presume that for the production of voiceless nasals air is passing through the nasal cavity before the vibration of vocal folds for the murmur portions.

Next, the measurements of the fundamental frequency (Fo) value were made at the vowel onset, and Table 5.7 presents the mean values for each word.

Table 5.7 Mean Fo value at vowel onset (Hz)
(N=6 for bilabial, alveolar, and velar)

Bilabial	hma'	214 Hz	ma'	158 Hz
Alveolar	hna-	196	na-	155
Velar	hŋa:	204	ŋa:	157

As seen in the above, the Fo value following voiceless nasals is markedly higher than that following voiced nasals, and there is a difference of about 40 to 60 Hz in the contrast between the two types of nasals. Generally the Fo value following voiceless sounds is higher than that following voiced ones, and a relatively higher onset of Fo in voiceless nasals may be viewed as an attribute of a higher airflow rate by [h].

Since it is considered that the characteristics of nasal murmur portions contain the relevant information for the distinction between voiceless and voiced nasals and for the distinction of places of articulation, spectral analysis was made in the 25 msec time window at the onset of nasal murmur portions. Figures 5.6 present the smoothed spectral envelopes of the contrast, and as seen in these examples, the spectral energy for voiceless nasals appears to be higher than that for voiced nasals, and in alveolar and velar nasals the higher energy is more apparent than in bilabial ones. As another measurement shows, there is a difference in the smoothed slope of the spectra; for voiceless nasals a falling slope is observed as the frequency goes up, while a level or less damped slope is observed for voiced nasals. Furthermore, there is a difference in the build-up of intensity associated with the two types of nasals. Apparently, as shown in the intensity patterns of Figure 5.4, a sharp rise or shorter amount of time is seen in voiceless nasals than in voiced nasals.

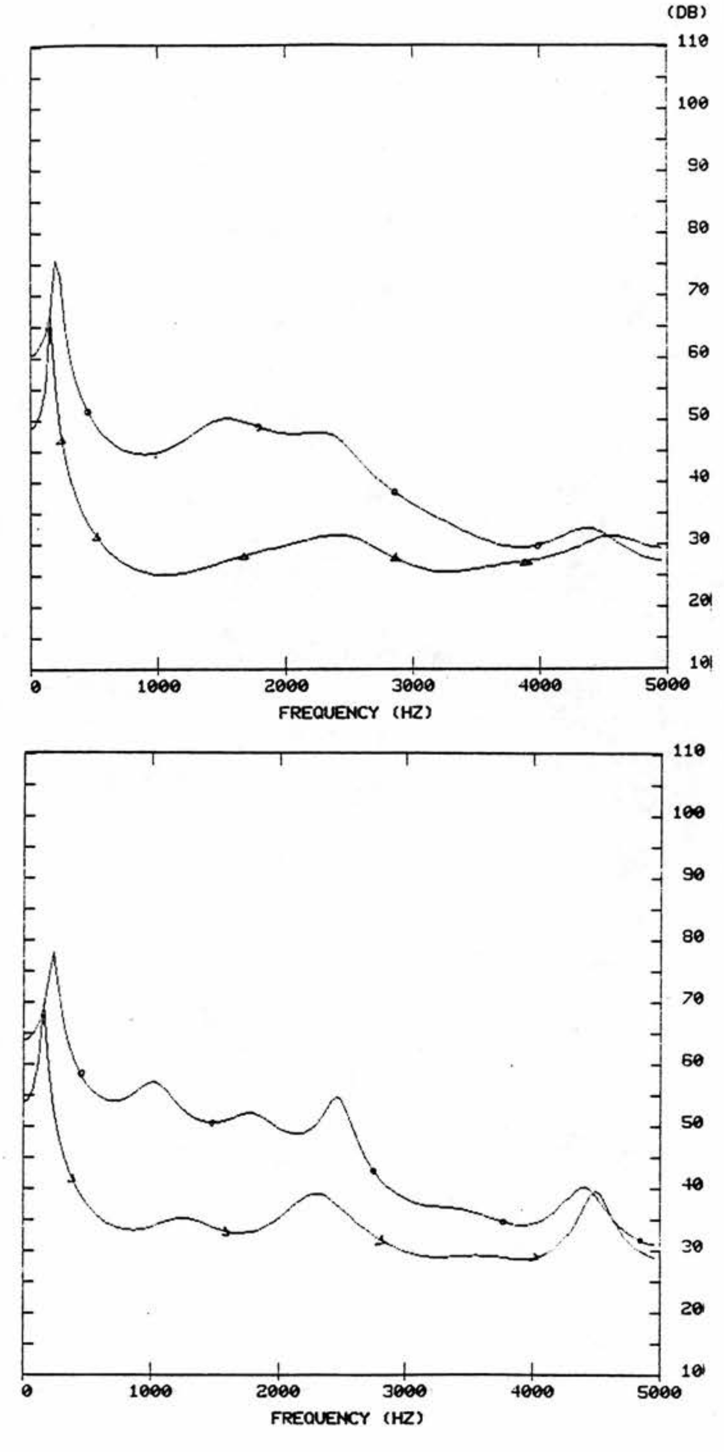


Figure 5.6 Smoothed spectra of voiced - voiceless nasals sampled at the onset of nasal murmur portions (25 ms). ∘ [hna-] △ [na-] (top) ∘ [hŋa:] △ [ŋa:] (bottom)

5.3.4. Discussion

The examination of acoustic properties of both types of nasals in Burmese reveals several differences in the duration of nasal murmur portions, in the Fo value at vowel onset, and in the spectral analysis. There is a clear difference in the duration of nasal murmur portions. As shown in Table 5.6, the duration of murmur portions for voiceless nasals is shorter than that of voiced nasals and is 45 - 60 % of the portions for voiced nasals. Based on these durational differences and Ladefoged's remark (1971), we may consider the articulatory events for the production of voiceless nasals as follows; while the articulatory configuration is being achieved in the oral cavity, the velum is lowered and the air goes through the nasal cavity. Then the vocal folds begin to vibrate for 60 - 90 msec (duration of murmur portions) before release of the oral closure, and at the time of oral release, the velum is raised for closure, resulting in no airflow in the nasal cavity. This durational analysis indicates that "voicelessness" of Burmese nasals consists of a very short voiceless portion in which air goes from the nose and the (voiced) nasal murmur portions before the oral release. This voiceless portion is considered to affect the Fo value and spectral characteristics.

The measurement of fundamental frequency at vowel onset also showed a difference between voiceless and voiced nasals. The Fo value following voiceless nasals is considerably higher than that following voiced nasals. Voicing contrast of nasals gives the same effect as stop consonants. This has a rather interesting implication for the voicing effect on Fo. According to Hombert (1978), it is pointed out that a voicing distinction of obstruents in prevocalic positions has an effect on Fo; voiceless stops raise Fo, while voiced stops lower Fo. Hombert (1978) argues the voicing effect of stops on Fo as follows: Although there are mainly two factors which affect Fo (namely, aerodynamic effect and vocal fold tension), it is the lowering of the larynx which causes the lowering of Fo.[6] The larynx lowers to increase the volume of supralaryngeal cavity to sustain the voicing during the oral

---

[6] See section 1.5 in Chapter 1.

closure. Although Hombert's argument is based on several experiments, the situation is rather difficult to explain in the case of voiced nasals. In the production of nasals,there is no need to enlarge supralaryngeal cavity to maintain sufficient airflow for voicing, since the air goes through the nose even if the oral cavity is closed. It is not clear at this stage whether there is any difference in the larynx height in Burmese nasals, but if the larynx lowers in the production of voiced nasals in Burmese (for the purpose of enlarging the cavity), then Hombert's argument that the need to enlarge the supralaryngeal cavity results in lowering of the larynx may need further revisions. Although there are several theoretical considerations on the relation between voicing contrast and Fo perturbations, the voicing contrast of nasals in Burmese will present some implications for the relation between a change of supralaryngeal cavity and Fo perturbations.

5.4. Summary

To summarize, the three types of voicing categories differ with respect to several acoustic dimensions and there are some variations between the two subjects in these dimensions. VOT serves to distinguish three types of voicing categories from each other in subject 1, but not in subject 2. The subjects differed considerably with respect to voicing lead in voiced stops: subject 1 always showed a realization of voicing lead, while subject 2 showed no "lead" realization. There is also a difference in the VOT values of voiceless aspirated stops between the subjects. Subject 1 showed a medium length of the VOT values and this indicates that the aspirated stops are not forcefully articulated, but rather loosely articulated by subject 1. From these, though there are some subject variations, the timing of voice onset can be an important cue for distinction of voicing categories from each other. There is also a difference in Fo values between voiced stops and voiceless stops, but it is not possible to ascertain the differences in Fo values for distinguishing voiceless unaspirated stops from aspirated ones. In the analysis of power spectra, as a weak trend, there is a difference in overall intensity between voiced and voiceless stops, and voiceless stops show a relatively higher intensity than that of voiced ones. Further, it can be noticed in subject 1 that voiceless aspirated stops are not forcefully articulated as in those in Korean or Hindi.

For voiced and voiceless nasals, there are several significant cues associated with the two types of nasals. The first cue is the duration of nasal murmur portions before release; the duration for voiceless nasals is considerably shorter than that of voiced nasals. The second cue is the Fo value at vowel onset; the Fo value for voiceless nasals is much higher than that of voiced nasals. So the voicing contrast of nasals exerts the same effect on Fo as the voicing contrast of stops. As discussed earlier, the lowering of Fo in voiced nasals has an interesting effect on the relation between Fo perturbations and larynx height. Although further evidence is needed, the explanation relating Fo perturbations to the difference in supralaryngeal configuration may not be supported, and the relation between Fo perturbations and supralaryngeal configuration may be revised. In the analysis of power spectra, the overall intensity of voiceless nasals is much higher than voiced nasals and the intensity patterns for voiceless nasals rise more abruptly than those of voiced nasals.

Chapter 6

Stops in Thai

6.1. Introduction

This chapter deals with stops in Standard Thai (hereinafter referred to as Thai). Thai has a three-way voicing contrast of stops, and the contrast involves two manner features - voicing and aspiration. The three-way contrast of initial stops in three places of articulation serves to separate eight stop consonant phonemes from each other; voiced bilabial and alveolar stops /b, d/; the voiceless unaspirated bilabial, alveolar, and velar stops /p, t, k/; and the voiceless aspirated bilabial, alveolar, and velar stops /p^h, t^h, k^h/.[1] Thai lacks the initial voiced stops in velar position, and this is known as a hole in the consonant inventory. All eight consonants occur in word initial position. Syllable-final stops are unexploded, and there is no opposition between three types of voicing in the word-final position. Abramson (1972) calls this a neutralization of the manner features at the end of a syllable. Like other languages in the South-East Asian linguistic area, Thai is a tonal contrast language and has five lexically contrasting tones which are usually termed mid, high, rising, low, and falling.

Thai has a number of phonetically interesting features which have attracted the attention of several linguists and phoneticians: what is the phonetic nature of word-initial and word-final stops, how are the tones and rhythm described, and what are the phonetic details of semi-vowels ? Considerable work on the description of Thai speech sounds has been done by Haas (1956) and Henderson (1964). Henderson investigated several aspects of Thai stop consonants, and, based on her auditory impressions, stated that the voiceless unaspirated

---

[1] Gandour (1974) pointed out that voiceless aspirated stops in Thai can be split into two groups: breathy type and plain type. In the present study, however, we will not divide them into two groups, since the material collected consists of all breathy type stops.

stops are articulated with tense phonation, while the voiceless aspirated stops are articulated with lax phonation.[2] Furthermore, a great deal of experimental work has been done by A. S. Abramson. Abramson (1962) made an extensive study on the acoustic properties of vowels and tones in Thai. Lisker and Abramson (1964) examined the three-way voicing contrast of stops in the word-initial position and presented the results of measurements of voice onset time (VOT). Abramson (1972) also examined the acoustic properties of the word-final stops which have been considered to be problematic in Thai phonetics. Through this descriptive and experimental work, the phonetic characteristics of Thai stops are now fairly well understood. Gandour and Maddieson (1976) examined the vertical movement of the larynx for varying pitch patterns and phonation types of consonants. They mentioned in their study that the larynx position for /p/ and /p^h/ is higher than that for /b/, and the pitch following /p^h/ is closer to the one following /b/ rather than /p/.

The present study will further examine the acoustic properties of the initial stops in Thai, and, based on the detailed acoustic analysis, will examine the underlying mechanism of the voicing contrast of stops. The study was undertaken as a part of research to examine the phonation types of some Asian languages and to investigate some cross-language characteristics of the voicing contrasts of stops.

6.2. Experimental Procedure

6.2.1 Subjects

The subjects in the present experiment are two native speakers (female) of Standard Thai, and are speakers from Bangkok. One of the subjects is a Thai phonetician, a visiting professor to the Department of Linguistics, and the other is a postgraduate student of the University of Edinburgh.

6.2.2. Linguistic Materials

---

[2] See Henderson (1964:418).

6.2.2. Linguistic Materials

As mentioned above, Thai has a three-way voicing contrast of stops, and has eight stop consonant phonemes in the consonant inventory. It is known that prevocalic stops /p, t, k/ are somewhat pharyngealized and tend to lower and centralize the following vowel.[3] The linguistic material in the experiment consists of 29 words in minimal and near minimal pairs, and the syllables are of the shape of CVV or CVN where C differs. VV represents the long low back unrounded /aa/ and the long high front vowel /ii/. The words in the experiment are written in Thai scripts by a Thai phonetician and were read five times, three times in isolation and two times in the carrier sentence: pu F k^ham M wə M_____ , "Say the word_____". The recording of the material was made in the sound-proof recording studio of the Phonetics Laboratory, University of Edinburgh. Three tokens (two items in isolation and one item in the carrier sentence) were recorded. Each subject was instructed to place a short pause before each test item in the carrier sentence. It does not appear that there is any significant difference in results between items in isolation and those in the carrier sentence. The material can be shown as follows:

---

[3] See Abramson (1962), p.4.

Table 6.1 List of Thai words

Five tones	(1) Mid	kʰaa M	"a kind of grass"
	(2) Low	kʰaa L	"galangal"
	(3) Falling	kʰaa F	"to kill"
	(4) High	kʰaa H	"to trade"
	(5) Rising	kʰaa R	"leg"

baa L	"shoulder"	paa L	"forest"	pʰaa L	"to cut into two"
dam M	"black"	tam M	"to pound"	tʰam M	"to do"
		kam M	"to hold in one's hand"	kʰam M	"word"
bin M	"to fly"	pii M	"year"	pʰii M	"fat
dii M	"good"	tii M	"to hit"	tʰii M	"time(as in one time)
		kii L	"how many"	kʰii L	"to ride"
bon M	"on"	pon M	"to mix"	pʰon M	"energy"
don M	"to inspire"	ton M	"the self"	tʰon M	"to endure"
		kon M	"trick"	kʰon M	"to stir"

6.2.3. Acoustic Analysis

The procedure of acoustic analysis is exactly the same as described in the general procedure in Chapter 1.

6.3. Results

6.3.1. Analysis of Waveforms

Some of the examples of waveforms for the three types of voicing contrasts are shown in Figure 6.1. We can see that the initial stops in Thai can be divided into several acoustic portions such as voicing lead, burst, and aspiration. In some tokens of voiceless unaspirated stops, the burst portions are not easily distinguishable from the aspiration portion following the burst. The waveforms clearly manifest the differences of the three voicing categories: the voiced stops show the voicing lead before release, while the voiceless aspirated stops show a considerable portion of aspiration between the oral release and the onset of voicing of the following vowel. It can be said that the voiceless unaspirated stops tend to be articulated with almost simultaneous oral and glottal closure and release, and the amplitude at the onset is quite high. On the other hand, the voiceless aspirated stops as observed in the recording of the materials are articulated more laxly, and it can be shown in the waveforms that there is a lengthy portion of breathy release and the rather gradual onset of the following vowel.

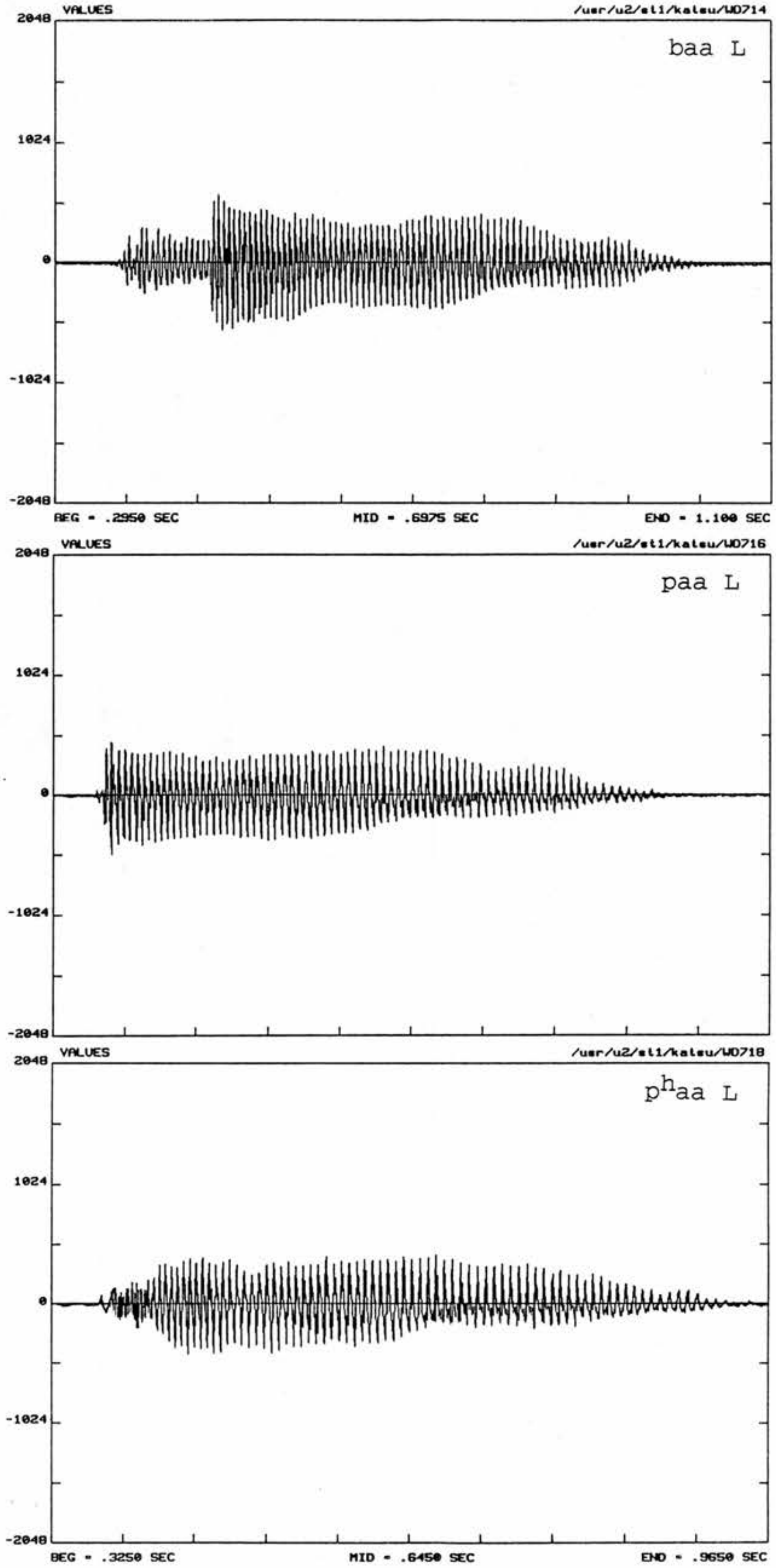


Figure 6.1 Examples of waveforms of three-way Thai stops

6.3.2. Voice Onset Time (V.O.T.)

The mean VOT value for each stop consonant phoneme is shown in Table 6.2.

Table 6.2 Mean VOT Values of Thai Stops (ms)
(N = 12 for bilabial, alveolar and velar)
(s.d. in parenthesis)

/b/	-104 (18.5)	/p/	5 (1.6)	/pʰ/	73 (22.2)
/d/	-106 (41.3)	/t/	8 (3.2)	/tʰ/	76 (13.2)
		/k/	23 (4.3)	/kʰ/	95 (4.7)
x	-105.2		12.1		81.4

The grouped mean values clearly show a trimodal distribution of VOT along a single timeline for three types of voicing categories. The differences between major voicing categories are statistically significant ($p < 0.01$); the two-tailed t-value is 6.16 for the difference between voiced and voiceless unaspirated stops, and the t-value is 11.0 for the one between voiceless unaspirated stops and voiceless aspirated stops. The range of voiced categories is rather extensive as shown in standard deviation, but there is no overlapping with other categories. The voiceless unaspirated stops are distributed in a narrow range of VOT.

As to the VOT values in relation to place of articulation, it can be seen that for voiceless unaspirated and aspirated categories the value is greater for velar stops than for bilabial or alveolar stops. For the voiced categories there is no difference between bilabial and alveolar stops. The tendency for velar stops to show a longer voicing delay is found in Thai too as in other languages under investigation (see p.27).

Furthermore, in order to examine the vowel effect on VOT, the data was analyzed as a function of the following vowel. Table 6.3 presents the mean value for each consonant followed by three vowels.

Table 6.3 Mean VOT value as a function of the following vowel (N = 6 for each consonant)

	/a/	/i/	/o/
/b/	-123	- 94	- 96
/d/	- 91	-121	-106
/p/	5	4	8
/t/	5	10	8
/k/	16	24	30
/p^h/	59	70	91
/t^h/	78	73	78
/k^h/	74	115	96

As shown in Table 6.3, it does not appear that there is a consistent relation between vowel height and VOT. For the voiceless unaspirated stops, there is a weak tendency for VOT to be somewhat longer before the vowel /o/ than before the vowels /a, i/. In the general sense, however, the vowel effect is less apparent and is not consistently manifested.

Furthermore, the VOT values in relation to the differences of tone types were examined. The mean VOT values of /k^{h}aa/ are presented in Table 6.4 for the five different tones.

Table 6.4 Mean VOT value as a function of tones (ms) (N = 6 for each tone type)

/k^{h}aa/	M	"a kind of grass"	75 ms
/k^{h}aa/	L	"galangal"	74
/k^{h}aa/	F	"to kill"	74
/k^{h}aa/	H	"to trade"	77
/k^{h}aa/	R	"leg"	84

As seen in Table 6.4, there does not seem to be any correlation between tone types and VOT of voiceless aspirated stops, except that Rising tone shows a rather longer VOT than others. Tone effect is considered to appear in the stretch of the syllable-final portions and there is no satisfactory explanation of a longer VOT for the rising tone, except to say that there might be some difference in the breath force between the five tones and the rising tone may require a greater breath force than the others.[4]

6.3.3. Fundamental Frequency (Fo) and the Contour

As mentioned in the previous chapter, the voicing contrasts of stops are closely related to pitch perturbations in the following vowels. There have been several experimental studies on tones (Fo perturbations) in Thai. Abramson (1962) examined the fundamental frequency contours of five tones; Hiranburana (1971) presented the data on the phonetic shapes of tones in fast and casual speech. Furthermore, Gandour (1974) presented the results of the acoustical investigation of the consonant types and tones. It was pointed out in his study that the Fo contour after voiceless obstruents is relatively high and falling and after voiced obstruents the Fo contour is relatively low and rising.

The present section presents the results of the measurements of Fo in the following vowel of the words in Table 6.1. First the measurements were made at the vowel onset immediately after oral release and the results are presented in Table 6.5.

Table 6.5 Mean Fo value at vowel onset (Hz)
(N = 12 for bilabial, alveolar, and velar)
(s.d. in parenthesis)

/b/	183 (16.1)	/p/	179 (14.5)	/p^h/	204 (14.1)
/d/	184 (14.6)	/t/	186 (8.9)	/t^h/	206 (15.7)
		/k/	192 (13.1)	/k^h/	205 (13.1)
X	183		186		205

From Table 6.5, it can be found that initial Fo value of voiceless aspirated stops is much higher than it is for voiced and voiceless unaspirated stops; the grouped mean Fo value for voiceless aspirated stops is 10 to 12 % higher than for the other two types of stops. The

---

[4] Henderson (1964) pointed out that the breath force must be increased in the pronunciation of the high tone, but did not mention about the force in the rising tone (p.420).

difference between voiceless unaspirated and aspirated stops is statistically significant (two-tailed t-value = 5.09 $p < 0.01$). But there is no marked difference in Fo value between voiced stops and voiceless unaspirated stops (two tailed t-value = 0.46, n.s.).

Next, the fundamental frequency contours were examined to see a change of Fo patterns. In examining the change of Fo contours, it can be said that the effect of tone types appears on the stretch of the latter half of the syllable and the initial stretch of the syllable seems to be less affected by the intrinsic form of the lexicon. The measurements of Fo were made at 20 ms intervals from vowel onset to 100 ms. The words examined are specified as mid tone and are listed below. The average Fo values were calculated at each interval and the frequency averaged curves on the following vowels are shown in Figure 6.2.

Table 6.6 A list of words examined for Fo contours

dam M "black"	dii M "good"	don M "to inspire"
tam M "to pound"	tii M "to hit"	ton M "the self"
tʰam M "to do"	tʰii M "time"	tʰon M "to endure"

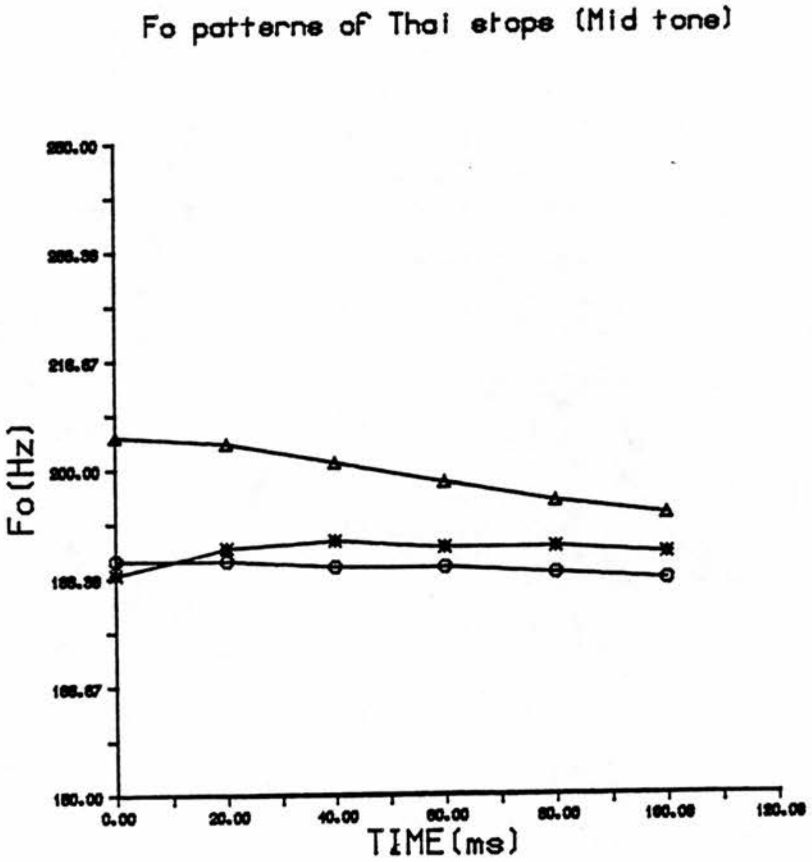


Figure 6.2 Averaged Fo contours of a three-way of Thai stops
* voiced ∘ vl.unaspirated ▵ vl.aspirated

It can be seen that there is a difference in the onset Fo and in the direction of change. Immediately after the voiceless aspirated stops, the contour starts in a higher region and shows a gradual falling. The voiced stops and the voiceless unaspirated stops begin at a similar Fo region, but show a difference in their change of direction. The voiced stops give a slight rise and then show a level pattern, while the voiceless unaspirated stops show a level pattern from the onset.

6.3.4. Spectral Analysis

The short-time power spectra were sampled in order to examine the spectral characteristics of voicing contrasts of Thai stops. A total of 16 tokens (one for each utterance for each subject for syllables followed by the vowel /a/) were sampled using a 25-ms time window beginning at the consonantal release. Because of the difference in the acoustic portions in waveforms mentioned above, the window portions actually sampled vary across the voicing categories. In the case of voiced stops and voiceless unaspirated stops, the 25-ms time window includes both burst and voicing onset of the following vowel, though the burst portion is not easily distinguishable from the voicing portion in some tokens. For the voiceless aspirated stops, the time window includes some portion of aspiration and does not extend to the onset of the voicing of the following vowel. Some of the examples of power spectra are shown in Figure 6.3.

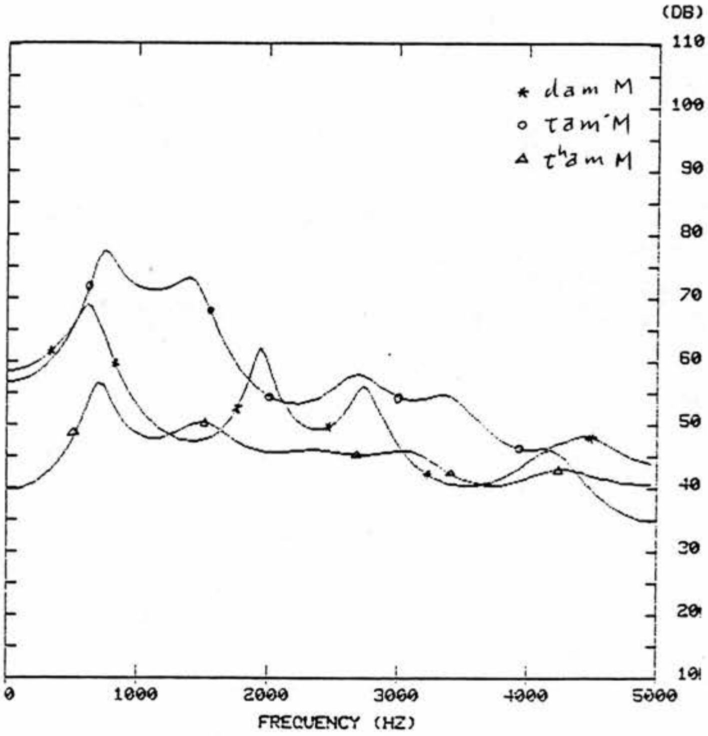


Figure 6.3 Examples of smoothed spectra of three-way Thai stops
* dam M ∘ tam M △ t^ham M

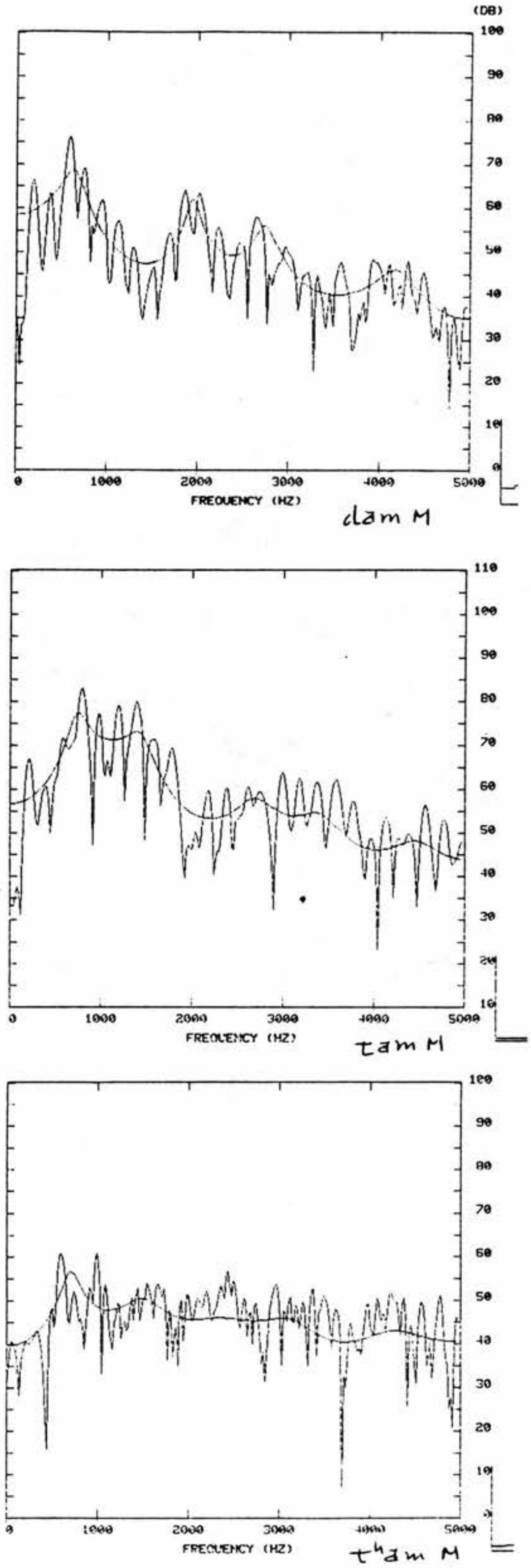


Figure 6.4 Examples of short-time power spectra sampled at the consonantal release for a three-way of Thai stops dam M (top) tam M (middle) t^ham M (Bottom)

Figure 6.3 shows that the intensity of voiceless unaspirated stops is greater than that of the other voicing categories, and that of voiceless aspirated stops is less than that of the others. This indicates that the initial air-flow of voiceless aspirated stops is rather weak, compared to other types of stops.

The examination of the individual power spectra in Figure 6.4 shows that the voiced stops and the voiceless unaspirated stops have regular energy distribution in the whole energy distribution and a slightly falling spectrum; on the other hand, the voiceless aspirated stops have rather irregular (aperiodic) energy presence and a level pattern of spectrum. Although the degree of regularity of energy distribution is somewhat difficult to quantify, it indicates the state of glottal pulsing, and the presence of regular energy distribution indicates the periodicity of glottal pulsing. Further, bandwidths of each peak are narrower in voiceless aspirated stops than other types of stops as far as bandwidths can be examined in these spectra.

6.3.5. The Onset Frequency of the First Formant

As in previous chapters, the measurements of F1 onset frequency were made on the words listed below. Table 6.7 summarizes the mean onset frequencies in Thai stops. Figure 6.5 shows some examples of formant patterns of Thai stops.

Table 6.7 Mean Onset F1 Frequency in Thai Stops
(N = 12 for bilabial, alveolar and velar)

baa L	739.3	paa L	806.0	p^{h}aa L	1189.3
dam M	613.0	tam M	807.5	t^{h}am M	1006.7
		kam M	665.5	k^{h}am M	965.0

As shown in Table 6.7, there is a difference in F1 onset frequency between the three types of stops, and the F1 onset value is higher in voiceless aspirated stops than in the other two types of stops. The differences in F1 onset frequency are what are expected from the difference in the onset time of periodicity. In voiceless aspirated stops, it is considered that the onset of periodicity occurs at some later time than other types of stops, as shown in VOT, and the presence of aspiration-related turbulent air-flow causes the delay of pulse excitation.

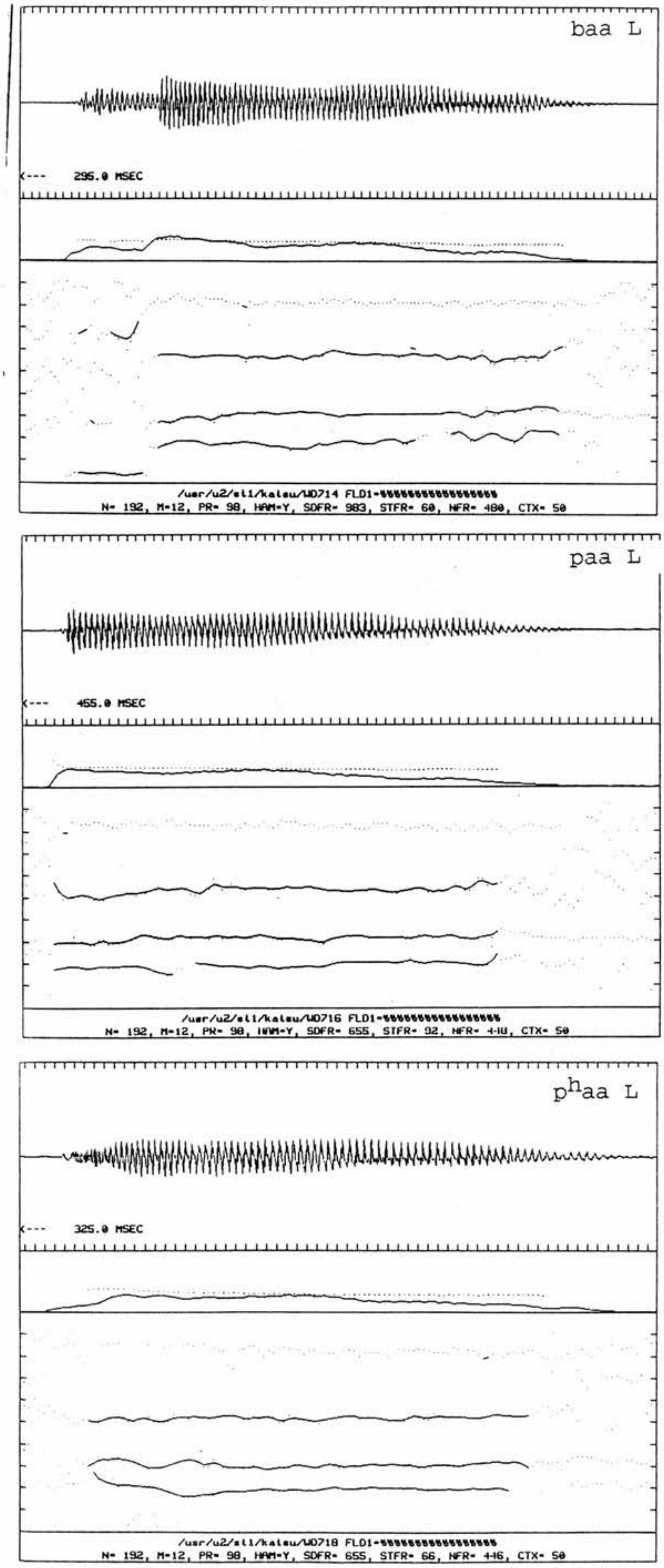


Figure 6.5 Examples of formant patterns of Thai stops

6.4. Discussion

Among several acoustic properties so far discussed, VOT appears to be a sufficient cue for the three-way voicing categories of stops in Thai, and the averaged VOT values more or less agree with previous studies (Lisker and Abramson, 1964). The mean values for the major categories are -105 ms for voiced stops, 12 ms for voiceless unaspirated stops, and 81 ms for voiceless aspirated stops. The values show a trimodal distribution for voicing lead, short lag, and long lag, and the ranges of voicing lead and long lag are rather extensive.

Although it may not be appropriate to compare the present VOT data to those from Lisker and Abramson (1964) due to different sampling and analytical procedures, an overall comparison can be made, and the averaged values for each consonant can be shown as follows:

	Present study	Lisker and Abramson (1964)
/b/	- 104	- 97
/d/	- 106	- 78
/p/	5	6
/t/	8	9
/k/	23	25
/p^h/	73	64
/t^h/	76	65
/k^h	95	100

Although there are some discrepancies between the two studies in the values, especially in the duration of voicing lead, the overall trends between the studies are consistent. Voiceless unaspirated stops /p, t/ show a very short lag time, and this indicates that they are articulated with almost simultaneous oral and glottal closure and release. The voiceless aspirated stops show a rather long delay of voicing, and are characterized by the breathy aspirated portion.

Furthermore, as shown in Table 6.4, VOT is not affected by the difference of tone types. Since the tone effect appears in the latter stretch of the syllable, the initial laryngeal timing is less affected. Henderson (1964) points out that there is a difference between the five tones in the degree of breath force, but the difference in breath force, if any, does not seem to have any effect on VOT.

The fundamental frequency values immediately after stop release appear to be an ambiguous cue for the three-way voicing categories, since the Fo values of voiced and voiceless unaspirated stops are almost in the same region. The mean Fo values at vowel onset are 183 Hz for voiced stops, 186 Hz for voiceless unaspirated stops, and 205 Hz for voiceless aspirated stops. The Fo value following the voiceless aspirated stops is higher than those of the other two types of stops. The Fo value immediately after the voiced stops shows the lowest one, but is quite close to that of voiceless unaspirated stops. The difference, as mentioned above, is not statistically significant (two-tailed t-value is 0.46, n.s.). The present data are not consistent with those from Gandour (1974). He found that the Fo value after a voiceless unaspirated stop is higher than that of voiceless aspirated stop. It is not clear at this stage why there exist differences between the two results. It can be said that voiceless aspirated stops are quite breathy, and a higher Fo value at vowel onset may be attributable to a greater volume of airflow.

To the finding that there was no marked difference in Fo values between voiced stops and voiceless unaspirated stops, we can speculate from historical changes of Thai. It is known in Thai (Nishida, 1987) that consonant types are closely related to the tone development. In many cases, lower tones tend to appear following voiced stops, and higher tones following voiceless stops. With the loss of voicing contrasts of initial consonants, intrinsic pitch perturbations tend to appear as tone types. In the process of tone development, there was a sound change from voiced stops to voiceless unaspirated stops and vice versa. Voiceless unaspirated stops which occur in present-day Thai might have been voiced stops in the earlier forms, and the opposite might have existed. This means that voiceless unaspirated stops, which were historically derived from voiced stops, may have similar characteristics and show a very similar Fo value to voiced stops, because of the historical development of tone.

As to a change of Fo contour, the Fo after the voiced stops shows a slight rise and this occurs in the initial stretch of the syllable regardless of the difference of tone types. The Fo contours after voiceless aspirated stops show a steady falling from a high region and the pattern is consistent to all types of tones. These results support the general hypothesis that voiced stops are associated with a rising pattern and voiceless stops with a falling pattern.

The results of spectral analysis show that the voiceless aspirated stops are clearly separated from the other two voicing categories; the voiceless aspirated stops show less intensity level and, energy distribution is less regular, which may be called aperiodic. However, the spectral differences between voiced stops and voiceless unaspirated stops are less apparent. Although there are some differences in such acoustic details as spectral peak and intensity level, they are not consistent in some tokens. As a weak trend, the voiceless unaspirated stops show a rather greater intensity and the voiced stops tend to have the first peak which is considered to be the first formant in the lower frequency region. This indicates that the initial articulatory state immediately after oral release in voiced stops and voiceless unaspirated stops is similar to each other and suggests that spectral characteristics are rather secondary for characterising the voicing categories of Thai stops.

The results of F1 onset frequencies appear to cue initial-stop voicing differences, and F1 onset frequencies are higher in voiceless aspirated stops than in other types of stops in Thai. The results generally agree with the ones in previous chapters which show that low F1 onset frequencies are associated with voiced stops and high F1 onset frequencies with voiceless ones.

Finally, as mentioned above, the voiceless unaspirated stops are referred to as having the tense quality (Henderson, 1964). Although there are a variety of acoustic and physiological correlates to the tense quality (e.g., the degree of muscle tension, width of pharynx, oral air pressure, larynx position, etc.), we may consider here, in addition to articulatory properties of voicing, that the tense quality of voiceless unaspirated stops in Thai may be associated with the sharp rise of intensity in the intensity curve in Figure 6.5. Unlike other the two types of stops, the voiceless unaspirated stops show a sharp rise of intensity at release.

6.5. Summary

To summarise, the examination of the acoustic properties of the three-way voicing contrasts in Thai reveals that voice onset time is an important factor in characterising the voicing categories. VOT shows a trimodal distribution (voicing lead, short lag, and long lag) along a single laryngeal timing dimension. The voiceless aspirated stops show different features in onset Fo, the contour, and spectral intensity level from other voicing categories. They show a higher onset Fo value, gradual falling pattern and lesser degree of intensity than the other categories. The voiceless unaspirated stops show rather similar characteristics to voiced stops in some aspects, but differ in the direction of Fo contour and intensity level. The voiced stops give a rise in the Fo contour regardless of types. The voiceless unaspirated stops show a very short VOT value and a higher intensity at vowel onset, and show a tense quality as seen in an abrupt rise of intensity at the vowel onset.

Chapter 7

Stops in Hindi

7.1. Introduction

Hindi, as in other Indo-Aryan languages, has a four-way phonological contrast of stop consonants in four points of articulation. The contrast involves both voicing and aspiration, and the stops are usually divided into four major categories: voiced unaspirated stops, breathy voiced stops (traditionally called voiced aspirated or murmured voice), voiceless unaspirated stops, and voiceless aspirated stops. The term for the second category (breathy voiced) is still problematic in Hindi phonetics, and we will here use the term "breathy voiced" following Schiefer (1988).[1]

There have been a number of studies which attempt to examine the phonetic characteristics of these voicing categories, especially those of the breathy voiced stops, and the studies have been made from three major areas in phonetics: acoustic, physiological, and aerodynamic areas. Acoustic studies have examined characteristics such as voice onset time (VOT), fundamental frequency (Fo), and intensity (dB). It has been shown that VOT serves to separate the three major categories of voiced, voiceless unaspirated, and voiceless aspirated stops, but does not separate the two types of voiced categories from each other (Lisker and Abramson, 1964; Poon and Mateer, 1985; Schiefer, 1986). Furthermore, it is understood in recent studies that Fo onset, the trajectory and the intensity of the formants play an important role for distinction between breathy voice and normal voice

---

[1] There has been a controversy over the term for the second category. Such terms as "voiced aspirated", "breathy voiced", "murmured", "murmured aspirated", "whispery voiced" and "voiced phonoaspirated" have been suggested (Benguerel and Bhatia, 1980). Ladefoged (1975) criticized the use of "voiced aspirated" since "they" (voiced aspirated stops) are neither voiced (in the sense of having regular vibrations of the vocal cords) nor aspirated (in the sense of having a period of voicelessness during and after the release of the closure)..."(p.127). For further discussions, see Benguerel and Bhatia (1980), Ladefoged (1975) and Schiefer (1988).

in Hindi stops (Schiefer & Kotten, 1983; Schiefer, 1986). In spite of all these experimental studies, there have been still arguments on the phonetic nature of the breathy voiced stops.

Physiological studies have examined the activities of the intrinsic laryngeal muscles such as the cricothyroid (CT), and it was found in EMG studies by Dixit (1975) and Dixit and MacNeilage (1980) that there is a high level of CT activity for the voiceless plosives but a suppressed activity for the voiced plosives. Investigations on glottal width were also made by Kagaya and Hirose (1975) and Benguerel and Bhatia (1980). It is reported in these studies that there are differences in glottal width; the glottis is maximally open for the voiceless aspirated stops, while it is about a half of that for the voiceless unaspirated stops. Dixit (1987) examined the timing relations of glottal opening-closing using photo-electric glottographic (PEG) techniques and reports that the timing of glottal opening-closing gestures in relation to supralaryngeal configurations is different for the major voicing categories in Hindi.

Furthermore, it is suggested by Chomsky and Halle (1968) that the voiced aspirated stops are produced with heightened subglottal pressure (Ps) and heightened subglottal pressure is one of the main features for the distinction between the two types of voiced stops. They also assume that heightened subglottal pressure is an independent variable and is not a function of laryngeal adjustment. These remarks are mainly based on the studies by other researchers on Hindi and Korean stops (Lisker and Abramson, 1964; Kim, 1965). They mention:

> "It must further be noted that heightened subglottal pressure may be used in the production of a speech sound without involving tenseness (in the supraglottal musculature). This is the situation in the aspirated voiced stops of languages such as Hindi, where, according to Lisker and Abramson (1964), voicing commonly occurs during the period of oral closure. ...We shall say, therefore, that voiced aspirated stops of Hindi are produced without tenseness but with heightened subglottal pressure." (Chomsky and Halle, 1968:326)

Ohala and Ohala (1972) and Dixit and Shipp (1985) examined these claims concerning subglottal pressure variations in these stops and reported that there is no systematic correlation between subglottal pressure and voicing and/or aspiration contrasts, but there is a momentary drop in Ps after the release of aspirated stops.

Through these studies, it can be said that the four types of Hindi stop differ in laryngeal gestures, muscle activities, the relative timing of glottal gestures to oral release, and transglottal pressure, though the details of specific aspects remain to be clarified. Especially, it is assumed that the production of the two types of voiced stops differs in the laryngeal gestures, the width and the relative timing of glottal dynamics, and these differences in production are reflected in the differences of the acoustic properties such as voice onset time (VOT), fundamental frequency (Fo), intensity (dB), and F1 onset frequency. The present study is a further attempt to provide and to examine detailed acoustic data for the four types of stops in Hindi and to examine the underlying mechanisms in the articulation of these stops.

7.2. Experimental Procedure

7.2.1. Subjects

The subjects of the present study are three native speakers of Hindi, all male, and are speakers of standard Hindi. Two of them are lecturers in the university, and the other one is a postgraduate student in the University of Edinburgh. Their ages range from 33 to 47 years and all of them reside in Edinburgh.

7.2.2. Linguistic Materials

A list of 16 minimal or near minimal pair words was prepared as shown below. These words were written in Hindi script by one of the Hindi speakers, and the word list consists of two sets of words: one is written in random order and the other is written in minimal pairs. Each word was read in isolation three times for recording in a natural way by the three speakers. The recording was made in the sound-proof recording studio in the Phonetics Laboratory, University of Edinburgh.

Table 7.1 A list of Hindi words

	Voiced stops	Breathy voiced	Voiceless unaspirated	Voiceless aspirated
Bilabial	**[bal]** **"hair"**	**[b̤ʱal]** **"forehead"**	**[pal]** **"Proper name"**	**[pʰal]** **"knife blade"**
Dental	**[d̪al]** **"lentil"**	**[d̪̤ʱal]** **"knife"**	**[t̪al]** **"beat"**	**[t̪ʰal]** **"plate"**
Retroflex	**[ɖal]** **"branch"**	**[ɖ̤ʱal]** **"shield"**	**[ʈal]** **"postpone"**	**[ʈʰal]** **"wood shop"**
Velar	**[gan]** **"song"**	**[g̤ʱan]** **"house"**	**[kan]** **"ear"**	**[kʰan]** **"mine"**

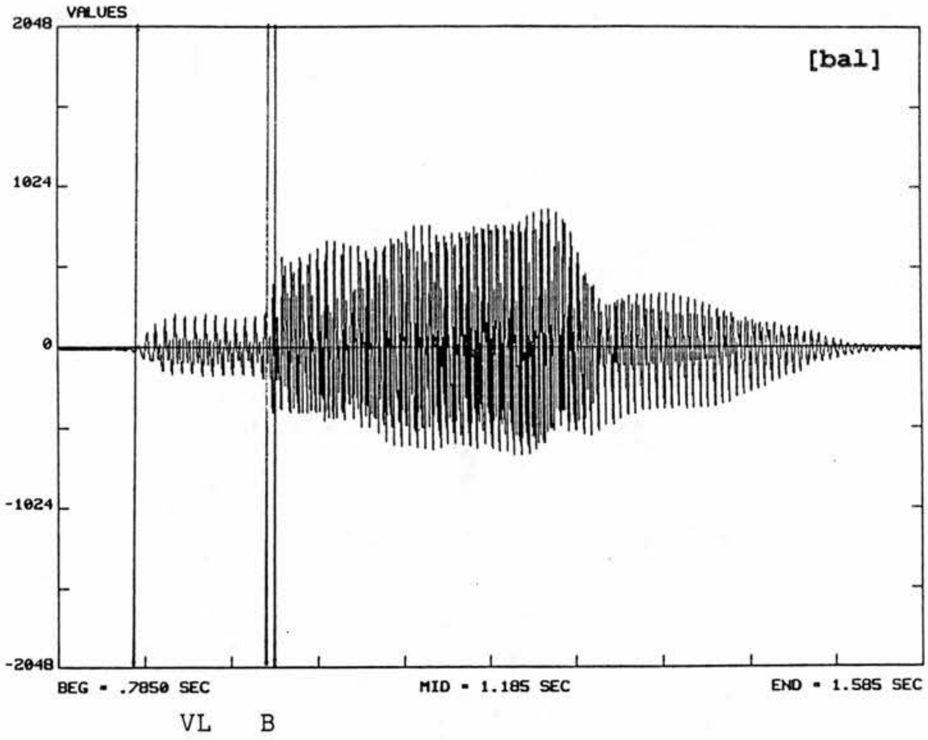
7.2.3. Acoustic Analysis

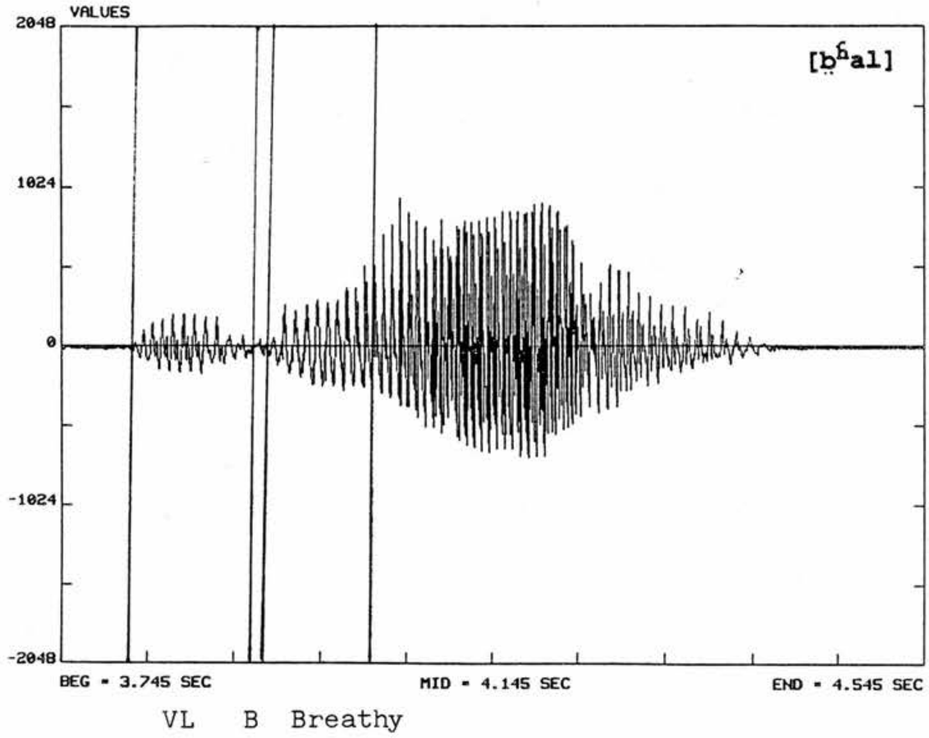
The procedure of acoustic analysis is exactly the same as described in the general experimental procedure in Chapter 1.

7.3. Results

7.3.1. Acoustic Portions in Waveforms

Word-initial stops in Hindi can be divided into several acoustic portions such as voicing lead, burst (or release of stop), and aspiration. Examples of waveforms divided into these acoustic portions are shown in Figure 7.1. Segmentation of waveforms into each acoustic portion was made by manually controlling the cursors on the waveforms. The beginning of the burst shows up as the sudden increase in the amplitude of the waveform, and breathy portion shows up as the sudden change of amplitude in the waveforms. In some utterances, the burst portions are not easily distinguishable from the aspiration portion following the burst. The average duration values for major categories are shown in Figure 7.2. Although it is pointed out in Schiefer (1989) that there is a subject variation in the duration of voicing lead for the voiced stops, all the subjects in the present experiment showed a voicing lead in the voiced stops and their variations are rather consistent with the group trend.





VL = voicing lead   B = burst

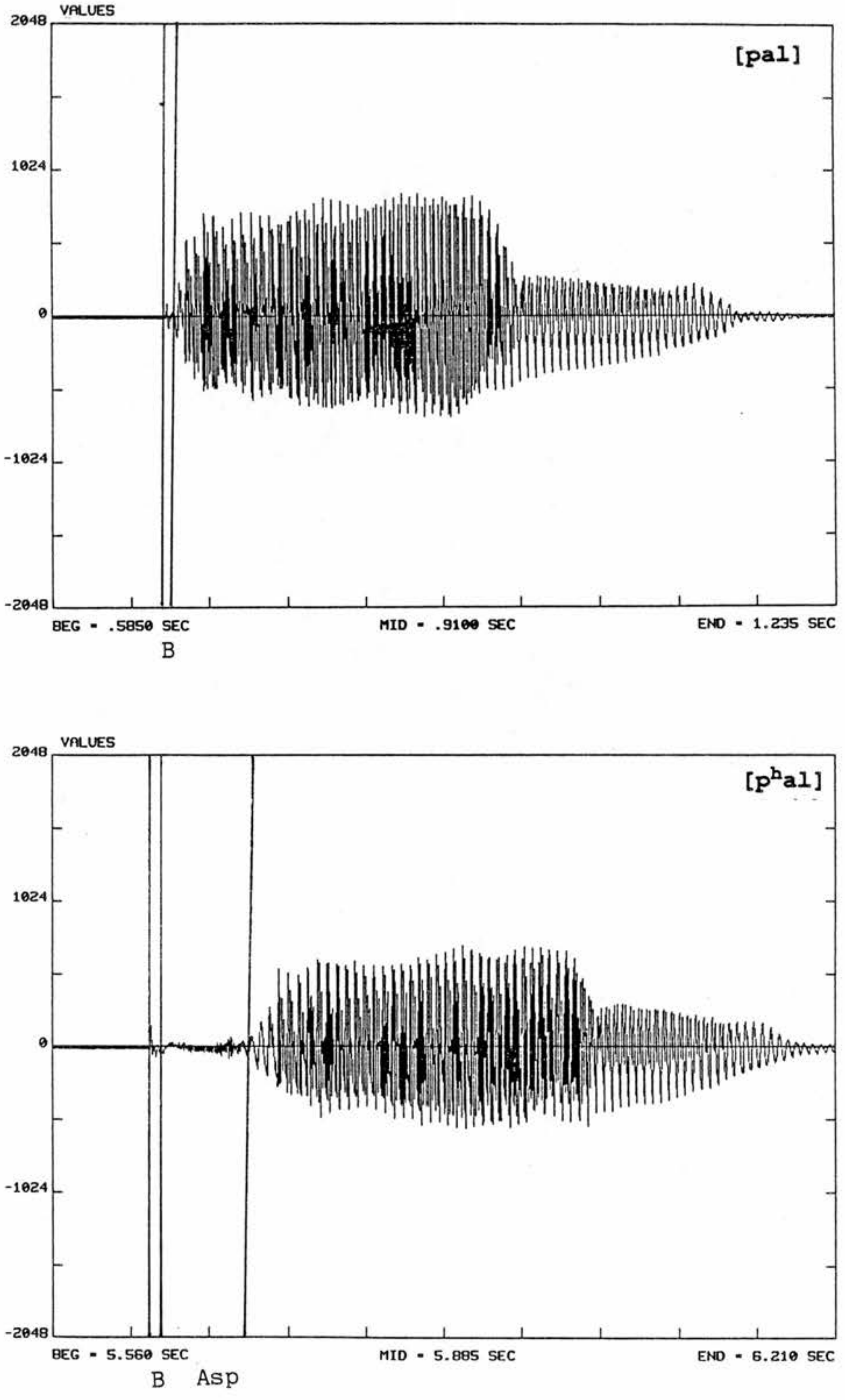


Figure 7.1 Examples of waveforms of four-way Hindi stops
VL = voicing lead, B = burst, Asp = aspiration

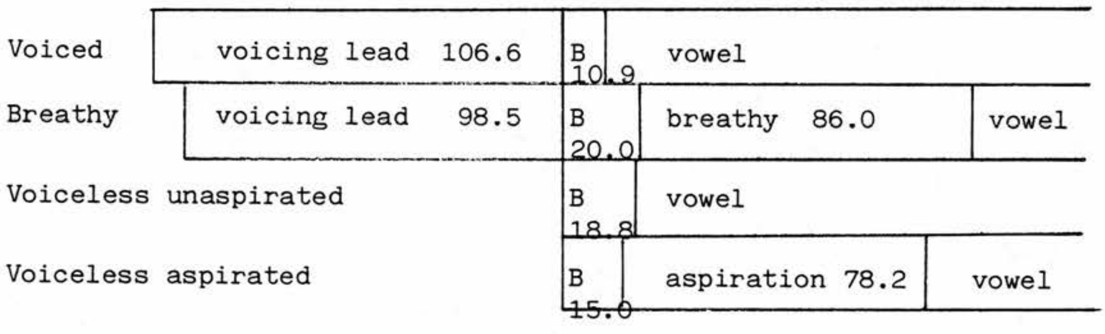


Figure 7.2 Average duration of acoustic portions of Hindi stops (ms)
(N = 36 for each acoustic portion)
B = burst

Table 7.2 Mean Duration of Breathy Portion (ms)
(N = 9 for breathy stops)

b̤ʱ	**69**	**(8.5) ms**
d̪̤ʱ	**103**	**(25.2)**
ɖ̤ʱ	**78**	**(21.4)**
g̤ʱ	**90**	**(32.8)**
X	**86.0**	

It is noted that the breathy voiced stops are divided into three acoustic portions, as shown above, and the breathy portion, traditionally called voiced aspiration, is considered to be the period in which vocal folds are vibrating without complete closure and to be significant for characterizing the stops. The mean values of breathy portion for each phoneme are presented in Table 7.2. As seen in the table, the breathy portion of the dental stop is longer than that of the others, but it does not appear that the difference in point of articulation gives an effect on the duration of the breathy portion.

7.3.2. Voice Onset Time (VOT)

As in other previous chapters, the measurements of VOT were made for each utterance in the recorded materials. Lisker and Abramson (1964) classified the voicing categories into three types according to the VOT; and VOT is an effective measure for distinguishing these voicing classifications. As mentioned above, however, it does not serve to separate the two voiced categories from each other in Hindi and other four-category type languages. Similar observations were made in Schiefer and Kotten (1983:518). Based on the measurements of acoustic portions, the average VOT values for each phoneme and the grouped mean for the major categories are shown in Table 7.3. The duration of voicing lead corresponds to VOT in the voiced stops, while the one of burst and aspiration corresponds to VOT in voiceless stops.

Table 7.3 Mean VOT values for Hindi stops (ms)
(N=9 for bilabial, dental, retroflex, and velar)
(s.d. in parenthesis)

	Voiced stops		Breathy stops		Voiceless unaspirated stops		Voiceless aspirated stops
/b/	-93(39.6)	/b̤ʱ/	- 93(28.9)	/p/	12(5.7)	/pʰ/	75(21.8)
/d̪/	-115(23.4)	/d̪̤ʱ/	- 90(21.8)	/t̪/	12(2.7)	/t̪ʰ/	82(16.4)
/ɖ/	- 98(35.1)	/ɖ̤ʱ/	-107(37.1)	/ʈ/	9(6.5)	/ʈʰ/	97(10.4)
/g/	-115(26.2)	/g̤ʱ/	-104(31.9)	/k/	34(8.2)	/kʰ/	119(6.5)
X	-106.6		- 98.5		18.8		91.3

As can be seen in Table 7.3, there is a clear-cut difference in the grouped mean of VOT between the three voicing categories, but there is no major distinction between the two types of voiced categories. The ranges of the two voiced categories are extensive and both categories overlap with each other. T-tests were applied to the major voicing categories, and the difference is significant for the two voiceless categories (two-tailed t-value = 11.6, p <.01), but is not significant for the two voiced categories (two-tailed t-value = 1.70, n.s.). These results are in agreement with the previous studies (Benguerel and Bhatia, 1980).

As to the VOT values in relation to place of articulation, as found in other languages under investigation, it can be seen that for voiceless categories the value is greater for velar stops than for bilabial or dental stops (see p.27). For voiced categories, it appears that retroflex and velar stops show a longer voicing lead than bilabial or dental stops.

7.3.3. Fundamental Frequency (Fo) and its Contours

It is generally understood that fundamental frequency is closely related to the state of the glottis and transglottal airflow conditions. In the present study, the measurements of the Fo were made at the breathy portion for breathy voiced stops, vowel onset and the steady-state portions of following vowel. Tables 7.4 (a - c) present the mean Fo value (Hz) of each stop.

Table 7.4a Mean Fo at vowel onset (Hz)
(N=9 for bilabial, dental, retroflex and velar)
(s.d. in parenthesis)

/b/	112.6(16.4)	/b̤ʱ/	110.3(19.9)	/p/	126.6(25.4)	/pʰ/	115.5(29.0)
/d̪/	113.0(17.6)	/d̪̤ʱ/	114.3(30.6)	/t̪/	130.7(20.0)	/t̪ʰ/	125.7(32.3)
/ɖ/	115.0(22.0)	/ɖ̤ʱ/	95.0(7.0)	/ʈ/	130.3(27.3)	/ʈʰ/	117.5(33.2)
/g/	122.3(22.1)	/g̤ʱ/	116.0(11.0)	/k/	134.0(28.7)	/kʰ/	130.6(29.0)
X	115.7		108.9		130.4		122.3

Table 7.4b Mean Fo at breathy portion (Hz)
(N=9 for breathy stops)
(s.d. in parenthesis)

/b̤ʱ/	104.0(9.5)
/d̪̤ʱ/	118.7(34.6)
/ɖ̤ʱ/	101.3(13.0)
/g̤ʱ/	115.0(9.5)
X	109.8

Table 7.4c Mean Fo at vowel steady state portion (Hz)
(N=9 for bilabial, dental, retroflex and velar)
(s.d. in parenthesis)

/b/	122.3(22.0)	/b̤ʱ/	121.1(16.3)	/p/	123.3(22.0)	/pʰ/	115.0(21.2)
/d̪/	121.0(23.0)	/d̪̤ʱ/	116.3(19.6)	/t̪/	122.3(15.6)	/t̪ʰ/	127.3(24.0)
/ɖ/	120.0(23.3)	/ɖ̤ʱ/	124.0(22.3)	/ʈ/	125.0(25.7)	/ʈʰ/	119.0(26.9)
/g/	127.7(14.6)	/g̤ʱ/	133.3(18.6)	/k/	130.3(22.5)	/kʰ/	133.3(19.0)
X	122.8		123.7		125.2		123.7

From Table 7.4a, it can be seen that there is a clear difference between voiced and voiceless categories in the Fo values. The grouped means of the Fo at the vowel onset following voiceless stops are higher than those following voiced stops. It can also be seen that the Fo value of the vowel onset following the breathy voiced stops is the lowest. The difference between the voiced stops and breathy voiced stops is not significant (two-tailed t-value = 2.76, n.s.), and that between voiceless aspirated and unaspirated stops is not significant either (two tailed t-value = 2.24, n.s.). The Fo values in relation to the place of articulation indicate that velar stops show the highest value in all voicing categories. Table 7.4b presents the mean Fo for the breathy portion of the breathy voiced stops. It can be noticed that the Fo of the breathy portion is markedly low and the grouped mean is almost the same with that at the vowel onset. Table 7.4c shows the mean Fo at vowel steady state portions. It can be found that there is no major difference between the major voicing categories. This indicates that influence of the preceding consonants on Fo has weakened or disappeared at the vowel steady state portions, in the period of 60 - 80 ms from the consonantal release.

In connection with the Fo at vowel onset and vowel steady state portions, the Fo was measured at 20 ms intervals from vowel onset to 100 ms to examine a change of direction of the Fo pattern. For the voiced categories, the measurements were made at 20 ms intervals before and immediately after the vowel onset. The average Fo values were calculated at each interval point and the frequency-averaged curves for the following vowel are shown in Figure 7.3. The onset point on the abscissa shows the onset of voicing and is not the same timing when the oral release occurs. The onset of voicing usually begins, as shown in the results of VOT, some time later after the release for voiceless categories.

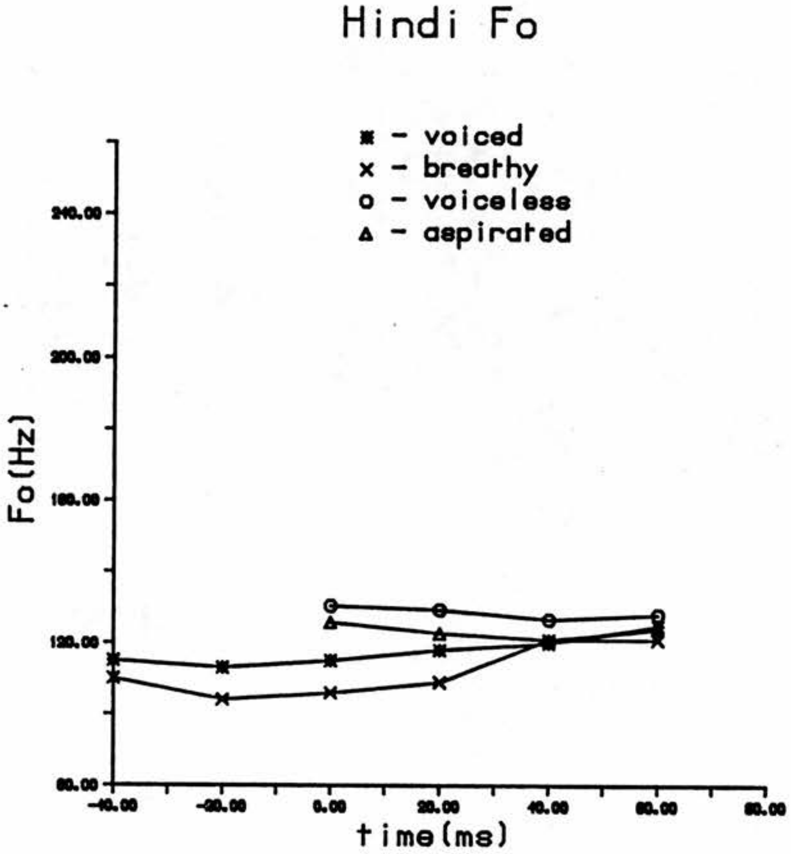


Figure 7.3 Fo curves of a four-way distinction of Hindi stops

From Figure 7.3, it can be seen that there is a difference in the Fo at vowel onset and in the direction of change. The Fo of voiceless categories begins in a relatively higher region, while the Fo of voiced categories begins in a lower region. Furthermore, it can be noticed that the voiceless categories show a steady falling pattern for the period of about 60 ms after vowel onset, while the voiced categories show a rising pattern in the same period after vowel onset. For the voiced categories, the vocal folds begin to vibrate before the release, and during this oral closure the Fo curve shows the falling pattern as the rate of vibration is presumed to be suppressed by the decrease in the transglottal pressure difference. Specifically, it can be noticed that the breathy voiced stops show a markedly low frequency and show a more distinctive fall-rise pattern. It can be said, therefore, that although the voiced categories show a similar effect on the Fo the breathy voiced stops will show a greater and more pronounced effect. The reason for this greater effect may be attributable to the loose vocal-fold vibration in the state of moderately open glottis; that is, vocal folds are vibrating without complete closure. As a final point it can be seen from Figure 7.3, voiceless aspirated stops show a considerable lower Fo than voiceless unaspirated stops. Although it is often mentioned that aspiration gives rise to a higher Fo, the Hindi data is not in the direction generally predicted.

7.3.4. Spectral Analysis

Spectral analysis was made by sampling power spectra for the initial period immediately after the stop release. A total of 36 power spectra (one for each word for three subjects excluding the words for retroflex) was sampled by ILS linear prediction for the initial time window period of 30 ms. This time window period is considered to contain relevant cues on place of articulation and voice quality. For the voiced unaspirated and voiceless unaspirated stops, this includes the burst and some breathy portion. For the breathy voiced stops and voiceless aspirated stops, this includes the burst and some portion of aspiration, but does not extend to the voicing. Some of the samples of power spectra are shown in Figures 7.4(a - b).

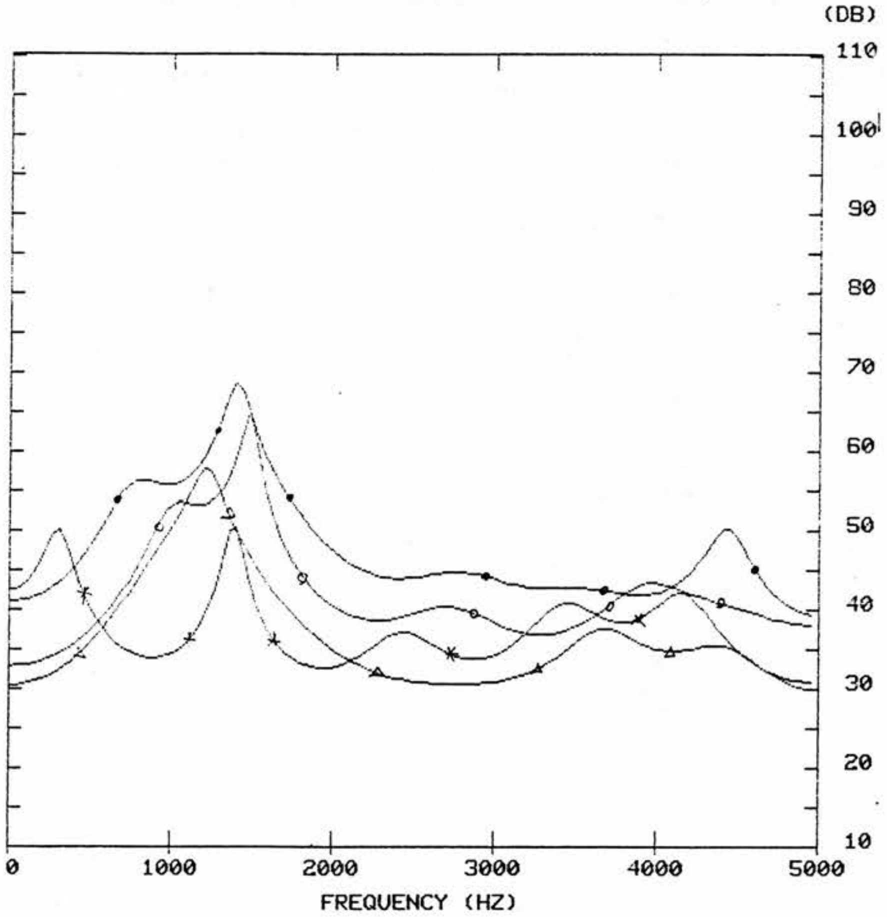


Figure 7.4a Smoothed power spectra of the four-way Hindi stops (30 ms)
* [gan] • [g̈ʱan] ∘ [kan] ▲ [k^han]

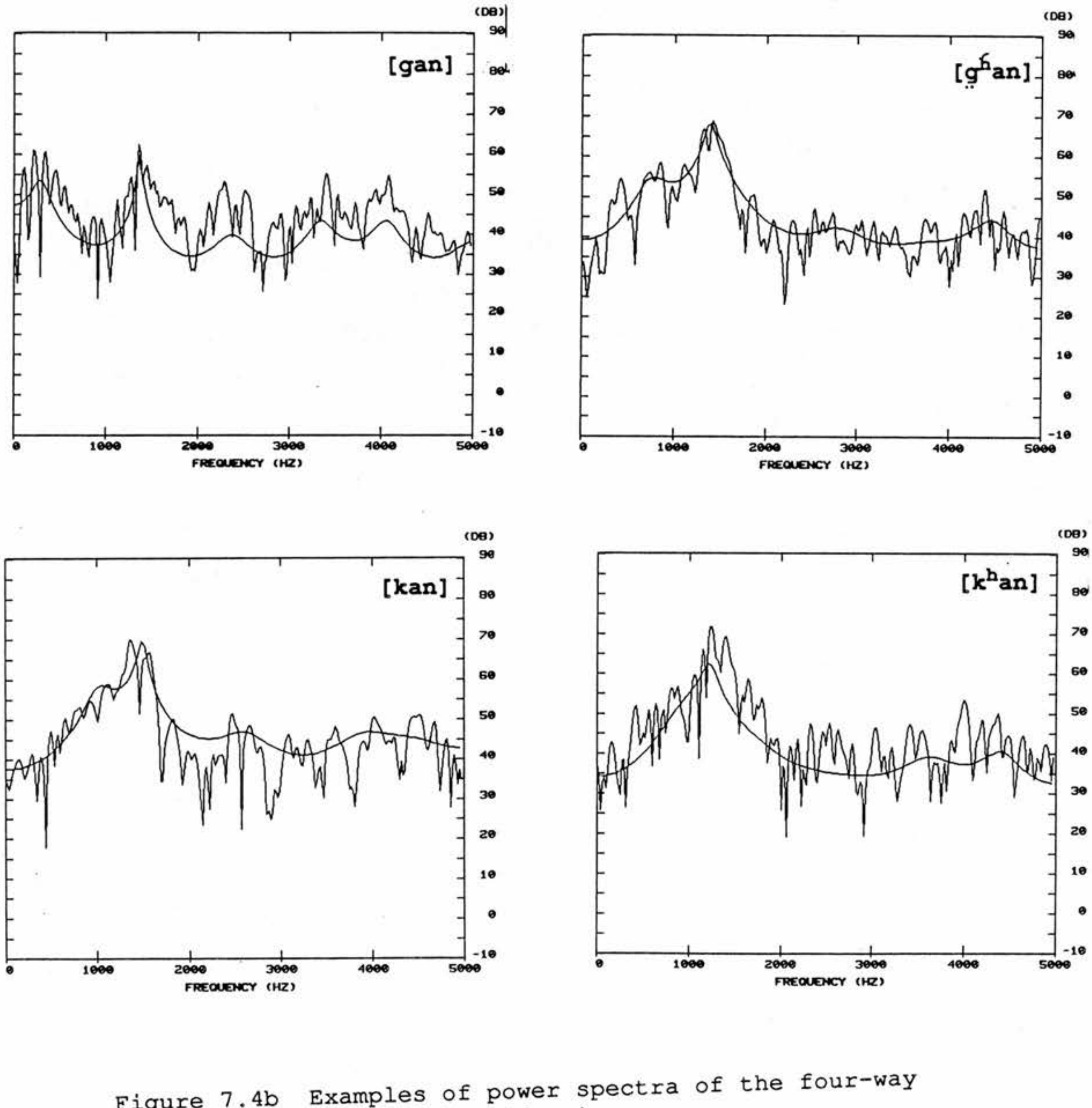


Figure 7.4b Examples of power spectra of the four-way Hindi stops (30 ms)

The examination of the overall intensity level shown in the smoothed slope of the power spectra indicates that the intensity level of voiceless unaspirated stops is as a weak trend somewhat greater than that of voiceless aspirated stops. For the two voiced categories, the intensity level of breathy voiced stops is considerably higher than that of voiced unaspirated stops. The articulatory reason for the greater intensity level may be attributed to the higher airflow rate, especially for the breathy voiced stops where the vocal folds are vibrating without complete closure, resulting in a higher airflow rate.

Furthermore, it can be seen in these figures that voiceless categories show a more irregular energy presence in the lower frequency region, while the breathy voiced stops have more irregularity in the higher frequency region. The regularity of energy distribution is not easy to quantify since it is influenced by several factors such as glottal pulsing and airflow rate, but it indicates differences of voice quality. As to the presence of irregular energy in the breathy voice, it can be considered that the vibration of the vocal folds with a slightly open status causes more turbulence in the airflow, resulting in a less regular and aperiodic energy presence in the higher frequency region.

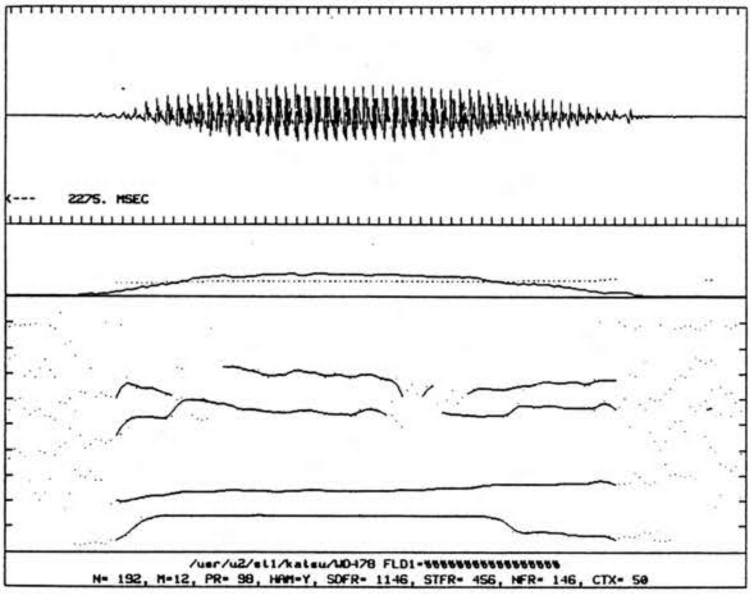
7.3.5. The Onset Frequency of the First Formant

The measurements of F1 onset frequency were made for the words listed below. Table 7.5 summarizes the mean F1 onset frequencies of four types of Hindi stops, and Figure 7.5 shows some examples of formant patterns of Hindi stops.

Table 7.5 Mean Onset F1 Frequency in Hindi Stops (Hz)
(N=9 for bilabial, dental, retroflex and velar)

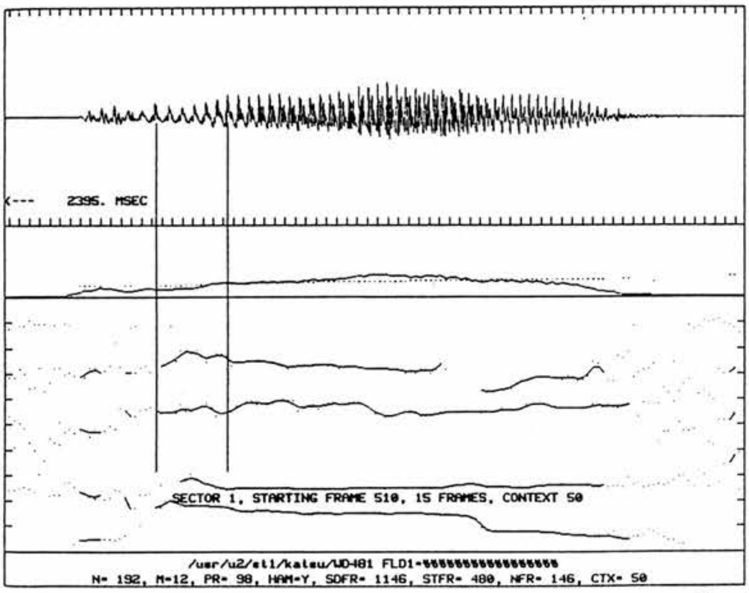
bal	374.7	b̤ʱal	857.0	pal	585.7	pʰal	847.0
d̪al	334.3	d̪̤ʱal	796.7	t̪al	485.3	t̪ʰal	878.3
ɖal	326.7	ɖ̤ʱal	891.5	ʈal	363.0	ʈʰal	720.0
gan	275.3	g̤ʱan	852.0	kan	408.7	kʰan	840.3
X	327.8		819.3		489.9		819.0

As shown above, F1 onset frequencies are higher in breathy stops and voiceless aspirated stops than in voiced stops and voiceless unaspirated stops, and both types of stops show exactly the same mean frequencies. Since the onset of voicing in breathy and voiceless aspirated stops occurs late and does not contain an F1 transition, F1 onset frequencies of these stops are higher than in voiced and voiceless unaspirated stops. It is apparent from this that high F1 onset frequencies at vowel onset are associated with breathy and voiceless aspirated stops, and low F1 onset frequencies with voiced stops. Therefore, F1 onset frequencies provide a cue for differentiating three major voicing categories of voiced stops, voiceless unaspirated stops, and breathy and voiceless aspirated stops.



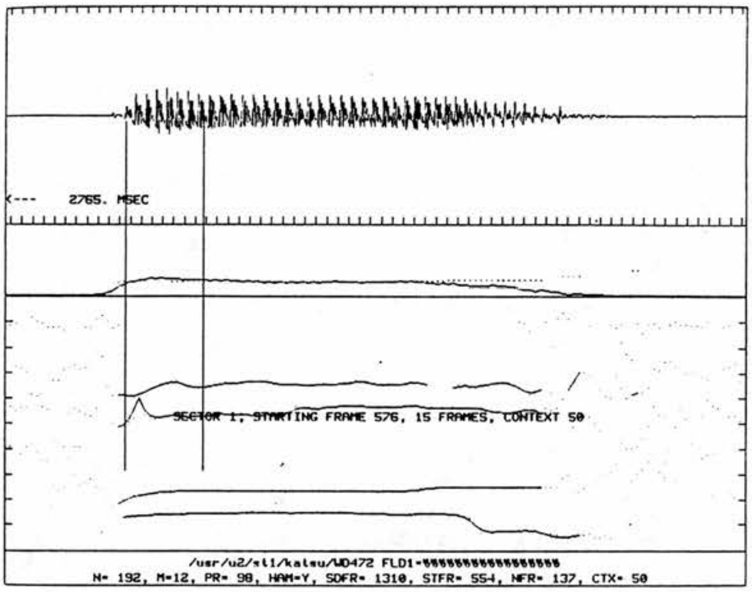
<---   2275. MSEC
/usr/u2/sl1/kalsu/WD478 FLD1-%%%%%%%%%%%%%%%%%%
N= 192, M=12, PR= 98, HAM=Y, SDFR= 1146, STFR= 456, NFR= 146, CTX= 50

[bal]

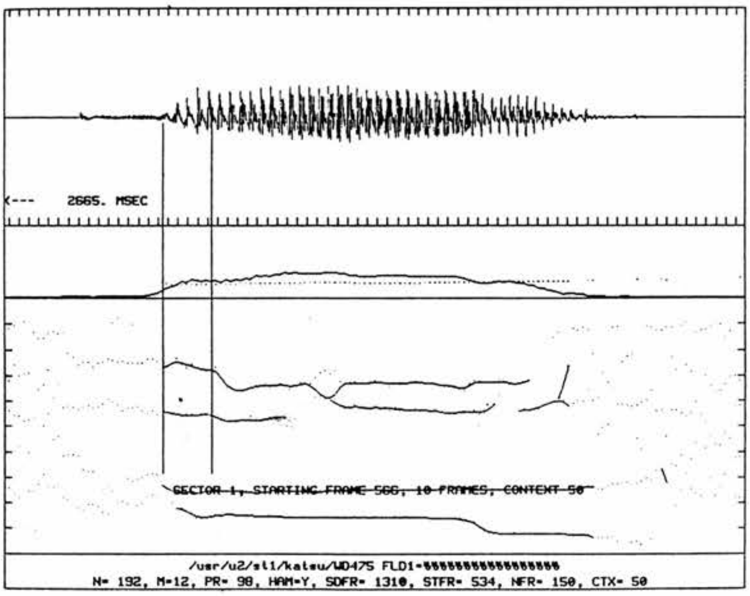


<---   2395. MSEC
SECTOR 1, STARTING FRAME 510, 15 FRAMES, CONTEXT 50
/usr/u2/sl1/kalsu/WD481 FLD1-%%%%%%%%%%%%%%%%%%
N= 192, M=12, PR= 98, HAM=Y, SDFR= 1146, STFR= 480, NFR= 146, CTX= 50

[b̤ʱal]



[pal]



[p^{h}al]

Figure 7.5 Examples of formant patterns of Hindi stops

7.4. Discussion and Summary

On the basis of the acoustic data which are summarized in Tables 7.1 - 7.5 and Figures 7.1 - 7.5, the acoustic properties of each major category in Hindi stops can be represented as follows:

1) Voiced unaspirated stop

The voiced unaspirated stops show a long voicing lead (x =-106.6 ms) and its range is rather extensive. The Fo at vowel onset after the release begins in a lower range (x=115.7 Hz) and shows a steady rise in the following vowel. The intensity level is lower than that of the voiced aspirated stops. The F1 onset frequencies are lower than those of other types of stops.

2) Breathy voiced stop

The breathy voiced stops show a markedly long voicing lead (x=-98.5 ms) and have a wider range of distribution of VOT. As shown in Fo trajectory, during the closure the Fo goes down sharply and the onset Fo frequency is the lowest among the four categories of stop. The Fo of the breathy portion is very low and steadily rises after the release. The Fo measurements in the 20 ms interval before and after the release show a distinctive fall - rise pattern among the four types of categories. The intensity level is generally higher than that of the voiced unaspirated stops. The F1 onset frequencies are higher in breathy stops than in voiced stops and voiceless unaspirated stops, and the values are as high as voiceless aspirated stops.

3) Voiceless unaspirated stop

The voiceless unaspirated stops show the shortest voicing lag (x= 18.8 ms) and show a very limited range of variation in VOT among the four types of stops. The Fo at vowel onset begins in a rather higher range (x=130.4 Hz) and shows a gradual fall to the steady state portion of the following vowel. The intensity level is slightly higher than that of the voiceless aspirated stops. The F1 onset frequencies are lower than those of voiceless aspirated stops, but higher than those of voiced unaspirated stops.

4) Voiceless aspirated stop

The voiceless aspirated stops show a long voicing delay (x= 91.3 ms) and have a rather limited range of variation. The Fo at vowel onset begins at a higher range, though slightly lower than that of voiceless unaspirated stops, and shows a gradual falling pattern. The intensity level is somewhat lower than that of voiceless unaspirated stops. F1 onset frequency is higher than those of voiced stops and voiceless unaspirated stops.

As shown in these acoustic characteristics, a single acoustic dimension does not necessarily distinguish the four-voicing categories of stops, and several acoustic dimensions are necessary to characterize the properties of the categories. One of the controversial issues in Hindi stop consonants is how to characterize the two voiced categories, especially the breathy voiced stops. VOT which represents the timing relation of voicing to the supralaryngeal configuration serves to differentiate the three major categories, as mentioned before. Although coordinated timing of voicing and supralaryngeal configuration is important in the production of these voicing categories, it is not functional in separating the two voiced categories from each other in Hindi.

As shown in Figure 7.1, breathy voiced stops possess the two timing events, one is voicing lead before release and the other is voicing lag after the release and the portion of voicing lag includes the burst and breathy portion. As mentioned earlier, it is known that the voicing lead can be missing in some cases. In connection with the voicing lead, Schiefer (1989) carried out the perception test of voicing categories by manipulating the durations of voicing lead and breathy portion in breathy voiced stops; i.e., how Hindi native speakers identify the manipulated stimuli of natural speech sounds. She concludes that the voicing lead is less important in the perception of breathy voiced stops and trading relations exist between the duration of voicing lead and that of the breathy portion. This implies that the identification of breathy voiced stops does not solely rely on the existence of voicing lead, and the existence of certain length of breathy portion is important. Although it is understood that the two voiced categories in Hindi stops show the similar length of the voicing lead and are not distinguished by VOT, the voicing lead itself is not mainly functional for the perceptibility of the breathy voiced stops.

Furthermore, one of the findings in the present study is the Fo curves related to the four types of stops. It is shown that the two voiced categories show a rather similar fall-rise pattern, but the breathy voiced stops represent the lowest Fo at the vowel onset and during the breathy portion, and show a distinctive and pronounced fall-rise pattern. The Fo at the onset shows the initial state of glottal adjustment and will be significant for characterizing the voicing categories which involve a change of initial glottal gesture. The physiological explanation for the markedly low Fo in breathy voiced stops can be considered as follows: the vocal folds are partially and moderately open and are vibrating without complete closure throughout the closure and noise interval of the voiced breathy portion. So the air flows out through the glottis and this vibration state is loose, and the rate and extent of vibration is considered to be low, resulting in the low Fo.

Chapter 8

Cross-Language Comparisons

8.1. Introduction

This chapter will examine some cross-language differences of voicing contrasts of stops in the six languages under investigation. Based on the results of the acoustic analysis described in chapters 2 - 7, cross-language examination will be made of laryngeal timing represented by voice onset time (VOT), fundamental frequency (Fo) and the Fo contour, spectral analysis,and F1 onset frequency. We would like to examine how language-specific properties differ from each other across languages, on what dimensions they are different, and how such differences are correlated with voicing categories of stops.

8.2. Classification of stops in each language

In order to examine some cross-language differences of initial stop voicing in six languages, it is imperative to have accurate phonetic information on the classificatory labellings of the stops. In two-category languages, Japanese utilizes the contrasts of stops which are phonetically voiced unaspirated stops and voiceless unaspirated stops. These have variants in word-initial position; /p/ has [p] and [pʰ] and /b/ has [b] and [b̥]. Mandarin Chinese, however, has voiceless unaspirated stops and voiceless aspirated stops, which have only one realization in word-initial position; /p/ is realized as [p] and /pʰ/ as [pʰ]. In three-category languages, Burmese and Thai have voiced unaspirated stops, voiceless unaspirated stops, and voiceless aspirated stops. This three-way phonemic contrast is maintained by the allophonic realization in initial position; /b/ is realized as [b], /p/ as [p], and /pʰ/ as [pʰ]. Thai has exactly the same contrast in word-initial and word-medial positions, but there is no three-way opposition in word-final position. Korean has a three-category contrast of stops, and they are all voiceless differing in aspiration and tensity. They are called voiceless tense unaspirated stop, voiceless lax unaspirated stop, and voiceless (tense) aspirated stop, and this contrast is maintained in allophonic realization; /p*/ is realized as [p*], /p/ as [p], and /p^{h}/ as [p^{h}]. The lax unaspirated stops are slightly aspirated in initial position and become voiced unaspirated stops in word-medial position. In four-category languages, stops are distinguished on the basis of contrasts of voicing and aspiration, and they are phonemically voiced unaspirated, voiced aspirated, voiceless unaspirated, and voiceless aspirated, and this phonemic contrast is maintained in allophonic realization. As we have shown in Chapters 2 - 7, these labelled stops do not always signify the same category and have in some cases qualitatively and quantitatively different values in several acoustic dimensions. The stop inventory (phonemic) in the six languages can be shown as follows:

Japanese b - p d - t g - k

Mandarin p - p^{h} t - t^{h} k - k^{h}

Korean p^{h} t^{h} k^{h}
/ \ / \ / \
p* - p t* - t k* - k

Burmese p^{h} t^{h} k^{h}
/ \ / \ / \
b - p d - t g - k

Thai p^{h} t^{h} k^{h}
/ \ / \ \
b - p d - t k

Hindi
b̤^{ɦ} - p^{h} d̪̤^{ɦ} - t̪^{h} ɖ̤^{ɦ} - ʈ^{h} g̤^{ɦ} - k^{h}
| | | | | | | |
b - p d̪ - t̪ ɖ - ʈ g - k

8.3. Voice Onset Time (VOT)

As mentioned before, voice onset time (VOT) is a timing dimension to represent the interval between the onset of voicing and oral articulatory release. It is usually understood that voiced stops have prevoicing, i.e., voicing precedes stop release; voiceless unaspirated stops have a short lag of voicing, i.e., voicing follows stop release; voiceless aspirated stops show a long voicing delay. Although recent studies indicate that several simultaneously occurring non-temporal cues may play a role in the voicing distinction, VOT is considered to be an efficient dimension for distinguishing initial stops. Table 8.1 presents the mean value and range of VOT in six languages.

Table 8.1 Mean VOT and range (msec) in Japanese, Mandarin, Korean, Burmese, Thai and Hindi

Japanese

	VOT	Range		VOT	Range
/b/	-89	(-65 - -125)	/p/	41	(15 - 65)
/d/	-75	(-40 - -135)	/t/	30	(15 - 50)
/g/	-75	(-35 - -125)	/k/	66	(50 - 100)

(N=72 for bilabials and alveolars, 120 for velars)

Mandarin

	VOT	Range		VOT	Range
/p/	7	(5 - 10)	/pʰ/	96	(80 - 115)
/t/	12	(10 - 15)	/tʰ/	98	(80 - 120)
/k/	19	(15 - 25)	/kʰ/	112	(90 - 130)

(N=18 for bilabials, alveolars and velars)

Korean

	VOT	Range		VOT	Range		VOT	Range
/p*/	10	(0 - 30)	/p/	31	(20 - 40)	/pʰ/	86	(75 - 95)
/t*/	11	(5 - 20)	/t/	20	(15 - 25)	/tʰ/	85	(75 - 105)
/k*/	23	(15 - 40)	/k/	49	(20 - 70)	/kʰ/	96	(85 - 110)

(N = 24 for tense, 12 for lax and aspirated)

Burmese

	VOT	Range		VOT	Range		VOT	Range
/b/	-104	(-90 - -120)	/p/	3	(0 - 10)	/pʰ/	46	(35 - 70)
/d/	-106	(-85 - -125)	/t/	15	(10 - 25)	/tʰ/	67	(55 - 85)
/g/	---		/k/	31	(30 - 35)	/kʰ/	76	(55 - 95)

(N = 6 for bilabials, alveolars, and velars)
--- data was not obtained.

Thai

	VOT	Range		VOT	Range		VOT	Range
/b/	-104	(-85 - -120)	/p/	5	(5 - 10)	/pʰ/	73	(50 - 95)
/d/	-106	(-70 - -140)	/t/	8	(5 - 10)	/tʰ/	76	(65 - 95)
			/k/	23	(15 - 30)	/kʰ/	95	(85 -105)

(N = 12 for bilabials, alveolars, and velars)

Hindi

	VOT	Range		VOT	Range		VOT	Range		VOT	Range
/b/	-93	(-50/-110)	/b̤ʱ/	-93	(-60/-120)	/p/	12	(15/20)	/pʰ/	75	(50/90)
/d̪/	-115	(-85/-125)	/d̪̤ʱ/	-90	(-60/-110)	/t̪/	12	(10/15)	/t̪ʰ/	82	(60/100)
/g/	-121	(-95/-140)	/g̤ʱ/	-104	(-65/-120)	/k/	34	(25/40)	/kʰ/	119	(110/125)

(N = 9 for bilabials, dentals, and velars)

As is seen in Table 8.1, there is a clear difference in VOT between the voicing categories in Japanese, Mandarin, Burmese and Thai. In these languages, VOT serves to make a distinction between the categories. But in languages such as Korean and Hindi, there is an overlap in the range of VOT - between the tense and lax stops in Korean and between the two voiced stops in Hindi. These overlaps have also been pointed out by other investigators (Han and Weitzman, 1970; Benguerel and Bhatia, 1980) and this indicates that VOT is not functional for distinguishing between these voicing categories in these two languages. It can be considered that there exist another dimensions differentiating these voicing categories in the articulation of these stops.

It can also be noted that there is a difference in the range of distribution of the VOT values. In Japanese which is a two-category language, the VOT values for voiced and voiceless categories are distributed in a wide range and show some degree of variability. In languages which have three or four categories, the VOT values for voiceless unaspirated stops are distributed in a rather narrow range of the continuum, and in these languages there are cases where the two categories show similar VOT values. Although further evidence is needed, the range of VOT distribution is somewhat correlated with the number of phonetic categories.

As to the VOT value in relation to place of articulation, it can be seen that for voiceless categories the value is greater for velar stops than for bilabial or alveolar stops, though this tendency is not evident in the voiced categories. The tendency is often found in other languages such as English and, as pointed out in previous chapters, there are several reasons for this. First, the volume of the supralaryngeal cavity behind the constriction for velar stops is smaller than those for bilabial and alveolar stops, resulting in higher pressure in the cavity. This suggests that the onset of voicing takes longer than for other types of stop because of the higher pressure in the oral cavity. This is a commonly held view, and, as expected, there is a tendency for alveolar stops to have a longer VOT than bilabial stops because of the difference in volume of the supralaryngeal cavity behind the constriction. The other reason is the speed of the articulators. It is pointed out that the back part of the tongue moves more slowly than the tip of the tongue or the lips, and this means that it takes longer for velar stops to attain the transglottal pressure difference which is sufficient for vocal-folds vibration than alveolar or bilabial stops, resulting in a delayed voicing (Hardcastle, 1973:266).

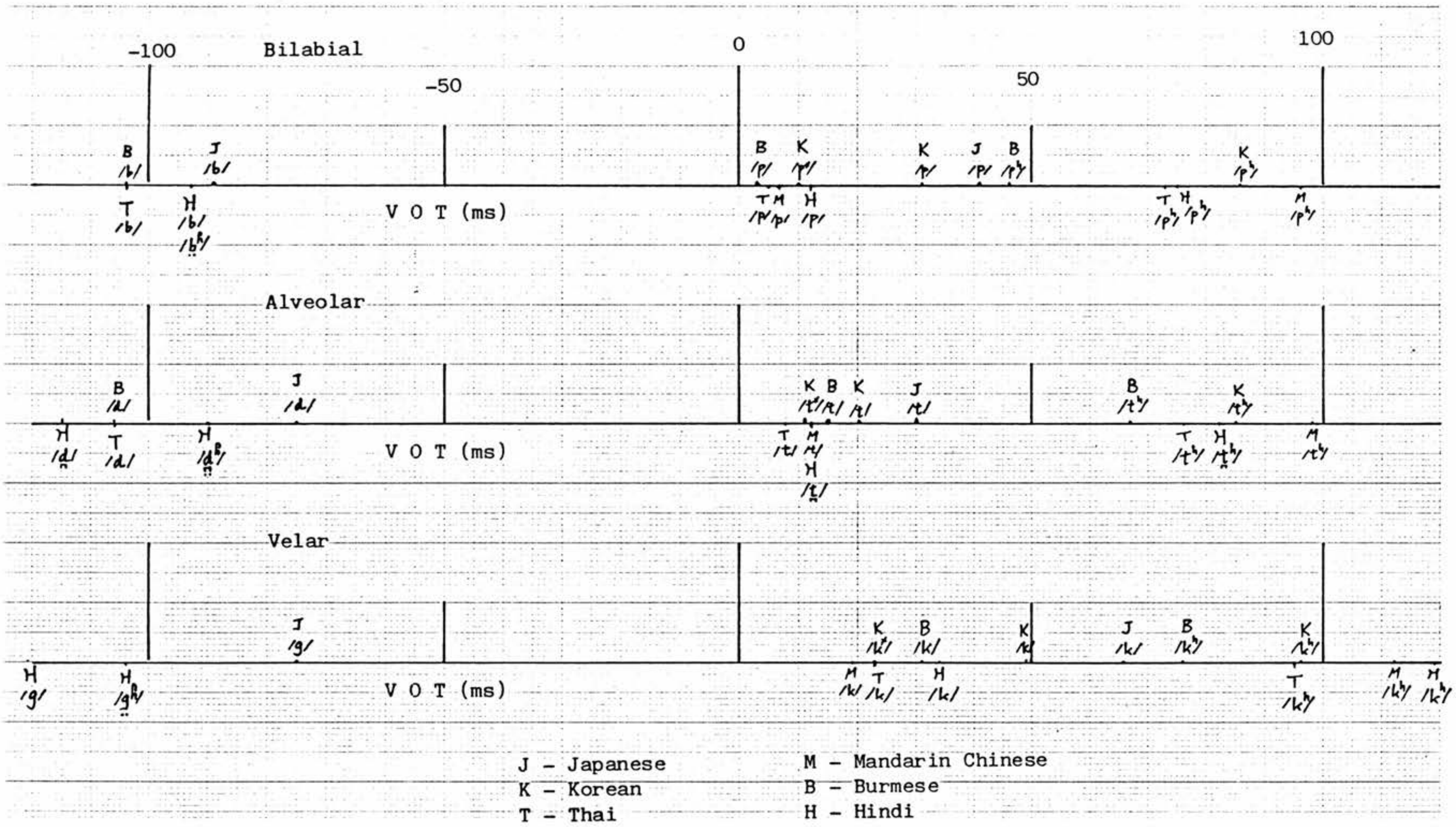


Figure 8.1 Voice onset time distributions based on the mean VOT values in six languages (bilabial, alveolar, and velar stops)

Based on Table 8.1, the averaged values of stop consonants in each language can be plotted in Figure 8.1. As shown in Figure 8.1, stops in different languages choose their own points in the timing continuum in a language-specific way. The bilabial stop /p/ in Burmese and Thai selects roughly the same timing point, while /p/ in Korean and Japanese selects a rather later point. It can be noted that /p/ in Japanese is medially aspirated and is close to /p^h/ in Burmese. Voiceless aspirated stops for three places of articulation are distributed in a rather limited range of delayed onset. Moreover, no types of stop consonant utilize the range from - 60 ms to zero as far as the mean value distribution is concerned. From the examination of Table 8.1 and Figure 8.1, it can be seen that the major voicing categories fall in the following three ranges, though discreteness between voiceless unaspirated and aspirated stops is not apparent:

voiced stops	-80 - -110 ms
voiceless unaspirated stops	5 - 45 ms
voiceless aspirated stops	70 - 100 ms

According to Lisker and Abramson (1964), the timing dimension of VOT falls into three categories: voicing lead, short lag, and long lag. Each category has three distinct ranges - one from about -125 to -75 ms, one from 0 to +25 ms, and a third from +60 to +100 ms, respectively. The ranges mentioned above are in general agreement with this and support these divisions of major voicing categories. It can be noted, however, that the category for voiceless unaspirated stops shows a wider range than that of Lisker and Abramson (1964) and the distinction between the two voiceless stops is rather fuzzy.

Next, in order to examine the cross-language differences of the voicing contrasts of stops, a one-way analysis of variance (ANOVA) was performed on the VOT values in these languages. The VOT values of each category were "standardized" (averaged) for different phonetic environments and different subjects in each language. For each phonetic category which is transcribed by the same phonetic symbols, ANOVA was applied to examine the language effect. The results can be shown as follows:

Table 8.2 One-way ANOVA Results: Language Effect on VOT in Six Languages

Segment	df	F ratio	Significance
/p/	F(5, 54)	23.21	$p < .01$
/t/	F(5, 54)	13.02	$p < .01$
/k/	F(5, 54)	12.44	$p < .01$
/b/	F(3, 36)	3.05	n.s.
/d/	F(3, 36)	4.34	n.s.[1]
/p^h/	F(4, 45)	2.85	n.s.
/t^h/	F(4, 45)	2.46	n.s.
/k^h/	F(4, 45)	3.28	n.s.

As shown in Table 8.2, stops /p, t, k/ in ANOVA yielded considerable differences among languages and the proportion of variance is significant. For the aspirated stops, however, the F ratios are small and there is no significant variance across the languages. For voiced stop /d/, there is a variance greater than those of aspirated stops, but the proportion of variance is still not significant. The ANOVA results indicate that in the production of aspirated stops, there is no significant difference across languages in the laryngeal timing events in relation to oral release, and similarly labelled aspirated stops are articulated in similar timing events in these languages. This means that aspiration requires a carefully adjusted timing between glottal and supralaryngeal events, and there is very little flexibility in the timing events in these languages. On the other hand, there is a considerable amount of variance in the production of voiceless unaspirated stops /p, t, k/ across languages, and this indicates that there is a wide range of timing points for producing these stops in these languages. As mentioned, the data on VOT values show that stops /p, t, k/ in Burmese and Thai are articulated in a rather narrow timing range, while those in Japanese are in a wide timing range.

---

[1] It is not significant at 1 % level, but is significant at 5 %.

To summarize, there is a difference between major voicing categories along the continuous dimension of VOT. VOT is functional in distinguishing major voicing categories in Japanese, Mandarin, Burmese, and Thai. In languages such as Korean and Hindi which have three or four voicing categories, two categories show similar VOT values. Furthermore, voiceless aspirated stops use a limited range of VOT, while voiceless unaspirated stops use a wider range of the variation in the continuum.

8.4. Fundamental Frequency (Fo) and its Contour

As has been demonstrated, voicing contrasts are closely related to pitch perturbations in the following vowels. The difference of pitch range is considered to be one of the properties for a voiced - voiceless distinction. In the present study, the measurements of Fo were made at the vowel onset following a stop release. Table 8.3 presents a pooled mean for major voicing categories in each language.

Table 8.3 Mean Fo for major categories at vowel onset (Hz)
(s.d. in parenthesis)

Language	Major Category	Fo(Hz)	Difference
Japanese	vd.stops	213.8(6.2)	
	vl.stops	248.5(19.5)	34.7(vl.- vd.)
(N=66 for voiced stops and voiceless stops)			
Mandarin	vl.unasp.	151.0(5.9)	
	vl.asp.	158.1(8.1)	7.1(asp.- unasp.)
(N=54 for vl.unaspirated and vl.aspirated stops)			
Korean	vl.tense	175.1(16.7)	33.9(tense - lax)
	vl.lax	141.2(30.5)	
	vl.asp.	180.8(24.6)	39.6(asp.- lax)
(N=48 for vl.tense, vl.lax and vl.aspirated)			
Burmese	vd.stops	167.0(6.9)	
	vl.unasp.	187.7(5.0)	20.7(unasp.- vd.)
	vl.asp.	186.8(6.6)	19.8(asp.- vd.)
(N=18 for vd. stops and vl.unaspirated and vl.aspirated)			
Thai	vd.stops	183.5(14.0)	
	vl.unasp.	186.0(9.9)	2.5(unasp.- vd.)
	vl.asp.	205.1(14.2)	21.6(asp.- vd.)
(N=24 for voiced stops,36 for vl.unaspirated and vl.aspirated)			
Hindi	vd.stops	115.7(19.2)	
	breathy	108.9(14.7)	6.8(vd.- breathy)
	vl.unasp.	130.4(27.6)	14.7(vl.- vd.)
	vl.asp.	122.3(28.8)	6.6(asp.- vd.)
(n=36 for vd.stops, breathy, vl.unasp. and vl.asp.)			

There are several points to note in the Fo data. Firstly, it can be seen as a general trend that Fo following voiceless unaspirated stops and voiceless aspirated stops is considerably higher than that following voiced ones. In Thai, however, there is no significant difference in Fo value between voiced stops and voiceless unaspirated stops (two-tailed t-value=0.46, n.s.). Secondly, voiceless tense stops in Korean show a higher Fo than voiceless lax stops. This implies that there is a difference in the initial glottal state between the two stops. According to Hirose et al. (1974), there is a sharp increase in the activity of thyroarytenoid for tense stops and this intrinsic muscle activity may be relevant for higher Fo. Thirdly, the Fo of breathy stops in Hindi is the lowest among the four categories of stops, and this agrees with the previous study (Kagaya and Hirose, 1975). Although the articulatory cause for this markedly low Fo of breathy stops needs further investigation, we can consider that the vocal folds do not close fully during vibration and vibrate loosely because of this incomplete closure. Additionally, the transglottal pressure difference may not be high enough to maintain a higher Fo. Lastly, it can be observed that there is a difference in the amount of variation between the voicing categories: in Japanese and Korean, the difference between the major categories is greater than that in Thai, Mandarin and Hindi. This implies that there is a difference in the initial glottal state and airflow rate between the categories in these languages.

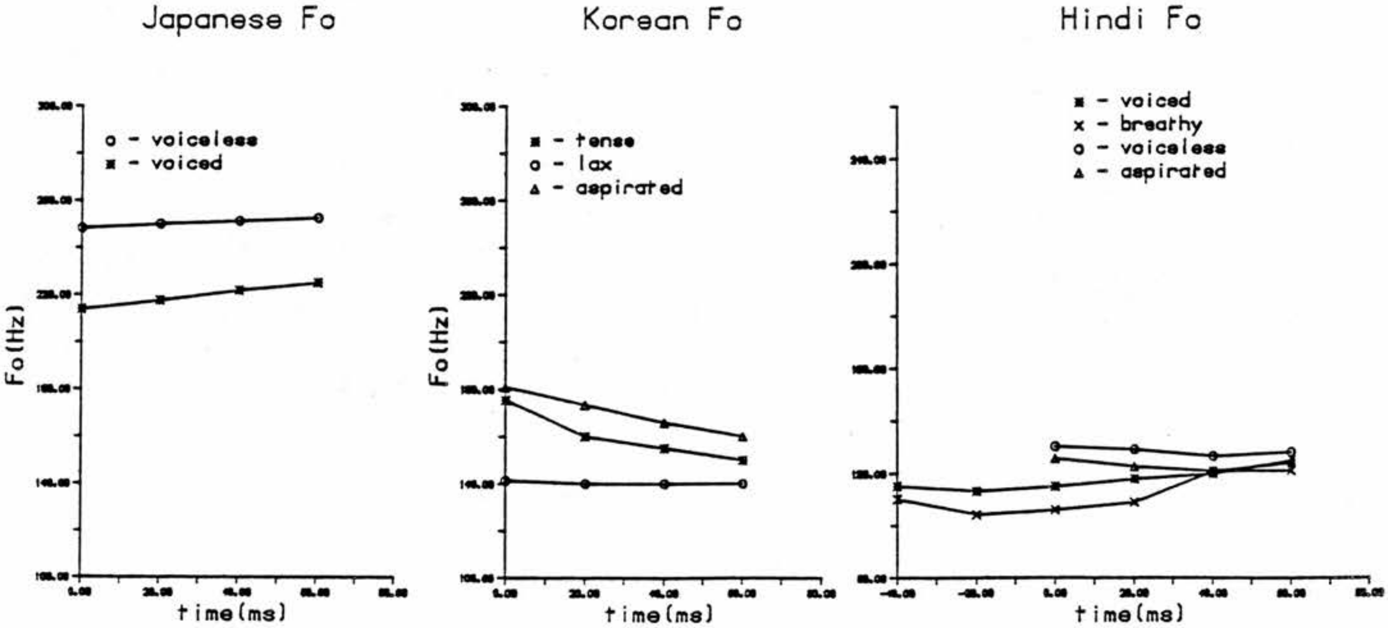


Figure 8.2 Schematized Fo curves in Japanese, Korean and Hindi[2]

Japanese(left) * voiced o voiceless
Korean(centre) * tense o lax ▲ aspirated
Hindi(right) * voiced x breathy
o vl.unaspirated ▲ vl.aspirated

---

[2] The zero point in Figure 8.2 indicates the onset of voicing, and it should be noted that the onset in the voiceless stops does not coincide with the one in real time, i.e., as shown in VOT. The diagram transfers the beginning of voicing to the zero point.

Associated with the differences in the onset Fo values, there can also be a difference in the Fo curves from the onset of vowels to the steady-state portion of vowels, as discussed by Hombert (1978). In the present study the Fo curves were examined for a period of 50 to 60 ms after the vowel onset. Figure 8.2 shows the schematized Fo curves for the vowels following the major voicing categories in Japanese, Korean, and Hindi. Burmese and Thai were excluded since they are tonal contrast languages, and it is rather difficult to minimize the effect of tone on the intrinsic Fo curve.[3] From the examination in Figure 8.2, it is seen that the Fo curves following the voiced stops in Japanese and Hindi show a gradual rising pattern, while the voiceless aspirated stops in Korean and Hindi show a gradual falling pattern in the same period. For voiceless unaspirated stops, the curve in Japanese shows a slight rising pattern within 50 to 60 ms, while the one in Hindi shows a steady fall. Furthermore, the Fo curve of the voiceless tense stop in Korean shows an abrupt drop immediately after the vowel onset. The Fo curve of breathy stops in Hindi shows a steady fall during closure, and upon release gradually goes up. This fall - rise pattern of Fo seems to be characteristic of breathy stops. Although the two voiced stops in Hindi show a similar direction of change, those of breathy stops are more distinctive and pronounced than those of voiced (unaspirated) stops.

8.5. Spectral Analysis

It is generally considered that there is a difference in the articulatory force, i.e., the rate of airflow, at the time of consonantal release between voiced and voiceless categories. In the present study, the short-time spectra were sampled by ILS linear prediction for the initial period of 25 or 30 ms immediately after the consonantal release. This period includes both the burst and some portion of the voicing onset. In the case of voiced stops, this period includes the burst and some portion of the vowel onset, whereas in the case of voiceless stops, this includes mainly the burst and some portion of aspiration, and it does not extend

---

[3] Japanese is a pitch-accented language and is termed a register-tone language, as opposed to a contour-tone language.

to vowel onset. Although this period is considered to contain some invariant cues for place of articulation (Blumstein and Stevens, 1979), the present study mainly examined the intensity and spectral characteristics. The power spectra and smoothed spectra were sampled for most of the tokens in each language and some of the examples of smoothed power spectra are presented in Figure 8.3.

Although spectral characteristics are rather difficult to quantify, the examination of power spectra in each language reveals some differences in intensity level and spectral characteristics. For Japanese, it can be noted that the intensity level of voiceless stops is slightly higher than that of voiced stops, but the difference is not consistent in some tokens. Therefore, the difference in the overall intensity level does not appear to be a major acoustic dimension for the voicing differences. Although the difference in the intensity level is not consistent, there is a small but noticeable difference in spectral shape between the two voicing categories. There is a spectral peak in the low frequency region for the voiced stops, but not in voiceless ones.

For Mandarin Chinese, voiceless unaspirated stops show regularly distributed peaks and narrower bandwidths than voiceless aspirated stops. It can also be seen that unaspirated stops show a higher dB than aspirated stops. This finding may raise a problem on the traditional distinction between fortis and lenis in Mandarin Chinese.

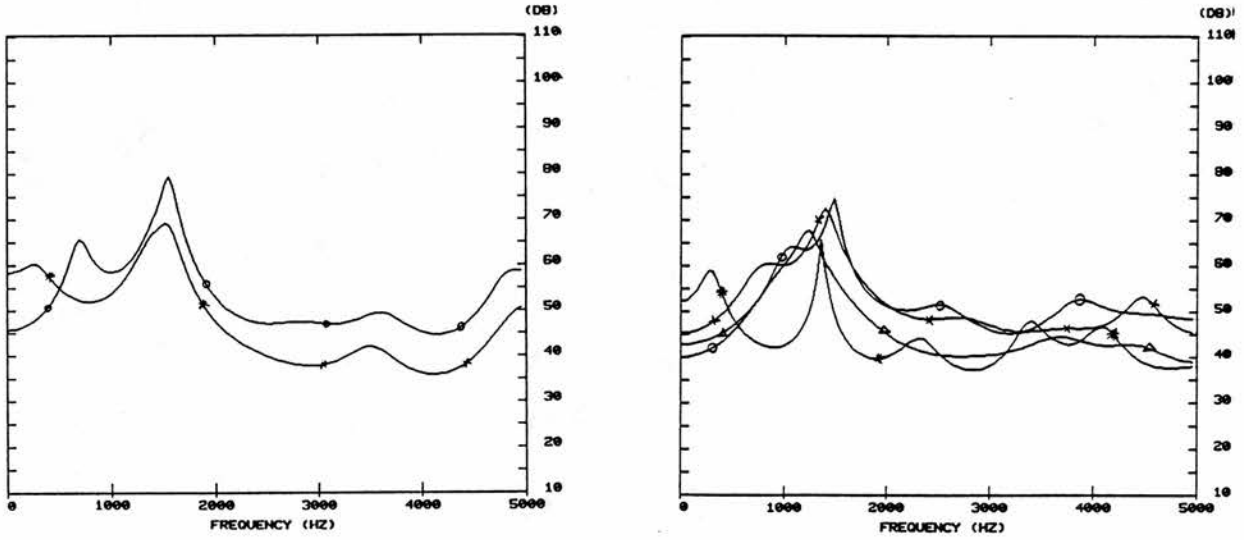


Figure 8.3 Examples of power spectra sampled at the time window(30) after release of Japanese(left) and Hindi(right) stops
Japanese ✳ [ga] o [ka]
Hindi ✳ [gan] × [g̤^{ɦ}an] o [kan] △ [k^{h}an]

For Korean stops, the power spectra indicate that the intensity of aspirated stops is considerably higher than that of the other types of stop. Although it is pointed out that there is a difference in the time to build-up intensity; a shorter rise time for tense stops than for other stops, the results in the present study did not show a noticeable difference between the three voicing categories (Han and Weitzman, 1970).

For Burmese stops, the power spectra showed that there is a difference in the overall intensity between voiced and voiceless stops; the intensity for voiceless stops is relatively higher than it is for voiced stops, but there is no marked difference between the two voiceless stops.

For Thai stops, the voiceless aspirated stops are clearly distinct from the other two voicing categories; the voiceless aspirated ones show less intensity level and energy peaks are less regularly distributed, while in the voiced stops and voiceless unaspirated stops, energy peaks are apparent and are regularly distributed.

For Hindi stops, there is a marked difference between the two voiced categories. The intensity level of breathy stops is greater than that of voiced unaspirated stops. There is also a difference in the spectral shape which should be noted. As seen in Japanese, the voiced unaspirated stops show a peak in the low frequency region, but other types of stop do not.

8.6. The Onset Frequency of the First Formant

The onset frequency of the first formant (F1 onset frequency) is another dimension for distinguishing voicing categories of stops: vowels following voiced stops show a low F1 onset frequency and contain a rising F1 transition. On the other hand, vowels following voiceless stops show a high F1 onset frequency and may not contain the transition. The results of the measurements of F1 onset frequencies in the six languages can be shown as follows:

Table 8.4 Mean F1 Onset Frequency of Voicing Categories in Six Languages (Hz)

Japanese

/b/	430.9	/p/	550.0
/d/	400.9	/t/	441.4
/g/	295.7	/k/	391.2

(N=72 for bilabial and alveolar, 120 for velar)

Mandarin Chinese

/p/	401.2	/pʰ/	436.6
/t/	333.7	/tʰ/	418.8
/k/	604.0	/kʰ/	834.0

(N=18 for bilabial, alveolar and velar)

Korean

vl.tense	468.7	vl.lax	501.3	vl.aspirated	580.9

(N=27 for vl.tense and vl.aspirated, 36 for vl.lax)

Burmese

/ba/	830.0	/pa/	788.0	/pʰa/	935.0
/da/	228.5	/ta/	264.5	/tʰa/	936.0

(N= 6 for each item, based on the data of Subject 2)

Thai

/baa/	830.0	/paa/	806.0	/pʰaa/	1189.3
/dam/	613.0	/tam/	807.0	/tʰam/	1006.7

(N= 12 for each word)

Hindi

voiced stops	327.8	breathy voiced	819.3
vl.unaspirated	489.9	vl.aspirated	819.0

(N=36 for voiced, breathy, vl unasp. and vl.asp.)

As is shown above, there is a tendency for F1 onset frequency to increase as the voicing category changes from voiced stops to voiceless aspirated ones; that is, the frequency proportionally increases as the VOT increases in the VOT continuum. The bases for the proportional relationship can be explained from articulatory mechanisms in production. If there exists voicing, i.e. onset of periodicity, F1 is expected to begin very low and will rise to the steady state. In voiceless stops, however, the onset of periodicity begins at some later time than in voiced stops because of the presence of aspiration between release burst and the onset of periodicity, and results in the absence of F1 onset transition.

Furthermore, although it is not shown in the above data, it should be pointed out that the effect of voicing categories on F1 onset frequency is not equally seen for all vowels followed. As shown in the results on F1 onset frequency in Japanese (Chapter 2), the effect on low vowels such as /a/ is more apparent than that on high vowels such as /i, u/. Low vowels are considered to have higher F1 frequencies compared to high vowels. It can be said that words containing low vowels or with high F1 frequency show larger voicing-related changes than words containing high vowels or with low F1 frequency.

8.7. Discussion

The acoustic results given above indicate that there are some differences in the way the voicing categories are characterized in the six languages, and there are also some similarities shared by the same categories in these languages. The use of each acoustic dimension for voicing distinction is summarized in Table 8.5. A slash in Table 8.5 indicates that the voicing categories are separated by the acoustic feature, and a dotted slash indicates a weak tendency of separation.

Table 8.5 The use of acoustic features which differentiate the voicing categories in six languages

Language	**VOT**	**Onset Fo**	**Fo curve**	**Power spectra**	**F1 onset**
Japanese	Vd/Vl	Vd/Vl	Vd/Vl	---	Vd/Vl
Mandarin	Unasp/Asp	Unasp/Asp	---	Unasp/Asp	Unasp/Asp
Korean	T L/Asp	L/T Asp	L/T/Asp	L T/Asp	L T/Asp
Burmese	Vd/Vl/Asp	Vd/Vl Asp	---	Vd/Vl Asp	Vd Vl/Asp
Thai	Vd/Vl/Asp	Vd Vl/Asp	Vd/Vl Asp	Vd Vl/Asp	Vd Vl/Asp
Hindi	B Vd/Unasp/Asp	B Vd/Unasp Asp	B/Vd/Unasp Asp	B Vd/Unasp Asp	Vd/Unasp/B Asp

(Vd=voiced, Vl=voiceless, T=voiceless tense
L=voiceless lax, Asp=Voiceless aspirated,
Unasp=voiceless unaspirated, B=breathy)

Voice onset time (VOT) is an efficient dimension for distinguishing voicing differences in Japanese, Mandarin Chinese, Burmese and Thai. In these languages, there is a clear-cut difference for the voicing categories. As seen in the range of VOT, Japanese shows a rather wide variability of VOT for each category, while voiceless unaspirated stops in Burmese, Mandarin Chinese and Thai show a restricted range of VOT variation. It can also be seen that Burmese aspirated stops show a medium range of voicing delay, compared to those in Korean and Hindi. As pointed out in earlier studies (Han and Weitzman, 1970; Benguerel and Bhatia, 1980), VOT is not sufficient to separate the voicing categories from each other in Korean and Hindi, and there should be other additional dimensions to characterize them. VOT represents a timing relationship relative to oral release. So VOT serves to separate the categories from each other if the voicing distinction is based on the laryngeal timing events, but if the voicing categories involve other laryngeal features such as glottal tension in Korean and glottal gesture in Hindi, then, other acoustic dimensions are needed for the distinction.

Furthermore, as is known, the relative timing of laryngeal and supralaryngeal events is important in distinguishing the voicing categories. As mentioned above, there are three timing domains for major voicing categories, and among these domains, the one for voiceless unaspirated stops appears to show language specific timing events and a variable domain across languages. For instance, /p, t, k/ in Japanese show a wide range of timing variation, while those in Thai and Burmese show a rather limited range. Some languages such as Korean and Hindi use other laryngeal features than timing relation for the voicing distinction, and it appears that such features are incorporated or "added" within the domain of the timing continuum for voiceless unaspirated stops. On the other hand, voiceless aspirated stops show little variance in the range of timing among languages. In my view this is because aspiration requires a wide open glottis, a timing adjustment between oral release and peak glottal opening, and sufficient airflow. There seems to be little flexibility in the timing event in the production mechanisms of aspiration.

The Fo data indicate that the frequency at vowel onset is relatively higher in the voiceless stops than in the voiced stops. The Fo curves are rising for the voiced categories, while those for the voiceless ones are level or falling. The voiceless aspirated stops show an initially high and then a gradual falling pattern. There are several points to note on the Fo and its curves. Firstly, the Fo curves of the voiceless tense stops in Korean begin in a higher range and abruptly fall within a period of 30 ms after the vowel onset. As mentioned, physiological studies made by Hirose et al. (1974) indicate that there is a sharp increase in the activity of the thyroarytenoid, and this increase of tension and the following relaxation are considered to be responsible for abrupt movement of the Fo. Furthermore, Kagaya (1974) indicates that aspirated stops in Korean are produced with a maximally opened glottis and high airflow rate. From these studies, there appear to be two different causes for a higher Fo in tense stops and aspirated stops in Korean. Secondly, it should be noted that the Fo at the breathy portion in Hindi is the lowest among the four types of stops, and the breathy stops show a distinctive fall-rise pattern of Fo. To explain this markedly low Fo, it can be speculated that the vocal folds vibrate loosely with a narrowly opened glottis after the articulatory release, resulting in the low Fo at the onset. Although the influence of breathy stops on the following vowel was examined by Schiefer (1986), the pattern of fall-rise itself should be correlated with the breathy state of the glottis.

The physiological reason for the Fo perturbations of the vowel immediately following a consonant has been examined by several investigators (Hombert, 1978, Hombert et al., 1979; Ohala, 1978). Among these studies, the most comprehensive examination has been made by Hombert et al.(1979). They examined the two hypotheses in detail; namely "aerodynamic" and"vocal-cord tension" hypotheses, and mentioned that these hypotheses are not sufficient to explain all aspects of Fo perturbations. Based on their experiments, they proposed a plausible explanation in terms of the vertical position of the larynx. As stated earlier, Hombert et al. (1979) showed that there is a difference in larynx height between voiced and voiceless consonants and it is the larynx height which is most compatible with the

Fo perturbations. The lowering of the larynx causes the lowering of the Fo, and expands the supralaryngeal cavity and by this expansion voicing is maintained. Although no measurement was made in the present study of the position of the larynx, it seems that vertical movement of the larynx is more relevant for the causes of the perturbations, though the relation between Fo perturbations and a change of supralaryngeal configurations needs to be further examined.

Furthermore, it is interesting to note that the Fo value does not differ as a function of place of articulation in these languages. A change in the place of articulation affects the size of the supralaryngeal cavity, and this should result in a change of airflow rate in the cavity. The fact that the Fo does not differ as a function of place of articulation means that the aerodynamic hypothesis is rather weak for explaining the Fo perturbations at the consonantal release.

As to the spectral analysis, it is rather difficult to quantify objectively, but it presents intriguing results for the voicing distinction. Although in general the intensity level is relatively higher for the voiceless stops than it is for the voiced ones, it is not the case in Japanese. This implies that the articulatory force relates not only to airflow rate, but also to the timing of glottal opening in relation to the oral release. Furthermore, aspiration which is generally characterized by the timing differences in VOT and the presence of turbulent airflow and aperiodic noise is considered to give a high intensity, and this holds true in Korean, but is not true for Hindi and Thai voiceless stops. Another point that should be noted here is that there appears to be a difference in spectral shape in the initial portion immediately after the release. As shown in Figure 8.3, the voiced category /g/ shows a peak in the low frequency region, but this is not present in breathy voiced stops and voiceless stops. Although further evidence is needed, the spectral characteristics in the low frequency region may serve to signal the voiced category.

8.7.1. Phonetic Aspects between Supralaryngeal and Glottal Events

Based on the acoustic results mentioned above and other studies on glottal adjustments, the phonetic characteristics between supralaryngeal and glottal events for voicing categories can be summarized as follows:

(1) Voiced unaspirated stops

In the present analysis, all voiced unaspirated stops in Japanese, Burmese, Thai, and Hindi show a long voicing lead and show low Fo values of following vowels after voiced stops. In Thai, however, there is no marked difference in Fo value between voiced unaspirated and voiceless unaspirated stops. Based on these observations and other studies on glottal adjustments (Japanese: Sawashima and Niimi, 1974; Hindi: Dixit, 1987), it can be considered that regardless of what allophonic forms appear the vocal folds remain in a voicing state and continue vibrating throughout the closure. There should be a sufficient airflow through the glottis.

(2) Breathy voiced stops

In the present analysis, Hindi exemplifies this type of stop. The acoustic analysis shows a long voicing lead, the lowest Fo of following vowel among the four types of stops, and a relatively long breathy duration. Dixit (1987) shows that this type of stop is produced with a moderately open glottis, and that the opening gesture of the glottis starts during the closure, and the peak of glottal opening occurs during the breathy noise period. It can be considered, therefore, that the vocal folds remain in the state of vibrating during the closure. Vibration continues after articulatory release with a slightly open glottis, and turbulent air flows during the breathy period. We can consider that the size and open - closed timing of glottis with respect to oral release are essential for this type of stop.

(3) Voiceless unaspirated stops

Voiceless unaspirated stops in the six languages show a wide range of variation in VOT values and a higher Fo than their voiced counterparts. This type of stop is the most unmarked among several types of phonetic categories of stop consonants. It can be considered that these stops are produced with an open glottis and the closing of the glottis begins somewhat after the articulatory release. But the stops in Thai show slight differences in the glottal events; as shown in the VOT values, they tend to be produced with simultaneous oral and glottal closure and release. There should be some differences in the size of glottal opening, and it is pointed out that there is some degree of variability with respect to the glottal opening (Dixit, 1987). Although there is no data available on the size of the glottal opening in Mandarin Chinese, Burmese, and Thai, this variability of the glottal width may lead to the variability in the onset of voicing.

(4) Voiceless tense unaspirated stops

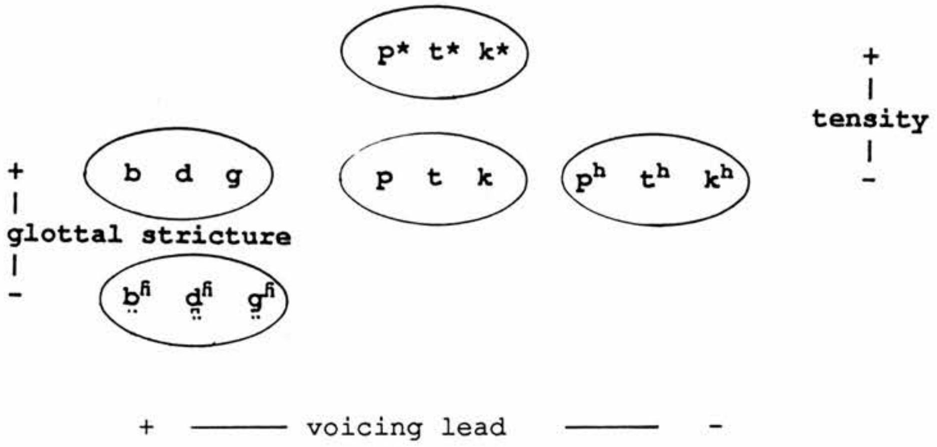
Korean has this type of stop, which shows a very short VOT value and is different from the voiceless lax unaspirated stops. Kagaya (1974) shows that they have a very small glottal opening during closure and the opening period is very brief. What is characteristic of this type is that it shows tight closing of the glottis immediately before the release.

(5) Voiceless aspirated stops

This type of stop is found in Mandarin Chinese, Korean, Burmese, Thai and Hindi in the present analysis. In Korean, the stops are quite heavily aspirated, whereas in Thai they are "laxly" aspirated, i.e., with breathy release. As expected, all voiceless aspirated stops show a long voicing delay except the one in Burmese which shows a rather medium range of voicing. It can be considered that these stops are produced with a wide open glottis during closure and peak of the glottal opening is to occur at around articulatory release.

8.7.2. Voicing Categories of Stops

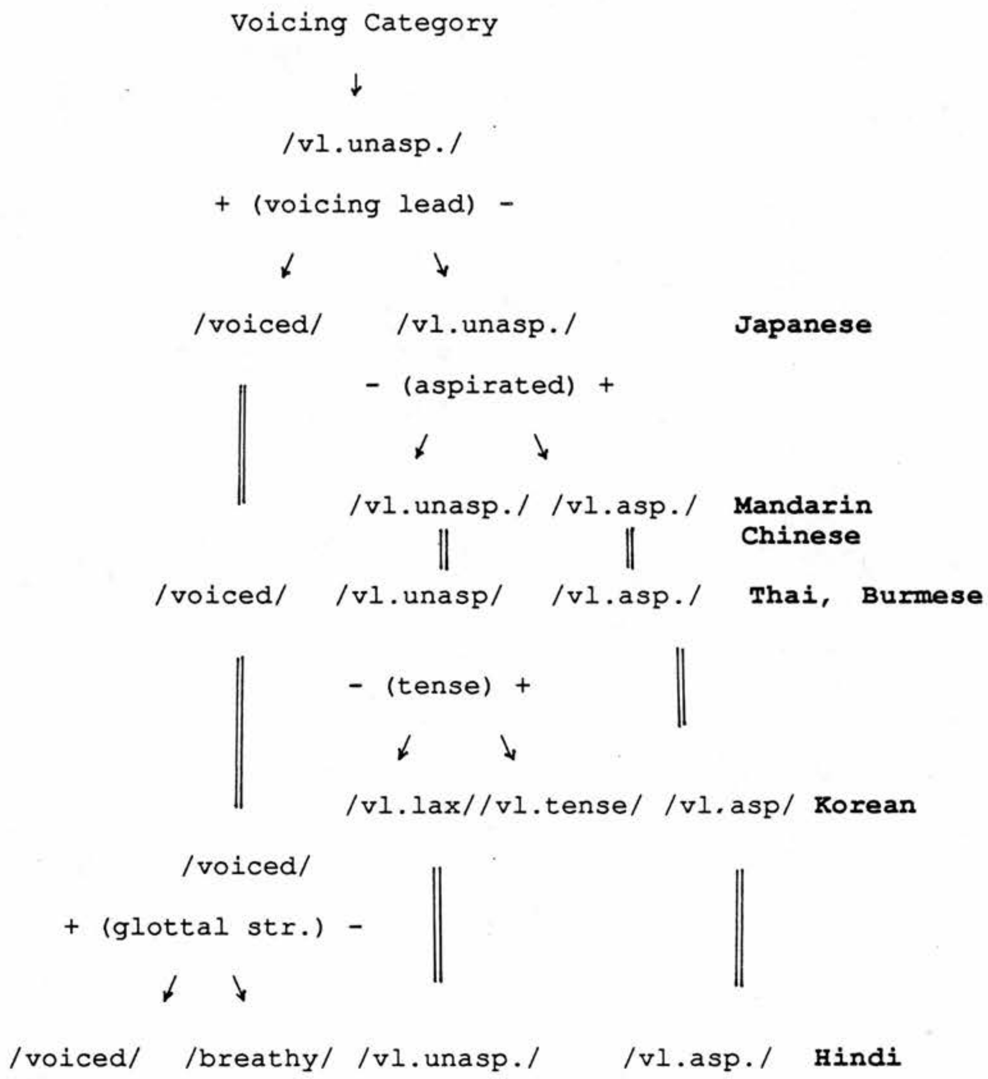
As is known, languages differ in the number of voicing categories. Some languages such as Hawaian have only one stop series, while some other languages such as Hindi have four stop series, and the majority of languages have two stop series: a distinction between voiced and voiceless stops. Through the examination of the acoustic results in the present study, it has become clear that the relation between glottal and supralaryngeal timing events is essential for characterizing the voicing categories. As is seen in Table 8.1 and Figure 8.1, VOT dimension can be divided into three major divisions which correspond to voiced, voiceless unaspirated, and voiceless aspirated stops, and these divisions support Lisker and Abramson's three basic voicing categories. As is known, VOT represents a timing dimension, and if voicing contrast involves features other than laryngeal timing, then, VOT is not functional for such contrast, and other properties are needed. In order to examine the categories in the present investigation, we can consider three parameters: laryngeal timing, laryngeal tensing, and glottal stricture, and the voicing categories are arranged as follows:[4]



The above relationship can be schematized in the following hierarchical structure:

---

[4] Several features are proposed to characterize the breathy voiced stops; e.g., [distinctive release], [delayed release] and [heightened subglottal]. We will here use [glottal stricture] following Ladefoged (1971) and Ohala (1979).



Arranged in this way, we can examine how voicing categories of stops might be formulated. In examining the number of voicing categories, it should be noted that languages tend to keep their contrast distinctively. First, the most common category of stop voicing is voiceless unaspirated stop, since it is considered to be most natural in terms of laryngeal and aerodynamic factors, and therefore, it is placed at the top of the hierarchy. This can also be substantiated in that almost all languages possess voiceless unaspirated stops in their consonant inventory. If a language has two series of stops, then, the contrast for voicing lead is applied, and this results in the contrast between voiced stops and voiceless stops. This contrast holds in Japanese. For Mandarin Chinese which is also a two-category language, the contrast for aspiration is applied, resulting in the contrast between voiceless unaspirated stops and voiceless aspirated stops. Next, for Thai and Burmese which are three-category languages, three distinctively opposite categories between voiced, voiceless unaspirated and voiceless aspirated series account for the contrast. However, for Korean which is also a three-category language but involves a contrast of laryngeal tensing, the contrast for tensing is applied to the voiceless unaspirated category, and this results in voiceless contrasts of tense, lax and aspirated stops. Finally, for Hindi which is a four-category language, the contrast for glottal stricture is applied, and as shown in the hierarchy, this results in the four voicing categories: voiced, breathy, voiceless unaspirated and voiceless aspirated. Thus we can consider the selection of phonetic categories of stop voicing involving several laryngeal parameters in terms of the hierarchical structure of the laryngeal features. Keating (1984a) argues that in explaining the number of phonetic categories, there is a polarization principle by which languages keep the categories further apart within the possible VOT continuum. While this principle can explain the two-category and three-category series if the voicing categories are based on laryngeal timing, it is unable to explain those languages which have more than two series and which involve laryngeal features other than timing. We think that a hierarchical structure of phonetic parameters is a more plausible way to explain the number and selection of the voicing categories where more than two categories are concerned.

8.8. Summary

The languages in the present study use several acoustic dimensions for voicing categories in different ways, and the "same" voicing categories have a language-specific variability, as well as features which are common to many languages. VOT functions for the distinction if the voicing categories are based on the laryngeal timing in relation to the oral release. If other laryngeal features are involved, other dimensions are needed for distinction. Among the major voicing categories, voiceless unaspirated stops show a wide range of variability in the glottal and supralaryngeal timing events, while voiceless aspirated stops do not. The Fo at the vowel onset represents the initial state of the glottal adjustments and is significant for characterizing the voicing categories which involve a change of initial glottal gesture as seen in Hindi breathy stops. Spectral analysis such as intensity level and spectral shape does not appear to provide a useful cue for distinguishing major voicing categories.

Chapter 9

Cross-Language Phonetics: Voicing Contrasts of Stops

9.1. Introduction

Through the examination of cross-language characteristics of initial-stop voicing in six languages, it can be seen that the "same" or similar sounds show some differences along several acoustic dimensions as well as similarities which are common to some or all the languages under investigation. It is not uncommon to find that the sounds in question in one language are different from those of other languages as a language-specific difference. For instance, as has been shown, voiceless unaspirated stops /p, t, k/ in Japanese are medially aspirated and show a wide range of VOT variation, whereas /p, t, k/ in Burmese and Thai are very weakly aspirated and show a rather limited range of variation. In the present chapter we would like to discuss some theoretical issues in cross-language phonetics, based on the data described in chapters 2 - 7. The issues we have in mind are (1) how differently manifested phonetic variation across languages can be treated in terms of the relation between phonetics and phonology, and (2) how generalized phonetic patterns which are commonly found in the six languages can be explained in terms of universal phonetics.

9.2. Relation between phonetic variation and phonological contrasts

The problem here is how the differences of stop voicing in phonetic and acoustic dimensions are related to phonological contrasts in six Asian languages; Japanese, Mandarin Chinese, Korean, Burmese, Thai and Hindi. That is, we would like to examine whether surface phonetic differences of stops are systematically related to "underlying" contrasts, and if they are, how they are derived from phonological contrasts. This is a problem which lies in the relationship between phonology and phonetics. There have been several attempts to examine how the relation should be explored since Chomsky and Halle (1968) made a rather explicit proposal. In their SPE model, they set up two levels of representation; namely systematic phonemic representation and systematic phonetic representation, and phonological rules relate two representations by changing (or deleting/adding) feature values. The systematic phonemic level is the one which represents what an ideal speaker knows about their sound systems and is considered to be more abstract than the phonemic level. The output of phonological rules is the phonetic representation which is considered to contain all idiosyncratic information supplied by a language-particular grammar. The representation then, adjusted by some very low level (universal) articulatory constraints, is converted to actual articulation which is continuous in time. As to the particular case of stop voicing, Keating (1984a:315) argued the relation between the phonological feature [voice] and its phonetic implementation, and mentioned that in a slightly modified SPE model certain cross-language phonetic differences can be expressed as a difference in the realization of the phonological feature contrast. She proposed a phonological feature [voice] to account for phonetic variation and also proposed phonetic categories such as {voiced}, {vl.unasp.}, and {vl.asp}. Although her paper has contributed to clarifying some aspects of the relation between them, she only dealt with the feature [voice] for languages which contrast no more than two phonetic categories.We think it necessary to elaborate further the model to deal with voicing contrasts involving several laryngeal and glottal features.

9.2.1. A Proposal

When we examine the phonetic and acoustic variation of initial-stop voicing in the six languages, we should note that initial-stop voicing varies along several dimensions, and that the variation involves several laryngeal and glottal states. For instance, some languages such as Hindi use another glottal state in addition to the voiced - voiceless glottal state, and voiceless tense stops in Korean use some different glottal state(s) from voiceless aspirated stops in Thai, and the difference is not the degree in the realization of the same glottal feature, but lies in the different aspect of glottal events. Therefore, it is necessary to consider phonetic variation across languages in terms of several features rather than a single feature, as far as the languages under investigation are concerned.

In examining the relation between phonetic variation and phonological contrasts, we would like to use the model with two levels of representation proposed by Chomsky and Halle (1968). We consider as follows: there are several abstract phonological features which can be used to characterize the glottal and laryngeal properties of human vocal organs. Each language selects some of the phonological features which are relevant for characterizing language-particular aspects of initial-stop voicing. At the phonetic level in each language, these selected phonological features are converted into articulatory maneuvers according to language-particular constraints. Based on the acoustic analysis in previous chapters, we consider that there are several articulatory maneuvers which are acoustically realized in such physical scales as VOT, Fo, power spectra and F1. Keating (1984a) considers that the physical scale appropriate to voicing contrasts is VOT scale in the examination of two-category languages such as English and Polish, but as we have seen in Korean and Hindi stops, VOT is not sufficient for distinguishing the voicing contrasts in such languages. For these languages, other laryngeal features than VOT are needed for distinction. The organization between phonetic patterns and phonological contrasts can be schematized as follows:

```
          Phonological Features

                    |
                    | Selection of features
                    | for specific languages
                    |
     L1 Features   L2 Features ... Ln Features
         |             |              |implemented by
         |             |              |language-specific
         |             |              |constraints

      [   ]         [   ]          [   ]
```

So there are three levels of representations in the above organization. The first is an abstract level of phonological features which characterize the voicing contrasts of stops in languages. The second is a level of language-particular phonetic features which characterize voicing categories in each language. The third is an articulatory-maneuver level which is attained by language-particular articulatory maneuvers. Then the third level can be converted into physically continuous speech after being subjected to low-level universal phonetic constraints.

9.2.2. Phonological Features

There has been a great deal of discussion on features in phonology; what are the features for, what are the physical bases for the features, and what are the formal properties and conventions in phonological descriptions, etc. The feature is a basic unit in phonology and is considered to capture a phonological relation among segments, phonological processes, and natural classes on the one hand, and to describe phonetic and phonological contents of segments on the other hand. The feature should, therefore, be identical at both the phonetic and the phonological level. The feature which is most relevant for distinguishing voicing contrasts is [voice], and a binary opposition distinguishes voiced from voiceless sounds. As we mentioned above, however, we need several phonological features to characterize voicing contrasts of initial stops in general, since phonetic variation involves more than variation of a single phonological feature. In order to deal with cross-language voicing contrasts, we would like to propose four phonological features, that is, /VOICE/, /ASPIRATED/, /TENSE/, and /BREATHY/. These features are considered to characterize differences in voicing contrasts of stop consonants. /VOICE/ is for characterizing the basic opposition between voiced and voiceless. /ASPIRATED/ is for another distinction between aspirated and unaspirated. As pointed out in Chapter 8, aspiration requires a carefully adjusted timing between glottal and supralaryngeal events, and should be considered separately from /VOICE/. Ladefoged (1989) discusses some aspects of laryngeal features and sets an independent feature [aspiration] for characterizing aspirated consonants.[1] Although it is well known that Halle and Stevens (1971) propose four laryngeal features to describe the glottal states of several types of voicing contrasts, we will not use them here. Their system is capable of relating pitch differences to voicing ones with a single laryngeal mechanism, but it has been criticized for inappropriateness in the description (Keating, 1984a).[2]

The features used here are abstract in nature and are to have binary values of plus and minus. It is assumed that other specific features can be added to the list of the features to make a more accurate description, if needed. For the present purpose, these four features are sufficient. Each language selects some different combinations of these features, and the selected features are considered to represent the idiosyncratic characteristics of each stop's voicing and to have a classificatory function in each language. For instance, the stops /p, t, k/ in Japanese are specified as [-voice], and the stops /p, t, k/ in Korean as [-voice, -aspirated, -tense]. This implies that voicing contrasts in Japanese stops involve only one feature, whereas those in Korean stops involve three features. That is, /p, t, k/ in Korean are different from /p, t, k/ in Japanese because they require more feature specifications; Korean stops /p, t, k/ are not voiced, not aspirated, and not tensed. These features are variously converted by language-particular articulatory maneuvers into several acoustic dimensions which are physically continuous. The information on the conversion of these features into physical scales is supplied by acoustic analysis data in each language described in previous chapters. The relation between phonetic patterns and phonological contrasts in Japanese stops can be schematized as follows:

---

[1] Ladefoged (1989:62-64) discusses some features relating to laryngeal properties and sets four terminal features; [voice], [glottal aperture], [aspiration] and [pitch] which are dominated by laryngeal node.

[2] Halle and Stevens (1971) propose four features [stiff], [slack], [constricted], and [spread] to characterize voicing contrasts of stops. Keating (1984a:288-289) points out that specification by these features is not adequate in some cases. Although the difference between fully voiced stops and voiceless lax /b̥, d̥, g̊/ is characterized by the degree of slackness, the phonetic fact lies in the difference of the pressure in the oral cavity.

```
            Phonological Features

                    /VOICE/
                    /ASPIRATED/
                    /TENSE/
                    /BREATHY/
                         |
                         |  (selection of features)
                         |
             [+voice]          [-voice]
                 |                 | implemented by language-
                 |                 | particular articulatory
                 |                 | maneuvers
Japanese     [b, d, g]         [p, t, k]
```

The second and third levels in other languages can be shown as follows:

```
Mandarin Chinese

             [-voice]        [-voice]
             [-asp  ]        [+asp  ]
                 |               |
                 | Chinese       | Chinese
             [p, t, k]       [pʰ, tʰ, kʰ]

Korean       [-voice]        [-voice]      [-voice]
             [+tense]        [-tense]      [+tense]
             [-asp  ]        [-asp  ]      [+asp  ]
                 |               |             |
                 | Korean        | Korean      | Korean
             [p*,t*,k*]      [p, t, k]    [pʰ, tʰ, kʰ]
Burmese
             [+voice]         [-voice]     [-voice]
             [-asp  ]         [-asp  ]     [+asp  ]
                 |                |            |
                 | BRM            | BRM        | BRM
             [b, d, g]        [p, t, k]   [pʰ, tʰ, kʰ]


Thai

             [+voice]         [-voice]      [-voice]
             [-asp  ]         [-asp  ]      [+asp  ]
                 |                |             |
                 |  Thai          |  Thai       | Thai
              [b, d]          [p, t, k]    [pʰ, tʰ, kʰ]

Hindi

     [+voice]      [+voice]      [-voice]      [-voice]
     [-asp  ]      [+asp  ]      [-asp  ]      [+asp  ]
     [-bre  ]      [+bre  ]      [-bre  ]      [-bre  ]
         |             |             |             |
         | Hindi       | Hindi       | Hindi       | Hindi
     [b, d̪, g]     [b̤ʱ,d̪̤ʱ,g̤ʱ]  [p, t̪, k]    [pʰ,t̪ʰ,kʰ]
```

Thus, the differences of initial stop voicing in each language are characterized by the selection of phonological features and conversion of those features into physical scales by means of language-particular articulatory maneuvers. Japanese stops are characterized by only the /VOICE/ feature, whereas Korean stops require three features /VOICE, TENSE, ASPIRATED/. Hindi selects three features, but different combinations of features. That is, cross-language phonetic differences can be expressed as a selection of phonological rules and the conversion of these features into articulatory maneuvers by language-particular constraints. As has been mentioned, Japanese stops show a wide range of variation on the VOT continuum. This is attributed to the fact that Japanese stops are specified by only one feature, which can therefore have a wide difference in value. On the other hand, voiceless stops in Burmese and Thai show a limited range of variation in VOT and this is because these languages are specified by a greater number of features. Therefore, we can consider that the number of features specified is relevant for variation in the realization of phonological features.

9.2.3. The Phonetic Bases of Phonological Features

It is understood that phonological features have some physical and concrete bases even if they represent an abstract level of representation (Chomsky and Halle, 1968). Some studies define all features in terms of acoustic characteristics (Jakobson et al., 1952), whereas other studies define them in terms of both acoustic and articulatory correlates (Chomsky and Halle, 1968). We will not go into details of these definitions and motivations described in these studies, but we will examine here some acoustic as well as articulatory correlates for features /VOICE, ASPIRATED, TENSE, BREATHY/ based on the acoustic analysis in previous chapters. The relationship between these features and physical correlates is not straightforward, and it is necessary to specify a phonological feature by several physical and acoustic scales.

Then what are the physical features for /VOICE/? The feature /VOICE/ is related to the state of modal voicing, that is, vocal folds are in the appropriate state for voicing and there should be appropriate transglottal airflow. Acoustically it is represented by the timing dimension between glottal and supralaryngeal events; that is, VOT. Cross-language analysis in six languages indicates that the VOT continuum can be divided into three discrete regions and gives support to Lisker and Abramson's three basic divisions. Although there are some differences between Lisker and Abramson's and the present study in the exact values, there is a basic agreement on the major division for categories. Based on the three divisions of the VOT continuum, /VOICE/ can be defined as a feature indicating whether there is a long voicing lead or not (-50 - -150 ms).

The feature /ASPIRATED/ also represents a relationship between glottal and supralaryngeal events, and involves some timing relation between the two events. The glottis is wide open at around the oral release. It is defined on the VOT continuum and on a certain range of Fo and F1 frequencies. The binary distinction of the feature /ASPIRATED/ may be the presence or absence of a long voicing lag on the VOT continuum (more than 50 ms), a higher range of Fo and onset F1 frequencies at vowel onset, a greater level of intensity and aperiodic noise after the release of the closure.

The feature /TENSE/ is specially set for voiceless tense stops in Korean. Although much has been done to clarify the physiological and acoustic properties of the tense stops, the physical correlates are still not well understood. Hirose et al. (1974) noted that there is a sharp activity for the thyroarytenoid muscle before the oral release. As seen in the acoustic analysis in Korean, the feature /TENSE/ here is defined by the presence or absence of an abrupt movement of the Fo curve from the onset of periodicity to the steady-state.

The feature /BREATHY/ is specially set for breathy voiced stops in Hindi which may be called murmured voice. Articulatorily, this type of stop is defined as having approximated vocal folds and is produced with a moderately open glottis. The vocal folds vibrate without complete closure during the interval called the breathy portion (or murmur) after the release. These glottal states result in a markedly low Fo in a fall-rise curve before and after oral release in acoustic terms. Therefore, the feature /BREATHY/ is acoustically defined by the presence or absence of a distinctive fall-rise pattern of the Fo curve before and after oral release, a lower Fo at vowel onset and aperiodic noise after the release. The relationship between phonological features and physical correlates can be summarized as follows:

Table 9.1 Articulatory and Acoustic Correlates of Phonological Features

Feature	Articulatory Correlate	Acoustic Correlate
VOICE	State of modal voicing	Voicing lead with more than 50 ms, 20 Hz Fo difference between vd. and vl. at vowel onset
ASPIRATED	Wide open glottis at release	Voicing lag with more than 50 ms, higher intensity, and aperiodic noise
TENSE	Sharp activity of laryngeal muscle	Abrupt drop of Fo curve from vowel onset
BREATHY	Vocal-folds vibration without tight closure	Sharp Fall-Rise Fo curve before and after vowel onset, lower Fo, and aperiodic noise

9.2.4. Implementation by language-particular constraints

As a next step, we should consider how selected features are implemented by language-particular constraints. By examining acoustic data in previous chapters, we have seen that initial-stop voicing in each language is differently manifested in several acoustic dimensions and shows language-particular values in the physical scales of the several acoustic dimensions. Consider the case of Japanese stops. Japanese stops have a mean VOT value of -80 ms for voiced stops, and 46 ms for voiceless stops, and show a rather wide range of variation. Implementation of [+voice] in Japanese means that sounds specified as [+voice] are articulatorily maneuvered so that they have the VOT values distributed in the range which centers on the mean values, and they are permitted to have a wide range of variation. [+voice] is mostly realized with a voicing lead, but sometimes with zero or near zero voicing lead within the range of VOT variation. Furthermore, [-voice] is implemented to have a higher Fo at vowel onset than [+voice]. Therefore, implementation of selected features here means assigning some articulatory maneuvers according to language-particular constraints. Although there is no major restriction on the implementation of these features, [-voice, +asp] stops are never realized as having the VOT values of voiced ones. Implementation of the features into physical scales in Japanese stops can be shown as follows:

Japanese

[-voice] ---> Articulatory Maneuvers ---> VOT centering at 46 ms

[+voice] ---> Articulatory Maneuvers ---> VOT centering at -80 ms

Fo difference between two types of
stops should be more than 35 Hz.

The organization proposed here is intended to characterize some cross-language differences and similarities of initial-stop voicings in six languages. We have shown that several phonological features are needed, and cross-language differences and similarities can be expressed in the selection and implementation of these features by language-particular constraints. The model is motivated by the assumption that cross-language phonetic differences can be captured at the language-particular phonetic feature level and the concrete implementation level. It can be considered, therefore, that the shared features in two languages may be differently manifested at the surface phonetic level because of language particular differences.

9.3. Some Generalized Aspects in Phonetic Patterns

Through the examination of cross-language analysis in Chapter 8, it can be noticed that some acoustic and phonetic characteristics exist in many languages or even virtually all languages, whereas others are less common. It can also be noted that although there are subtle differences in VOT, Fo, and F1 and amount of variation, there exist some generalized trends in phonetic patterns. In the present section, we would like to examine such aspects and to attempt to explain them in terms of universal phonetics.

The main phonetic generalizations which can be found in the present study are as follows:

(1) Languages (like Japanese) with two contrasting categories, one voiced and one voiceless, show a wide range of VOT values, whereas languages with more than three categories show a rather limited range of VOT variation.

(2) Voiceless aspirated stops in five of the languages under investigation show a limited range of VOT values and show little variation.

(3) Comparing VOT values in relation to the place of articulation, voiceless velar stops show the longest VOT value among the three places of articulation.

(4) Fo values following voiceless unaspirated and voiceless aspirated stops are considerably higher than those following voiced ones.

(5) In spectral analysis, there is a difference in the overall intensity between voiced and voiceless stops; there is a tendency for voiceless stops to have a higher intensity than voiced ones.

(6) In the analysis of F1 onset frequency, the onset frequency is higher in voiceless aspirated stops than in voiced ones.

9.3.1. Phonetic Categories and VOT ranges

In the measurement of VOT values of initial-stop voicing, it can be seen that there is a difference in the range of VOT variation and there is some correlation between the number of phonetic categories and VOT variation; languages with more voicing categories tend to show rather limited range of variation. As shown in Table 8.1 in Chapter 8, this is specially apparent in Japanese (two-category language) and in Thai and Burmese (three-category languages). We can examine other data of VOT of voicing categories in other two-category languages such as English and Spanish whose voicing contrasts involve a basic distinction between aspirated/unaspirated and voiced/ voiceless. Although a detailed comparison of the data in the present study and these data is difficult because of differences in the experimental design, an overall examination of the range of variation can be made. Lisker and Abramson (1964:392-394) present data on variation of VOT values of English and Spanish stops (bilabial), and they can be shown as follows:

	/b/[3]	/p/
English	-130 - -20	20 - 120 ms
Spanish	-235 - -60	0 - 15

As seen above, the range of variation is extensive in these languages, though the one for voiceless stop /p/ in Spanish is rather limited. The data by other investigators (Williams, 1977) on Spanish stops are in general agreement with the above. Meanwhile, in three-category languages such as Korean, Hardcastle (1973:265) presents the following range for three categories of bilabial stops, and the range of variation is rather limited.

	/p*/	/p/	/pʰ/
Korean	4.5 - 12.0	24.0 - 61.0	65.5 - 111.0

---

[3] It is known that /b,d,g/ in English have two phonetic realizations: voiced [b,d, g] and voiceless [b̥, d̥, g̊] and utilize two timing regions of the VOT continuum. The timing range stated here refers to the one for voiced [b].

Thus, the number of phonetic categories of initial stops seems to constrain the possible range of variation within the VOT continuum. This may be correlated with allophonic variations of stops. Keating et al. (1983) noted that languages with more contrasting categories show a trend to have fewer allophonic variations, and attempted to explain the observations by an aerodynamic model.

9.3.2. Aspiration and VOT variation

In the measurement of VOT values, voiceless aspirated stops in five languages (Mandarin Chinese, Korean, Burmese, Thai, and Hindi) show little variation in the VOT values, as shown in ANOVA analysis. In Chapter 8, we attributed this trend to the fact that aspiration requires carefully adjusted timing events between the articulatory release and the peak of the glottal width, accompanied with a sufficient air-flow. Although it is considered that these timing events may vary across languages and show some language-specific properties, our data reveal that there is little variation in these timing events across languages. Aspiration in general is described in terms of two phonetic grounds; namely the delayed timing in the VOT values (Lisker and Abramson, 1964) and the size of the glottal opening (Kim, 1970). In my opinion, these two grounds should not be considered as separate events but as a unified event, and the glottal timing of the peak opening in relation to articulatory release seems to be important in the production mechanism of aspiration. Although all the five languages have voiceless aspirated stops, there are some differences in voice quality which can not be captured by the difference in the VOT values. The voiceless aspirated stops in Thai are "breathily" aspirated, whereas the ones in Korean are "strongly" aspirated.

9.3.3. VOT on velar stops

With the VOT values in relation to the difference in place of articulation, it was consistently observed in the six languages that the VOT values associated with voiceless velar stops are always longer than those associated with bilabial and alveolar stops. This observation holds true in other languages such as Spanish, Hungarian, Dutch, and English (Lisker and Abramson, 1964). There are mainly two reasons which have been often pointed out in the literature; namely (1) the volume of the supralaryngeal cavity and (2) the motion speed of the tongue body (Hardcastle, 1973; Zue, 1980). The details of these two reasons are already discussed in section 8.3 in Chapter 8.

If we follow the reason (1), we can expect that VOT values will be successively longer as the tongue moves from bilabial to velar position. This kind of predicted VOT difference was found in Mandarin (voiceless unaspirated stops), Burmese (voiceless unaspirated and aspirated stops) and Hindi (voiceless aspirated stops), but was not found in Japanese (voiceless unaspirated stops), Korean (voiceless unaspirated and aspirated stops) and Hindi (voiceless unaspirated stops). That is, unlike the predicted difference in VOT values, /p/ and /t/ showed the same values in Japanese. In Korean, /p/ showed a longer VOT values than /t/ and, in Hindi /p/ and /t/ showed the same VOT values. This suggests that, if based on (1), there isn't any difference in the volume of the supralaryngeal cavity between bilabial and coronal stops (dental and alveolar stops) in Japanese, Korean and Hindi.

9.3.4. Fo and initial stop voicing

As has been noted, there is a difference in the Fo value of the following vowel between voiced and voiceless stops. It is evident in the investigation of these languages that the Fo of a following vowel is higher after voiceless stops than after voiced ones. But Thai showed a rather exceptional case; i.e., there was no difference in Fo value between voiced stops and voiceless unaspirated stops, and we attributed this observation to the historical changes involving these stops. The relationship between Fo and voicing contrasts of initial stops has been discussed in detail in Hombert et al.(1979) and several hypotheses have been proposed. As mentioned above, there are several physiological and aerodynamic factors to influence pitch perturbations. According to Hombert et al. (1979), the most relevant physiological factor among several ones is the larynx height. They hold the view that larynx height can explain the effect on the Fo of the following vowel. It is also reported that there is considerable evidence that the lowering of the larynx facilitates the expansion of the supralaryngeal cavity to maintain voicing (Bell-Berti, 1975). Thus there are several physiological and aerodynamic conditions related to Fo perturbations for voicing contrasts, and they can be shown as follows:

Voiced stop: Lower Fo	Voiceless stop: Higher Fo
Lax larynx	Tense larynx
Lowered larynx	Raised larynx
Lower airflow rate	Higher airflow rate
Distended supra-laryngeal cavity*	Reduced supralaryngeal cavity*

*from Matisoff (1973)

The case of voiced and voiceless nasals in Burmese presents a intriguing problem. As pointed in Chapter 5, there is no need to expand the supralaryngeal cavity in the production of voiced nasals for maintaining voicing, since airflow goes through the nose. Although there is no data on larynx height for voiced and voiceless nasals in Burmese, a question will be raised as to the relevance of larynx lowering if the lowering effect is observed in voiced nasals. That is, voiced stops and voiced nasals showed the same effect of the lowering of Fo, and the explanation of larynx lowering (to sustain voicing) which can be applied for voiced stops may not be applied for voiced nasals, because voicing can be sustained by the airflow through the nose. Then, vertical movement of larynx and the resulting supralaryngeal cavity configurations may not have an effect on pitch perturbation of consonants, and the hypothesis on larynx height proposed by Hombert et al.(1979) may need to be revised. In any case, there are several factors to influence pitch effects in consonants, and at the present stage there seems to be no convincing model to explain the relationship between pitch perturbations and the various factors.

9.3.5. Intensity in Spectral Analysis

In the spectral analysis, it was observed that the intensity level is relatively higher for voiceless stops than it is for the voiced stops. Especially voiceless aspirated stops show a higher intensity than other types of stops. It can be considered that difference in intensity is related to the air-flow rate in the supralaryngeal cavity, and voiceless stops have more pressure build-up behind the closure than the voiced ones. In the production of voiced stops, airflow is maintained so as to generate excitation of vocal-folds vibration, whereas in the production of voiceless stops, it is maintained so as to generate noise turbulence in the supralaryngeal cavity. The difference in the pressure build-up and air-flow rate results in the difference in spectral feature. Although there is some general trend that voiceless stops show a higher intensity than voiced ones in several languages under investigation, there are some cases which are not consistent with this trend.

9.3.6. The Onset Frequency of the First Formant

In the measurement of onset F1 frequency, it was consistently observed that there is a difference in the onset F1 frequency, and the onset frequency is in most cases higher in voiceless stops than it is in voiced stops. Therefore, it can be said that differences in onset of F1 provide a cue for distinguishing voicing categories of stops. The difference can be explained from differences in production. In the production of voiced stops, it is observed that there is voicing during closure and F1 is very low in frequency, and upon release of oral closure F1 rises rapidly from a low frequency region. On the other hand, in the production of voiceless stops, the onset of periodicity is delayed, as seen in VOT values, and there is in most cases an absence of rising F1 transition because of aspiration, and this results in higher onset frequency for voiceless stops. It should be noted, however, that the difference in onset F1 frequency between voiced and voiceless stops may not occur equally for all following vowels. The present study shows much larger differences for utterances containing /a/ or other non-high vowels than for utterances containing /i/ or /u/. This means that stops followed by relatively high F1 frequencies show larger changes in F1 onset frequency than stops followed by low F1 frequencies. From these, it may be said that onset F1 differences are more apparent in distinguishing voicing categories for stops followed by low vowels than those followed by high vowels.

9.4. Summary

We have examined how cross-language differences in initial-stop voicing can be treated in terms of relationships between phonetics and phonology. We proposed a model in which such cross-language differences can be related to the selection of phonological features and implementation of these selected features into language-particular articulatory maneuvers. For examining initial-stop voicing in six languages, four features /VOICE, ASPIRATED, TENSE, BREATHY/ are proposed, from which each language selects the features which are relevant for characterizing voicing contrasts in that language. Some physical bases have been proposed for each feature from articulatory and acoustic points of view. We have shown that implementation of these features can be made in such a way that they produce appropriate acoustic scales which are characteristic of stop consonants in each language.

With respect to some generalized trends in phonetic patterns, we have examined six examples which can be found as an overall trend. It can be hypothesized that these trends are not due to language-specific control, but to universal constraints. Although there is variation in several dimensions across languages, there are some trends which reflect universal principles. These trends are important in understanding physiological and physical aspects in speech production. Cross-language phonetics presents cues as to which aspects in variation are under universal constraints, and which other aspects are under language-specific ones.

Chapter 10

Conclusion

The purpose of the present study has been to examine some cross-language differences of voicing contrast of stop consonants in six Asian languages and to explore the ways to describe some cross-language characteristics. In Chapters 2 - 7, the results of acoustic analysis and their examination were presented. In Chapter 8, some cross-language comparisons of the results were made, and a one-way ANOVA was applied to VOT values in each language to examine the language effect on the laryngeal timing. In Chapter 9, a model was presented to deal with the relation between surface phonetic differences and phonological contrasts of stops across languages.

Based on the results of acoustic analysis, it can be said that the languages under investigation use several acoustic dimensions for distinguishing voicing categories of stops in different ways, and that the "same" or similar sounds represented by the same phonetic symbol in these languages show some language-specific properties, as well as features which are common to these languages. It has become clear that VOT functions for distinguishing voicing categories if they are based on the timing event of glottal and supralaryngeal movements. Laryngeal timing in relation to articulatory release is considered to be a continuum of time values and can be divided into three major regions. The division into three distinct regions has already been noted by Lisker and Abramson (1964), and the present acoustic analysis in the six Asian languages gives further supporting evidence for their basic divisions, though the ranges for major voicing categories show slightly different values; the ones for voiceless unaspirated stops show a wider range of distribution. In three or four category languages using laryngeal features such as glottal tensing or glottal stricture, VOT is not sufficient for distinguishing the categories. In Korean and Hindi, for instance, VOT is unable to distinguish tense from lax, and voiced stops from breathy voiced stops, respectively. This means that VOT is not sufficient if laryngeal features other than glottal timing are involved in the voicing distinction, and acoustic dimensions other than VOT are needed for distinguishing these categories.

Among the three major voicing categories, as seen in one-way ANOVA results, it has been found that voiceless unaspirated stops show a wide range of variability in the articulatory timing events, while voiceless aspirated stops show little variability in these six languages. This means that voiceless unaspirated stops are articulated with language-particular characteristics and have some flexibility in choosing the articulatory timing points in the VOT continuum. The reason that voiceless aspirated stops show little variability is that aspiration requires a carefully adjusted timing of glottal width and articulatory release; i.e., the timing when glottal width reaches its maximum opening must be adjusted with articulatory release, and since there is little flexibility in choosing these timing events in order to make the sounds aspirated, the timing of the glottal events in relation to the release and the glottal width should be considered as a unified event rather than as two separate events.

In voiced and voiceless unaspirated categories, there are some differences in their choice of timing points in VOT continuum in six different languages. These different choices of their points are also considered to be a language-specific aspect. Some languages such as Japanese, choose the category domain for /p, t, k/ with greater variability, while Thai and Burmese choose the domain for /p, t, k/ with less variability. As examined in Chapter 9, there appears to be some correlation between the number of phonetic categories and VOT variation.

With regard to Fo and the curve, it has been demonstrated in the acoustic results that voiced and voiceless distinctions have a different effect on the Fo perturbations of the following vowels, and voiceless stops are generally associated with a higher Fo, while voiced stops are associated with a lower Fo. In Korean, however, all stops are phonemically specified as voiceless, and the distinction between tense and lax stops affects the Fo perturbations, so that voiceless lax stops show the lowest Fo values compared to voiceless tense stops and voiceless aspirated stops. Furthermore, in Thai, there was no significant difference in Fo values between voiced stops and voiceless unaspirated stops. As has been mentioned, voiceless unaspirated stops in Thai are articulated with almost simultaneous glottal closure and release, and the effect of this articulation may weaken the contrast between the two types of stops. Since onset Fo reflects the initial state of the glottis, the observation in Thai suggests that the overall laryngeal conditions in voiceless unaspirated stops at release are comparable to those of voiced stops. Further, it is pointed out (Hombert, et al., 1979:41) that there is a tendency in tone languages to minimize the intrinsic Fo perturbing effect of prevocalic consonants, but this tendency is not always apparent. In Chinese, there was no marked difference in Fo values between the two categories, but in Burmese and Thai (for the contrast between voiceless unaspirated stops and voiceless aspirated stops), there were significant differences between the categories, and the scale of differences in Fo values were not different from those of non-tonal languages such as Korean. For these observations, we can consider as follows: (1) the minimizing tendency depends on the types of tone systems of language, i.e., whether the tone types are contour or register, and whether the tone unit is syllable or word, and (2) the tone effect is considered to appear on the latter portion of tone unit, and the pitch perturbations in the initial portion are less affected by the difference in tone types.

It has also been demonstrated that there is a difference in the Fo curve from the onset to steady-state portion. As expected, voiceless stops tend to show a lowering pattern, and voiced stops a rising one. In Japanese, the effect of voiceless stops is not apparent, and a level pattern was observed. In Korean, a clear-cut distinction between voiceless tense stops and voiceless lax ones is observed, and the tense stops show an abrupt falling Fo curve. Furthermore, in Hindi, Fo curves of the breathy stops show the lowest values and demonstrated a characteristic Fo pattern of fall - rise.

Next, the examination of power spectra in each language reveals some differences in intensity level and spectral characteristics. In Chapter 8, general characteristics of power spectra are described. The spectral characteristics can be examined in the regularity of peak energy distribution, level of intensity, bandwidths and the spectral shape. Although the degree of regularity is difficult to measure, it can be said as a general trend that voiced stops show more regularly distributed energy peaks, while voiceless aspirated stops tend to show less regularly distributed energy peaks. Furthermore, although it is generally known that voiceless stops show a greater articulatory force; i.e., higher rate of airflow, than voiced stops, and hence show a higher intensity level, this trend was not consistently observed, and some languages such as Japanese did not show any marked differences. Further, it was observed, as an overall trend, that the intensity is greater for the voiceless aspirated stops than it is for voiced stops.

With respect to F1 onset frequency, it was found that there is a difference in F1 onset frequency between voiced and voiceless stops; the onset frequency is in most cases higher in voiceless stops than it is in voiced stops. Differences in the onset frequency reflect differences in the speech production of these types of stops. It was also found that the difference in F1 onset frequency is affected by the ones of following vowels, and stops followed by low vowels show greater changes than those followed by high vowels.

Lastly I have discussed how cross-language differences can be treated in the relation between phonetics and phonology, and presented a model in which these differences can be expressed as the ones in the selection of phonological rules and implementation of these selected features. The model has basically three levels; an abstract phonological feature level, a phonetic level, and an articulatory maneuver level, and language-particular properties are implemented at phonetic and articulatory-maneuver levels. I have also discussed some generalized phonetic patterns and pointed out that some variations are under universal constraints, while others are under language-specific ones. Each language under investigation has its own phonetic characteristics, some of which may be similar to those of other languages, but which may have different articulatory targets in the realization of phonetic details. Cross-language study in the present thesis has shown that some aspects of variations are fundamentally important, i.e., the three basic divisions of voicing continuum, and are under universal constraints.

Through the studies of phonetic comparisons of stop consonants in six Asian languages, it can be seen that there are both similarities and differences across languages, and that there are some general phenomena of speech sounds which are considered to be based on universal principles. These similarities and general phenomena may be due to the common physical mechanisms in speech production and to the linguistic functions of a language as a means of communication. On the other hand, some of the differences may be due to the physical differences between the groups of speakers of languages; e.g., differences in the lip characteristics at oral release. The other differences may be due to the different use of articulatory organs and different use of timing relations between glottal and supralaryngeal events. That is, some languages use some particular properties for voicing distinction, while other languages use other properties. It depends on individual languages which aspects of properties are more predominantly used than others for distinction. The similarities and differences across languages reflect some phonetically structured patterns, and further clarification of these patterns will contribute to the understanding of underlying speech mechanisms and of the linguistic functions which language possesses as a self-organizing system.

References

Abercrombie, D. 1967. *Elements of General Phonetics,* Edinburgh University Press.

Abramson, A. S. 1962. The Vowels and Tones of Standard Thai: acoustical measurements and experiments, *International Journal of American Linguistics* 28 (2).

Abramson, A. S. 1972. "Word-final stops in Thai," *Tai Phonetics and Phonology* ed. by J.G. Harris and R.B. Noss, Central Institute of English Language, Office of State Universities, Bangkok.

Bell-Berti, F. 1975. "Control of pharyngeal cavity size for English voiced and voiceless stops," *J.Acoust.Soc.Am.* 57, 456-464.

Bengeurel, A.-P. and Bhatia, K. T. 1980. "Hindi stop consonants: an acoustic and fiberscopic study," *Phonetica* 37, 134-148.

Blumstein, S. E. and Stevens, K. N. 1979. "Acoustic invariance in speech production: Evidence from measurements of the spectral characteristics of stop consonants," *J.Acoust.Soc.Am.* 66, 1001-1017.

Blumstein, S. E. and Stevens, K. N. 1980. "Perceptual invariance and onset spectra for stop consonants," *J.Acoust.Soc.Am.* 67(2), 648-662.

Borden, G. J. and Harris, K. S. 1980. *Speech Science Primer Physiology, Acoustics, and Perception of Speech,* Williams & Wilkins, Baltimore/London.

Burling, R. 1967. *Proto Lolo-Burmese,* Indiana University Research Center in Anthropology, Folklore, and Linguistics, Publication 43, Bloomington.

Caramazza, A. and Yeni-Komshian, G. 1974. "Voice onset time in two French dialects," *J.Phonetics* 2, 239-245.

Catford, J. C. 1964. "Phonation types: the classification of some laryngeal components of speech production," *In Honor of Daniel Jones* ed. by Abercrombie, D. et al., Longmans, 26-37.

Catford, J. C. 1977. *Fundamental Problems in Phonetics,* Edinburgh University Press.

Catford, J. C. 1988. *A Practical Introduction to Phonetics,* Clarendon Press.

Cho, S-B. 1967. *A Phonological Study of Korean, With a Historical Analysis,* ACTA Universitatis Upsaliensis, Uppsala.

Chomsky, N. and Halle, M. 1968. *The Sound Pattern of English,* New York, Harper & Row Publishers.

Connell, B. A., Hogan, J. T. and Rozsypal, A. J. 1983. "Experimental evidence of interaction between tone and intonation in Mandarin Chinese," *J.Phonetics* 11, 337-351.

Cornyn, W. S. and Roop, D. H. 1968. *Beginning Burmese,* Yale University Press, New Haven and London.

Dart, S. N. 1987. "An aerodynamic study of Korean stop consonants: Measurements and modelling," *J.Acoust.Soc.Am.* 81, 138-147.

Dixit, R. P. 1975. *Neuromuscular aspects of laryngeal control: with special reference to Hindi*, Doctoral dissertation. Austin: University of Texas.

Dixit, R. P. 1979. "Inadequacies in phonetic specifications of some laryngeal features: evidence from Hindi," *Current Issues in the Phonetic Sciences* 9(1) ed. by Harry and Patricia Hollien, Amsterdam-John Benjamins B.V., 423-433.

Dixit, R. P. 1987. "Mechanisms for voicing and aspiration: Hindi and other languages compared," *UCLA Working Papers in Phonetics* 67, 49-102.

Dixit, R. P. and MacNeilage, P. F. 1980. "Cricothyroid activity and control of voicing in Hindi stops and affricates," *Phonetica* 37, 397-406.

Dixit, R. P. and Shipp, T. 1985. "Study of subglottal air pressure during Hindi stop consonants," *Phonetica* 42, 53-77.

Docherty, G. J. 1985. "The timing of voicing in stops and fricatives in British English," *Work in Progress*, Dept. of Linguistics, University of Edinburgh 18, 27-41.

Docherty, G. J. 1989. *An Experimental Phonetic Study of the Timing of Voicing in English Obstruents*, Ph.D. Thesis, University of Edinburgh.

Dow, F. M. M. 1972. *An Introduction to the Pronunciation of Chinese*, Edinburgh University Press.

Ewan, W. G. and Krones, R. 1974. "Measuring larynx movement using the thyroumbrometer," *J.Phonetics* 2, 327-335.

Fant, C. G. 1960. *Acoustic Theory of Speech Production*, Mouton, The Hague.

Fischer-Jørgensen, E. and Hutters, B. 1981. "Aspirated stop consonants before low vowels, a problem of delimitation, its causes and consequences," *Annual Report*, Institute of Phonetics, Univ. of Copenhagen.

Flege, J.E. 1979. *Phonetic interference in second language acquisition*, Indiana University dissertation.

Flege, J.E. and Eefting, W. 1986. "Linguistic and developmental effects on the production and perception of stop consonants," *Phonetica* 43, 155-171.

Flege, J.E. and Eefting, W. 1988. "Imitation of a VOT continuum by native speakers of English and Spanish: Evidence for phonetic category formation," *J.Acoust.Soc.Am.* 83(2), 729-740.

Fromkin, V. A. (ed.) 1978. *Tone: A Linguistic Survey*, New York, Academic Press, Inc.

Gandour, J. T. 1974. "Consonant types and tone in Siamese," *J.Phonetics* 2, 337-350.

Gandour, J. T. and Maddieson, I. 1976. "Measuring larynx movement in Standard Thai using the cricothyrometer," *Phonetica* 33, 241-267.

Gandour, J. T. and Weinberg, B. 1980. "On the relationship between vowel height and fundamental frequency: Evidence from esophageal speech," *Phonetica* 37, 344-354.

Haag, W. K. 1979. "An articulatory experiment on voice onset time in German stop consonants," *Phonetica* 36, 169-181.

Haas, M. 1956. "A brief description of Thai with sample texts," *Outline for Types of Linguistic Structure*, University of California, Berkeley, 2nd ed. 1-27.

Halle, M. and Stevens, K. N. 1971. "A note on laryngeal features," *QPR*, Research Laboratory of Electronics, MIT 101, 198-213.

Han, M.S. 1961. *Japanese Phonology:An Analysis Based upon Sound Spectrograms*, unpublished doctoral dissertation, University of Texas, Ann Arbor: University Microfilm.

Han, M. S. and Weitzman, R. S. 1970. "Acoustic features of Korean /P, T, K/, /p, t, k/ and /p^h, t^h, k^h/," *Phonetica* 22, 112-128.

Hardcastle, W. J. 1973. "Some generalizations on the tense-lax distinction in initial stops in Korean," *J. Phonetics* 1, 263-272.

Henderson, E.J.A. 1964. "Marginalia to Siamese phonetics studies," *In Honour of Daniel Jones* ed. by Abercrombie, D. et al., Longmans, 415-424.

Hiranburana, S. 1971. *The Role of Accent in Thai Grammar*, Unpublished doctoral dissertation, University of London (Abstract only).

Hirano, M. and Kakita, Y. 1985. "Cover-body theory of vocal fold vibration," *Speech Science* ed. by Daniloff, R.G., Taylor & Francis, London.

Hirose, H., Lee, C. Y. and Ushijima, T. 1974. "Laryngeal control in Korean stop production," *J. Phonetics* 2, 145-152.

Hirose, H., Niimi, S., Honda, K. and Sawashima, M. 1985. "The relationship between glottal opening and the transglottal pressure difference during consonant production," *Ann. Bull.RILP* 19, 55-64.

Hirose, H., Yoshioka, H. and Niimi, S. 1978. "A cross language study of laryngeal adjustments in consonants production," *Ann.Bull.RILP* 12, 61-71.

Ho, A. T. 1976. "The acoustic variation of Mandarin tones," *Phonetica* 33, 353-367.

Hombert, J.-M. 1978. "Consonant types, vowel quality and tone,"*Tone:A Linguistic Survey*, ed. by V.A. Fromkin, Academic Press. 77-111.

Hombert, J.-M., Ohala, J.J. and Ewan, W.G. 1979. "Phonetic explanations for the development of tones," *Language* 55(1), 37-58.

Howie, J. M. 1974. "On the domain of tone in Mandarin," *Phonetica* 30, 129-148.

Howie, J. M. 1976. *Acoustical Studies of Mandarin Vowels and Tones*, Cambridge University Press.

Huang, R. 1969. *Mandarin Pronunciation*, Hong Kong University Press.

Hutters, B. 1985. "Vocal fold adjustments in aspirated and unaspirated stops in Danish," *Phonetica* 42, 1-24.

Itoh, M., Sasanuma, S., Tatsumi, I.F. and Kobayashi, Y. 1979. "Voice onset time characteristics of apraxia of speech," *Ann.Bull.RILP* 13, 123-132.

Iwata, R. and Hirose, H. 1976. "Fiberoptic acoustic studies of Mandarin stops and affricates," *Ann.Bull.RILP* 10, 47-60.

Iverson, G. K. 1983. "Korean ***s***," *J. Phonetics* 11, 191-200.

Jakobson, R. Fant, G. and Halle, M. 1952. *Preliminaries to Speech Analysis: The Distinctive Features and their Correlates,* Cambridge, MA.

Javkin, H. and Maddieson, I. 1983. "An inverse filtering analysis of Burmese creaky voice," *UCLA Working papers in Phonetics* 57, 115-125.

Kagaya, R. 1974. "A fiberscopic and acoustic study of the Korean stops, affricates and fricatives," *J. Phonetics* 2, 161-180.

Kagaya, R. and Hirose, H. 1975. "Fiberoptic electromyographic and acoustic analyses of Hindi stop consonants," *Ann.Bull.RILP* 9, University of Tokyo, 27-46.

Keating, P. 1984a. "Phonetic and phonological representations of stop consonant voicing," *Language* 60 (2), 286-319.

Keating, P. 1984b. "Universal phonetics and the organization of grammars," *UCLA Working Papers in Phonetics* 59, 35-49.

Keating, P., Linker, W. and Huffman, M. 1983. "Patterns in allophonic distribution for voiced and voiceless stops," *J. Phonetics* 11, 277-290.

Keller, K. 1971. *Instrumental Articulatory Phonetics*, Summer Institute of Linguistics, University of Oklahoma.

Kemp, J. A. 1981. "Early descriptions of nasality," *Towards a History of Phonetics* edited by R. E. Asher and J. A. Henderson, Edinburgh University Press, 35-49.

Kim, C-W. 1965. "On the autonomy of the tensity feature in stop classification (with special reference to Korean stops)," *Word* 21, 339-359.

Kim, C-W. 1970. "A theory of aspiration," *Phonetica* 21, 107-116.

Klatt,D.H. 1975. "Voice onset time, frication, and aspiration in word-initial consonantal clusters," *J.Speech and Hearing Research* 18, 686-706

Ladefoged, P. 1971. *Preliminaries to Linguistic Phonetics*, The University of Chicago Press.

Ladefoged, P. 1975. *A Course in Phonetics* (1st Edition), New York, Harcourt Brace Jovanovich Publishers.

Ladefoged, P. 1978. "Phonetic differences within and across languages," *UCLA Working Papers in Phonetics* 41, 32-40.

Ladefoged, P. 1982. *A Course in Phonetics* (2nd Edition), New York, Harcourt Brace Jovanovich Publishers.

Ladefoged, P. 1982. "The linguistic use of different phonation types," *UCLA Working Papers in Phonetics* 54, 28-39.

Ladefoged, P. 1983. "Cross-linguistic studies of speech production,"*The Production of Speech* ed. by P. F. MacNeilage, Springer-Verlag, 177-188.

Ladefoged, P. 1989. "Representing phonetic structure," *UCLA Working Papers in Phonetics* 73.

Ladefoged, P. and Maddieson, I. 1986. *Some of the Sounds of the World's Languages, UCLA Working Papers in Phonetics* 64.

Laver, J. 1980. *The Phonetic Description of Voice Quality.* Cambridge University Press.

Liberman, A. M., Delattre, P. G. and Cooper, F. S. 1958. "Some cues for the distinction between voiced and voiceless stops in initial position,"*Language and Speech* 1, 153-167.

Lieberman, P. 1970. "Towards a unified phonetic theory," *Linguistic Inquiry* 1, 307-322.

Lieberman, P. and Blumstein, S. E. 1988. *Speech Physiology, Speech Perception, and Acoustic Phonetics*, Cambridge University Press.

Lisker, L. 1975. "Is it VOT or a first formant detector?," *J.Acoust.Soc.Am.* 57, 1547-1551.

Lisker, L. 1980. ""Voicing" in English: A catalogue of acoustic features signaling /b/ versus /p/ in trochees," *Language and Speech* 29(1), 3-11.

Lisker, L. and Abramson, A.S. 1964. "A cross-language study of voicing in initial stops: Acoustical measurements," *Word* 20, 384-422.

Lisker, L. and Abramson, A.S. 1967. "Some effects of context on voice onset time in English stops," *Language and Speech* 10, 1-28.

Lisker, L. and Abramson, A.S. 1970. "The voicing dimension:some experiments in comparative phonetics," *Proceedings of the 6th ICPhC*, 563-567.

Lisker, L. and Abramson, A. S. 1971. "Distinctive features and laryngeal control," *Language* 47(3), 767-785.

Löfqvist, A., Baer, T. and Yoshioka, H. 1981. "Scaling of glottal opening," *Phonetica* 38, 265-276.

Löfqvist, A. and Yoshioka, H. 1981. "Interarticulator programming in obstruent production." *Phonetica* 38, 21-34.

Maddieson, I. 1984. *Patterns of Sounds*, Cambridge University Press.

Maddieson, I. 1984. "The effects on Fo of a voicing distinction in sonorants and their implications for a theory of tonogenesis," *J. Phonetics* 12, 9-15.

Maddieson, I.(ed.) 1987. *UCLA Working Papers in Phonetics, Studies of Phonation Types* 67, UCLA Dept. of Linguistics.

Martin, S. E. 1954. *Essential Japanese*, Charles E. Tuttle Co., Tokyo.

Matisoff, J. A. 1973. "Tonogenesis in south-east Asia," *Southern California Occasional Papers in Linguistics* no.1, 71-96.

McDavid, R. 1945. "Burmese phonemics," *Studies in Linguistics* 3/1, 6-17.

Mohr, B. 1971. "Intrinsic variations in the speech signal," *Phonetica* 23, 65-93.

Nishida, T. 1987. "Higashi Asia ni okeru seichou gengo no hatten," (A development of tone languages in east Asia) *Gekkan Gengo* 16(7), 54-69.(in Japanese)

Norman, J. 1988. *Chinese*, Cambridge University Press.

Ohala, J. J. 1974. "Phonetic explanation in phonology," *Papers from the Parasession on Natural Phonology* Chicago Linguistic Society, 251-274.

Ohala, J. J. 1978. "Production of tone," *Tone: A Linguistic Survey* ed. by Fromkin, V. A., Academic Press, 5-39.

Ohala, J. J. and Jaeger, J. J. 1986. *Experimental Phonology*, Academic Press.

Ohala, M. 1979. "Phonological features of Hindi stops," *South Asian Languages Analysis* 1, 79-87.

Ohala, M. and Ohala, J. 1972. "The problem of aspiration in Hindi phonetics," *Project on Linguistic Analysis* 16, 63-70.

Ohnishi, M. 1987. "Douteki jinkou kougai ni yoru Nihongo shion no kousatsu,"(Some aspects of Japanese consonants by dynamic palatography) *Bulletin of Phonetics* 186, 19-24.(in Japanese)

Okell, J.A. O. 1969. *Reference Grammar of Colloquial Burmese*, 2 vols. Oxford University Press.

Painter, C. 1979. *An Introduction to Instrumental Phonetics*, University Park Press.

Palmer, A. 1969. "Thai tone variants and the language teachers," *Language Learning* 19(3-4), 287-299.

Pettorino, M. and Giannini, A. 1984. *A Study on Aspiration*, Speech Laboratory Report VI, Instituto Universitario Orientale, Napoli.

Pickett, J. M. 1980. *The Sounds of Speech Communication*, University Park Press, Baltimore.

Poon, P. G. and Mateer, C. A. 1985. "A study of VOT in Nepali stop consonants," *Phonetica* 42, 39-69.

Ren, H. 1986. "On the acoustic structure of diphthongal syllables," *UCLA Working Papers in Phonetics* 65.

Rothenberg, M. 1977. "Measurement of airflow in speech," *J.Speech and Hearing Research* 20, 155-176.

Rothenberg, M. and Mahshie, J. 1986. "Induced transglottal pressure variations during voicing," *J.Phonetics* 14, 365-371.

Sawashima, M. 1980. "Some notes on the physiology of speech production," *Language and Speech* 23(1), 31-37.

Sawashima, M. and Hirose, H. 1983. "Laryngeal gestures in speech production," *The Production of Speech* ed. by MacNeilage, P. F., Springer-Verlag, 11-38.

Sawashima, M. and Niimi, S. 1974. "Laryngeal conditions in articulations of Japanese voiceless consonants," *Ann.Bull.RILP* 8, 13-17.

Schiefer, L. 1986. "Fo in the production and perception of breathy stops: evidence from Hindi," *Phonetica* 43, 43-69.

Schiefer, L. 1988. "'Voiced aspirated' or 'breathy voiced': a case for articulatory phonology," Paper presented at the Sixth International Phonology Meeting at Krems.

Schiefer, L. 1989. "Trading relations in the perception of stops and their implications for a phonological theory," Paper presented at the 2nd Conference on Laboratory Phonology at Edinburgh in June, 1989.

Schiefer, L. and Kotten, K. 1983. "Amplitude envelope and the perception of breathy stops in Hindi," *Abstracts of the Tenth International Congress of Phonetic Sciences*, 516.

Shimizu, K. 1977. "Voicing features in the perception and production of stop consonants by Japanese speakers," *Studia Phonologica* XI, 25-34.

Shimizu, K. 1979. "Articulatory effects on speech perception: Adaptation test of voicing feature detectors," *Studia Phonologica* XIII, 1-8.

Skalickova, A. 1960. *Korean Consonants*, Nakladatelstvi Ceskoslovenske Akademie Ved, Praha.

Stewart, J.A. 1955. *A Manual of Colloquial Burmese*, Luzac, London.

Stevens, K. N. and Keyser, S. J. 1989. "Primary features and their enhancement in consonants," *Language* 65(1), 81-106.

Studdert-Kennedy, M. 1974. "The perception of speech," *Current Trends in Linguistics* ed. by T.A.Sebeok, Mouton, The Hague, 2349-2385.

Studdert-Kennedy, M. 1976. "Speech perception," *Contemporary Issues in Experimental Phonetics* ed. by N.J.Lass, Academic Press, 243-293.

Sudo, M. Kiritani, S. and Yoshioka, H. 1982. "An electropalatographic study of Japanese intervocalic [r]," *Ann.Bull.RILP* 16, 21-25.

Suomi, K. 1980. *Voicing in English and Finnish Stops: A Typographical Comparison with an Interlanguage Study of the Two Languages in Contact*, Publication of the Department of Finnish and General Linguistics of the University of Turku.

Summers, W. V. 1988. "F1 structure provides information for final-consonant voicing," *J.Acoust.Soc.Am.* 84(2), 485-492.

Trullinger, R.W. and Emanuel, F.W. 1983. "Airflow characteristics of stop-plosive consonant productions of normal speaking children," *J.Speech & Hearing Research* 26, 202-208.

Tun, U. T. 1982. "Some acoustic properties of tones in Burmese," *Papers in South-East Linguistics* No.8, 77-116.

Yoshioka, H., Löfqvist, A. and Hirose, H. 1982. "Laryngeal adjustments in Japanese voiceless sound production," *J.Phonetics* 10, 1-10.

Yoshioka, H., Löfqvist, A., Hirose, H. and Collier, R. 1986. "How voiceless sound sequences are organized in terms of glottal opening gestures," *Ann.Bul.RILP* 20, 55-67.

Weismer, G. 1979. "Sensitivity of voice-onset time(VOT) measures to certain segmental features in speech production," *J.Phonetics* 7,197-204.

Wenck, G. 1966. *The Phonemics of Japanese Questions and Attempts*, Otto Harrassowitz, Wiesbaden.

Westbury, J.R. and Keating, P.A. 1985. "On the naturalness of stop consonant voicing," *UCLA Working Papers in Phonetics*, 60, 1-19.

Williams, L. 1977. "The voicing contrast in Spanish," *J.Phonetics* 5, 169-184.

Zue, V. 1980. *Acoustic Characteristics of Stop Consonants*. Indiana University Linguistics Club.