Shekgalagari stop contrasts:
A phonetic and phonological study

By: Kemmonye C. Monaka
2001

Submitted to the Department of Phonetics and Linguistics in partial fulfilment of
the requirements for the Degree of Doctor of Philosophy

University College
University of London
2001
For Mom and Dad
Acknowledgements

I am greatly indebted to a number of people who have helped me in the preparation of this dissertation. First and foremost, I wish to express my sincere and deepest gratitude to my supervisor, Dr. Evelyn R. M. Abberton, for her constant and efficient guidance throughout the preparation of this thesis. It is difficult to imagine how this work could have been completed without her invaluable comments, insightful remarks and suggestions. Her thorough and detailed criticism of my work has helped me to learn scholarly responsibility. I also had the benefit of her unfailing support, encouragement, kindness, understanding and everlasting patience, and I needed these in ample proportions, for my stay in the United Kingdom was the most difficult of my entire life.

I am also thankful that she stayed with me both at the start of the research when it looked like I did not know what I came to UCL for, and at the end when I was struggling with exhaustion, almost believing that being a research student will be my eternal occupation. She also survived my earlier written pieces of work!!!!!!! I have never been able to express my heartfelt thanks to her verbally, although I still feel that these written few lines are still somehow inadequate.

I would also like to specially thank Prof. A.J. Fourcin for the times he spent with me in the laboratory teaching me how to interpret the Lx and Gx waveforms when I thought they were just nice pictures, and for reading and providing invaluable comments on Chapter Four, which is based on the laryngograph.

Many thanks also go to the technicians: Steve Nevard (UCL) for performing the acoustic recordings in the anechoic room and for providing information and the related diagram on the recording conditions for Chapters Four and Five, and also for the many computer related help he was kind to give me when I needed it; to Dave Cushing (UCL) for hours spent in the laboratory helping with the computer in order that I may analyse the data; starting with showing me what a computer was and where to switch it on, and to Bernard Howard (SOAS) for EPG recordings and technical assistance, saying ‘this is what I do, apart from living’.

I also wish to thank Dr. Katrina Hayward (SOAS) for reading earlier scripts on Chapter Four and providing useful suggestions and references for the discussion section on that chapter. She also offered invaluable discussion on my EPG data, and offered suggestion on areas to focus on for Chapter Six. I still
cannot comprehend the fact that she passed away during the preparation of this thesis.

I would like to express my sincere thank to the Graduate School and the Department of Phonetics at UCL for assistance in purchasing the EPG artificial palate, which is, naturally, very expensive, being made to fit.

Prof. Stuart Rosen at Wolfson House spent valuable time helping me with the statistical analysis for Chapter Five (VOT). Statistics was to me lots and lots of numbers only. I am very thankful to him for helping me see that numbers can communicate information. Many thanks also to Prof. Andy Faulkner who was on-call for this statistical analysis.

Many thanks also go to Prof. John Harris for suggestions on the presentation of the work in this dissertation. Molly Bennett also answered the many questions I asked in the Departmental Office.

I also like to express my sincere thanks to the people who kindly agreed to participate as the subjects of the acoustic recording: Mosiamisi Mokgolele, Tshiamiso Moumakwa, both of who came over from the University of Reading, and to Dr. Earnest Mosimanegape Nkwane who flew all the way from The Republic of Ireland for this recording. I appreciate the fact that they sacrificed their time for me when they themselves were academically hard pressed.

Many of my friends also made academic work and stay in The United Kingdom bearable. These include Thebe Basebi, Mosiamisi Mokgolele, Happy Masalila, Emmanuel Adjei, Victoria Lemieux, Ann & Richard Omedi, Theresa Davies, Gomotsang Tshoso, Rrenyane Dikole, Keitiretse Kalstveit, and many others.

Finally, much credit and tribute go to members of my family: my parents; Mr & Mrs Simon & Sedijame Monaka, sister; Kebaneilwe Patricia, and four brothers: Keo – Phil, Six – Al, Emma – E’man and Reuben – Gape. Boifang Gabanakgosi is also worth mentioning. Their unfailing support and belief in me kept me hanging on, and their many letters and phone calls kept me from missing home too much.

This project was sponsored by The Commonwealth Scholarship (first year) and The University of Botswana.
Abstract

The aim of this thesis is to investigate the voicing structure, articulatory and co-articulatory properties of stops in Shekgalagari, a South-eastern Bantu language of the Sotho group, which is discussed in details in Chapter Two. Particular reference is given to the Shengologa dialect. The study is based on spectrographic, electrolaryngographic (ELG) and electropalatographic (EPG) techniques. Spectrography and electrolaryngography provide acoustic information, (with ELG providing limited amount of information about the production of the sound) and electropalatography provides articulatory information.

The acoustic results are examined in the light of specified feature theories presented on Chapter Three, and of speech production, and electropalatographic results are examined against some theories of co-articulation, also presented in Chapter Three.

Shekgalagari, like all South-eastern Bantu languages, has traditionally been said to have ejectives in its stop system. In addition to deciding, amongst other things, the category of stop voicing contrast to which this language belongs, this thesis attempts to examine whether Shekgalagari does actually have ejectives. A literature review of Shekgalagari taxonomy is provided in Chapter Two of the thesis, also detailing the situation with ejectives in the language. A detailed analysis of larynx activity by means of both forms of the electrolaryngograph, Lx (which monitors vocal fold vibration) and Gx (which monitors larynx movement) is performed in Chapter Four. Quantitative examination, mainly on Voice Onset Time (VOT) is done in Chapter Five. The results obtained in these chapters are used to argue that these so-called 'ejectives' are, in fact, plain, voiceless unaspirated stops, and not ejectives.

Chapters Four and Five also examine laryngeal characteristics of the other stops in the language, the voiceless aspirated and the voiced stops, and establish Shekgalagari as a three-way contrasting language. Other factors are also investigated: the language-specific phonetic characteristics of its stop system, how the language-specific properties of voicing in Shekgalagari are different from or similar to the phonetic realisation of stops in other languages reported in the literature — e.g. English, Korean, Thai, Burmese, Hausa etc, in order to determine which feature system (presented in Chapter Three) can best describe the phonetic realizations of Shekgalagari stops. The implications of the results obtained in Chapters Four and Five for feature theories are discussed in Chapter Seven.
Articulatory and co-articulatory characteristics of the stops are examined in Chapter Six, as well as their implication for the theories of co-articulation, which, as mentioned above, are also presented in Chapter Three. Chapter Eight concludes by way of summarising the main findings of the thesis and also pointing out some suggestions for future research. Chapter One mainly introduces the dissertation.
# Table of contents

Dedication ............................................................................................................ 2  
Acknowledgements ............................................................................................. 3  
Abstract ................................................................................................................. 5  
Table of contents .................................................................................................... 7  
List of figures .......................................................................................................... 12  
List of tables ............................................................................................................ 29  

**Part one: general background theory** ................................................................. 24

**Chapter 1. Introduction** ..................................................................................... 25

1.1 The purpose of the study: aim and objectives .................................................. 25  
1.2 Scope and structure of the present study ......................................................... 27  
1.3 Statistical analysis ............................................................................................. 29

**Chapter 2. Shekgalagari: Genetic classification and linguistic issues** ............... 30

2.1 Introduction ........................................................................................................ 30  
2.2 Shekgalagari .................................................................................................... 30  
2.2.1 Classification ............................................................................................... 30  
2.2.2 Geographical considerations ....................................................................... 33  
2.3 Phonemic inventories ........................................................................................ 34  
2.3.1 Vowels ....................................................................................................... 34  
2.3.2 Consonants: stops ..................................................................................... 38  
2.4 Prosodic features .............................................................................................. 43  
2.4.1 Tone .......................................................................................................... 43  
2.4.2 Downdrift .................................................................................................. 43  
2.4.3 Length ....................................................................................................... 43  
2.5 Syllable structure .............................................................................................. 44  
2.6 Summary .......................................................................................................... 46
### Chapter 3. Stops: VOT and F0 perturbation, phonological representation and issues of co-articulation

3.1 Introduction .............................................................................47
3.2 Voice onset time & F0 perturbation .......................................47
   3.2.1 Voice onset time ........................................................47
   3.2.2 F0 perturbation ..........................................................48
3.3 Theories on the representation of the voicing structure of stops ..........................................50
   3.3.1 The Sound Pattern of English model (SPE) ..............50
   3.3.2 Lisker & Abramson (L&A) (1964): (VOT) ...............52
   3.3.3 Halle and Stevens (1971) (H&S): laryngeal features .......52
   3.3.4 Keating (1984,1990) ...............................................55
   3.3.5 Kohler (1984) fortis and lenis stops .........................61
   3.3.6 The Element based theory .......................................67
3.4 Studies on stop voicing ...........................................................67
3.5 Ejectives .................................................................................71
   3.5.1 The production of ejectives .......................................71
   3.5.2 Studies on ejectives ....................................................73
   3.5.3 Features of ejectives ..................................................73
      3.5.3.1 Features of ejectives: the problem ..........72
      3.5.3.2 Features of ejectives: general cues ......73
3.6 Co-articulation ........................................................................75
   3.6.1 Accounts of co-articulation .................................77
3.7 Co-articulation resistance .......................................................89
3.8 Summary .................................................................................91

### Part Two: Experimental study of Shekgalagari stops

4.1 Introduction .............................................................................93
4.2 Properties of the Gx, Lx, Sp signals and of the spectrogram: background information .................93
4.3 The Gx: An exploration with Xhosa stops .........................97
Chapter 5. Acoustics: Voice Onset Time

Chapter 6. Articulatory Analysis: Electropalatography
## Table of contents

6.4 Processing....................................................................................173
   6.4.1 Divisions of the palate......................................................173
   6.4.2 Criteria for analysing contact pattern and area...............175
   6.4.3 Criteria of analysing contact pattern and area and the
definition of maximum contact (MC) adopted in
   this study...............................................................................180
   6.4.3.1 Contact pattern and area.............................................180
   6.4.3.2 Maximum contact (MC)...............................................183

6.5 Results..........................................................................................184
   6.5.1 Spatial characteristic patterns of stop articulation:
voice types and place of articulation..............................184
   6.5.1.1(a) Voice types: number of on-electrodes
   at MC...............................................................................184
   6.5.1.1(b) Voice types: lingual-palatal contact profile -
   percentage of on-electrodes............................................186
   6.5.2 Articulatory dynamics of stops......................................190
   6.5.3 Characteristic of co-articulation: spatial changes.........196

6.6 Discussion of results.................................................................208
   6.6.1 The feature-spreading model...........................................209
   6.6.2 The window width theory..............................................210
   6.6.3 Co-articulation resistance theory.................................214

6.7 Summary......................................................................................215

---

**Part three: Representation of Shekgalagari stops**......217

---

**Chapter 7. Some implications of the acoustic investigation for general phonetics and phonology**.................................218

7.1 Introduction..............................................................................218
7.2 The phonetics-phonology correlation..................................218
7.3 Features..................................................................................220
   7.3.1 Features and the phonetic and phonology
connection..............................................................................220
   7.3.2 Distinctive features......................................................221
   7.3.2.1 Distinctive feature geometry: the issue of
   universality..........................................................................221
   7.3.2.2 Distinctive features and their correlates.................222
**Table of contents**

7.3.2.3 Distinctive features and their correlates for three category languages ..................224

7.4 Shekgalagari stops: distinctive features ..........................................................237

7.4.1 Review of acoustic experimental results ....................................................237

7.4.1.1 The spectrogram, speech and Lx waveforms ........................................237

7.4.1.2 Voice onset time ..................................................................................241

7.4.2 Evaluation of feature theories in the light of the Results obtained for Shekgalagari ........241

7.4.2.1 The Sound Pattern of English model (SPE) ..........................................242

7.4.2.2 Lisker & Abramson (L&A) (1964): voice onset time (VOT) ......................244

7.4.2.3 Halle and Stevens (1971) (H&S): laryngeal features ..................................244

7.4.2.4 Keating (1984, 1990) ...........................................................................246

7.4.2.5 Kohler (1984) fortis and lenis stops .......................................................254

7.4.2.6 The Element theory .............................................................................255

**Chapter 8. Conclusion** ......................................................................................258

8.1 Introduction .......................................................................................................258

8.1.1 The aim and objectives of the thesis, data and data processing ......................258

8.2 Summary of results ..........................................................................................258

8.2.1 Electrolaryngographic and VOT investigation ............................................258

8.2.2 Electropalatographic study: articulation and co-articulatory characteristics ....259

8.3 Suggestions for further research ........................................................................262

8.3.1 Acoustics .....................................................................................................262

8.3.2 Articulation ..................................................................................................262

**Appendices** .......................................................................................................264

**Bibliography** ......................................................................................................311
List of figures

Chapter Two

Figure 2.1 ~ The distribution of Shekgalagari speakers in Botswana.................34

Figure 2.2 ~ Shekgalagari vowel chart – Sheboloongwe dialect
             (du Plessis & Kruger 1943)............................................................35

Figure 2.3 ~ Shekgalagari vowel chart (Dickens 1986).....................................36

Figure 2.4 ~ Proposed Shekgalagari vowel chart...............................................38

Chapter Three

Figure 3.1 ~ The Lx shape for intervocalic ejective affricate [ts’]
             in the Hausa word [tsaats’âa] ‘rust’................................................69

Figure 3.2 ~ The initiation of egressive glottalic airstream mechanisms............72

Figure 3.3 ~ The co-articulatory behaviour of Italian consonants in the
             iCi and aCa environments (Farnetani (1997: 389)).........................81

Figure 3.4 ~ The window model of Keating (1990a).........................................83

Figure 3.5 ~ The representation of overlapping movements in the gesture
             Model of co-articulation (Fowler & Saltzman (1993))......................87

Chapter Four

Figure 4.1 ~ The positions of the signals for the waveforms to be described
             in Chapter Four...............................................................................97
List of figures

Figure 4.2 (a) ~ Showing upward deflection of the Gx (and Lx) trace for the Xhosa ejective [p'] in the word [p’enápʼe:ná] ‘throb with pain’, produced in isolation ..............................98

Figure 4.2 (b) ~ Showing upward deflection of the Gx (and Lx) trace for the Xhosa ejective [p'] in the word [p’enápʼe:ná] ‘throb with pain’, produced in a carrier sentence ..................................................99

Figure 4.2 (c) ~ Vowel onset after the release of the ejective [p'] in word initial position in the word [p’a:p’a] ‘porridge’........................100

Figure 4.3 (a) ~ The behaviour of the Gx signal in the Xhosa implosive [ɓ] in the word [ɓōːna] ‘see’, produced in isolation..............101

Figure 4.3 (b) ~ The behaviour of the Gx signal in the production of the implosive [ɓ] in the word [ɓalɛiŋa] ‘run/running’, produced in a frame sentence..................................................102

Figure 4.3 (c) ~ Showing modal voicing on the implosive and the vowel, but irregularity of vocal fold vibration prior to and in the initial portion of the stop in the word [ɓɛi̯tʰa] ‘hit/hitting’, produced in a frame sentence..............103

Figure 4.4 (a) ~ Showing the behaviour of the Gx signal in the production of the Xhosa pulmonic [b] in [bōːka] ‘praise’, produced in isolation........................................................................104

Figure 4.4 (b) ~ Gx behaviour for the stop [b] in [bɛbɛːtʰa] ‘hit/hitting’........................................................................105

Figure 4.4 (c) ~ Characteristics of vowel onset after the articulation of the pulmonic [b] in [bɔːːka] ‘praise’, produced in isolation........106

Figure 4.5 ~ A block diagram of the arrangement of the recording equipment for Chapters Five and Six.................................................109
List of figures

Figure 4.6 (a) ~ Prevoicing in the production of Shekgalagari voiced palatal stop [j], in the word [jāːqa] ‘sojourn’ produced in isolation (speaker EN).................................................................115

Figure 4.6 (b) ~ Showing voice decay in the production of a voiced [b] in the word [bīːna] ‘dance’ produced in isolation (speaker TM).........................................................................116

Figure 4.6 (c) ~ Continuous activity on the Lx and Sp signals for the [b] stop in the word [bīːna] ‘dance’ produced in a carrier sentence (speaker MK).................................................117

Figure 4.6 (d) ~ Voicing decay in the stop [b] in [bīːna] ‘dance’ produced in a carrier sentence (speaker TM)......................118

Figure 4.7 (a) ~ Modal phonation prior to and after the burst for the [t] stop in the word [tīːqa] ‘find find property’, produced in a frame sentence (speaker MK).................................119

Figure 4.7 (b) ~ Aspiration on the Sp signal after the release burst of the word initial palatal stop [c] in the word [cāːca] ‘chase someone away’, produced in isolation (speaker TM)...........................................................................120

Figure 4.8 (a) ~ Modal voicing prior to the stop [tʰ] in the word [tʰāːma] ‘bind’, produced in a frame sentence (speaker MK).............................................................................121

Figure 4.8 (b) ~ Breathy vowel onset after the production of the stop [tʰ] in the word [tʰāːma] ‘bind’, produced in a frame sentence...................................................................................122

Figure 4.9 ~ The movement of the Gx in the production of the voiced stop [d] in the Shekgalagari word [dɔːmo] ‘stupid’, produced in isolation (speaker EN).........................................................123

Figure 4.10 (a) ~ The behaviour of the Gx in the production of the voiceless unaspirated stop [p] in the Shekgalagari word [pʰiːna]
List of figures

'song', produced in a carrier sentence (speaker EN)...........124

Figure 4.10 (b) ~ Gx movement for the voiceless unaspirated stop [k] in the word [ki:ka] ‘mortar’, produced in a carrier sentence (speaker EN). .................................................................125

Figure 4.11 ~ Gx movement for the voiceless aspirated stop [tʰ] in the word [tʰá:ma] ‘mortar’, produced in a carrier sentence (speaker MK). .................................................................126

Figure 4.12 ~ The Lx shape for the intervocalic ejective affricate [ts’] in the Hausa word [tsaats’aa] ‘rust’. Adapted from Lindsey, Hayward and Haruna (1992:519).................................132

Figure 4.13 ~ Gradual build-up of amplitude on the Lx traces after the release of the so-called ‘tense’ stop [pp] in the Korean word [ppul] ‘horn’ (Abberton 1972:74) ......................133

Chapter Five

Figure 5.1 ~ The determination of voice onset time for a voiceless aspirated stop.................................................................145

Figure 5.2 ~ The determination of voice onset time for a voiceless unaspirated stop.................................................................145

Figure 5.3 ~ The determination of voice onset time for a voiced stop........146

Figure 5.4 ~ Exploration of the voicing contrast with VOT as a dependent Feature.................................................................160

Figure 5.5 ~ Exploration of the interaction between vowel quality and voicing contrast.................................................................161

Figure 5.6 ~ Exploration of the interaction between place of articulation and voicing contrast.................................................................163
List of figures

Chapter six

Figure 6.1 ~ The Reading EPG artificial palate.................................172

Figure 6.2 ~ Stylised EPG frame printout...........................................174

Figure 6.3 ~ Divisions of the EPG printout adopted in this study..........175

Figure 6.4 ~ Divisions of the hard palate according to
Dixit & Flege (1990:127).............................................................176

Figure 6.5 (a) ~ The cut-off method – the simple type showing
contact pattern for the Hindi retroflex stop
[t] produced in different vowel contexts and speed style
(Dixit & Flege 1990:220).............................................................177

Figure 6.5 (b) ~ The cut-off method – the simple type showing
contact pattern for the Hindi retroflex stops
[t], [d], [t] and [d] produced in different vowel
contexts and speed style (Dixit & Flege 1990:220)..............177

Figure 6.6 ~ Lingual-palatal contact for the English [kl] sequence
(Gibbon et al. 1993:272).............................................................178

Figure 6.7 ~ The overlap type of the frequency representing method..179

Figure 6.8 ~ A contour example of the average representing method,
showing consonants produced in different vowel contexts
(Recasens 1984: 130).............................................................179

Figure 6.9 (a) ~ Procedure for calculating activation percentage......181

Figure 6.9 (b) ~ Graph showing tongue contact profile at MC for a
hypothetical segment.............................................................182

Figure 6.10 ~ Showing maximum contact (MC) for the stop [t]
in the context /ata/ in Shekgalagari word [matapo:le]
‘potatoes’.............................................................184
List of figures

Figure 6.11 ~ Mean number of on-electrodes at MC for Shekgalagari
voice types across vowel contexts by place of articulation.....185

Figure 6.12 ~ Tongue contact profile: average percentage of on-electrodes
at MC for the stops by voice types and places of
articulation.............................................................................187

Figure 6.13 ~ Percentage of on-electrodes vs. time for the dental stops......192

Figure 6.14 ~ Percentage of on-electrodes vs. time for the palatal stops......193

Figure 6.15 ~ Percentage of on-electrodes vs. time for the velar stops......194

Figure 6.16(a) ~ The influence of flanking vowels on the production of
the voiceless aspirated stop [tʰ]..............................................197

Figure 6.16(b) ~ The influence of flanking vowels on the production of
the voiceless unaspirated stop [t].............................................198

Figure 6.16 (c) ~ The influence of flanking vowels on the production of
the voiced stop [d]................................................................199

Figure 6.17(a) ~ The influence of flanking vowels on the production of
the voiceless aspirated stop [cʰ].................................200

Figure 6.17(b) ~ The influence of flanking vowels on the production of
the voiceless unaspirated stop [c].................................201

Figure 6.17(c) ~ The influence of flanking vowels on the production of
the voiced stop [j]...............................................................202

Figure 6.18 (a) ~ The influence of flanking vowels on the production of
the voiceless aspirated stop [kʰ].................................203

Figure 6.18(b) ~ The influence of flanking vowels on the production of
the voiceless unaspirated stop [k].................................204
List of figures

Figure 6.18(c) ~ The influence of flanking vowels on the production of the voiced stop [g] .................................................................205

Figure 6.19 ~ Window widths for Shekgalagari stops ......................212

Chapter Seven

Figure 7.1 ~ The distinctive feature [± voice] and its correlates ..............223

Figure 7.2 ~ Basic and non-basic correlates for the distinctive, phonological features [± voice] and [± aspiration] in Thai .................................................................227

Figure 7.3(a) ~ Basic and non-basic correlates for the distinctive, phonological feature [± voice] in Korean ...............................................233

Figure 7.3(b) ~ Basic and non-basic correlates for the distinctive, phonological feature [± tense] in Korean ........................................234

Figure 7.3(c) ~ Basic and non-basic correlates for the distinctive, phonological feature [± aspiration] in Korean ..............................234

Figure 7.4(a) ~ Basic and non-basic correlates for the distinctive, phonological features [± voice] in Burmese ........................................236

Figure 7.4(b) ~ Basic and non-basic correlates for the distinctive, phonological features [± aspiration] in Burmese ..............................236

Figure 7.5 ~ The relationship between phonological representation and the three levels of phonetic output (Keating 1990) .......................249

Figure 7.6(a) ~ Representation for VOT lead – for the voiced category of stops (Keating 1990:236) .................................................................250

Figure 7.6(b) ~ Representation for the short VOT lag – for the voiceless unaspirated category of stops (Keating 1990:236) ..............250
List of figures

Figure 7.6(c) ~ Representation for the long VOT lag – for the voiceless aspirated category of stops (Keating 1990:236)...................251

Figure 7.7 ~ The representation of Shekgalagari stops based on Keating’s (1984) theory.......................................................252

Figure 7.8 ~ Shekgalagari stops in terms of Keating’s model of three levels of representation: phonological, categorical phonetic and articulatory and acoustic parametric.......................................................253
List of tables

Chapter Two

Table 2.1 ~ Shekgalagari stop taxonomy van der Merwe & Schapera (1943).................................38
Table 2.2 ~ Shekgalagari stop inventory (du Plessis & Kruger (1968))....39
Table 2.3 ~ Shekgalagari stop inventory (Doke 1954:243).........................40
Table 2.4 ~ Shekgalagari stop inventory: Sheshaga dialect
                 (Doke 1954:243)..............................................40
Table 2.5 ~ Shekgalagari stop inventory (Andersson & Janson 1997:52)....42

Chapter Three

Table 3.1 ~ Segment specification for obstruents under the Halle
                 and Stevens (1971) feature system..............................53
Table 3.2 (a) ~ Summary of the articulatory properties of fortis and
                 lenis stops in the world’s languages (Kohler 1943)........64
Table 3.2 (b) ~ Summary of the articulatory properties of fortis and
                 lenis stops in the world’s languages (Kohler 1943);
                 (continuation of (a)).............................................65
Table 3.3 ~ Element specification for some linguistic systems.............67

Chapter Four

Table 4.1 (a) ~ The classification of the waveform of type 1 category.....112
Table 4.1 (b) ~ The classification of the waveform of type 2 category.....113
Chapter Five

Table 5.1 (a) ~ Average prevoicing and VOT values for the stops..............148

Table 5.1 (b) ~ Effect of place of articulation and vowel context on prevoicing duration, showing the ranking order of factors......149

Table 5.1 (c) ~ T-test of the mean values for the voiced and the voiceless unaspirated stops.................................................................150

Table 5.1 (d) ~ VOT as a function of place of articulation and vowel context for the voiceless unaspirated stops, showing the ranking order of factors on VOT........................................150

Table 5.1 (e) ~ T-test of the mean values for the voiceless aspirated and unaspirated stops.................................................................151

Table 5.1 (f) ~ VOT as a function of place of articulation and vowel context for the voiceless aspirated stops, showing the ranking order of factors on VOT........................................152

Table 5.2 ~ Average prevoicing and VOT values and standard deviations (SD) for the individual speakers..............................................158

Table 5.3 ~ Exploration of voicing contrast: case-processing summary......159.

Table 5.4 ~ Exploration of interaction between vowel quality and voicing contrast: case-processing summary...............................161

Table 5.5 ~ Exploration of interaction between place of articulation and Voicing contrast: case processing summary.........................162

Table 5.6 ~ ANOVA test of between-subject effects..............................164
List of tables

Table 5.7 ~ Element specification for some linguistic systems .....................169

Chapter seven

Table 7.1 ~ Element specification for some linguistic systems.................256
List of Graphs

Graphs

Graph 5.A ~ Mean prevoicing and VOT for Shekgalagari stop contrasts...........153

Graph 5.B ~ The effect of vowel quality on prevoicing and VOT for
       Shekgalagari stops.................................156
Part One:
General Background Theory
1 Introduction

1.1 The purpose of the study: aim and objectives

The purpose of the present study is to examine the voicing structure and articulatory characteristics of Shekgalagari stop consonants with particular reference to the Shengologa dialect. Shekgalagari is a South-eastern Bantu language of the Sotho group of languages and is spoken in Botswana, southern Africa (cf. Chapter Two). The study is primarily phonetic, but phonological aspects are also addressed. The study is based on spectrographic, electrolaryngographic (ELG) and electropalatographic (EPG) techniques (all described in detail in Appendix 1A). Laryngographic analysis, using external monitoring of vocal fold vibration, provides useful and rather detailed information about the nature of vocal fold vibration which may not otherwise be obtained from speech pressure waveforms and spectrograms alone. Furthermore, this information can be processed both qualitatively and quantitatively (cf. Abberton 1976). Electropalatography (EPG) monitors lingual-palatal contacts in time by means of an artificial palate and is helpful in studying the nature of tongue to palate articulation involved in producing speech sounds. The spatio-temporal properties of sounds obtained by means of EPG may be rather difficult to extract from acoustic recordings alone (cf. Hardcastle et al. 1989: 13). This includes, for instance, qualitative analysis of the extent of frontness of an articulation — determined on the basis of the number of the most front row of electrodes of the frame on which contact is registered, and of the degree of closeness of an articulation — the sum of contacts in any given row(s) (cf. Ibid.). Simultaneous processing of the laryngograph waveforms: the Lx and the Gx, the speech waveforms, the spectrogram and electropalatograms is a feature of this study. For spectrographic and laryngographic analysis (cf. Chapters 4 and 5), data comprises homorganic oral stops in the context of word initial CV sequence (in a real word of the form CV:CV) produced by four native speakers. For the EPG study (cf. Chapter 6), real meaningful words of the form CV.CV.CV.CV produced by a native speaker of the language are used, where the highlighted C is the stop under investigation.

The main objectives of the investigation have been to study: the different voice types that Shekgalagari has in its stop systems and to decide the category of stop voicing contrast this language belongs to; the language-specific phonetic characteristics of its stop system; which feature system can best describe the
phonetic realizations of Shekgalagari stops; how the language-specific properties of voicing in Shekgalagari are different from or similar to the phonetic realization of stops in other languages reported in the literature — e.g. English, Korean, Thai, Burmese, Hausa etc; and whether Shekgalagari has 'ejectives' or plain voiceless unaspirated stops.

Consideration of the production of speech sounds must take into account at least the following three factors: the limitations imposed by the articulatory apparatus, the universal phonetic properties of speech sounds and the language-particular (phonetic and/or phonological) properties of sounds. The findings of the experiments in Part Two of this thesis are discussed in Part Three in the light of the above objectives as well as within the framework of speech production (cf. Appendix 3) and of the current theories on the representation of stops discussed below (cf. Chapter 3).

On a more specific level, it is hoped that as a result of this study, a contribution will be made towards the understanding of the nature of voiceless unaspirated stops, the so-called 'ejectives', in Shekgalagari. In particular: the behaviour of the larynx in the production of these stops is investigated by means of the Gx laryngograph, the behaviour of the vocal folds during their production is studied by means of the laryngograph, and acoustic and articulatory properties associated with them are also analysed. As reviewed in Chapter Two (cf. § 2.2), the classification of Shekgalagari within the Sotho group of the South-eastern Bantu languages still remains largely controversial. Perhaps an experimental study of its sound system (stop system in this case) may help to shed light on the nature of its sound system, provide laryngographic, acoustic and articulatory data for perceptual experiments which may be performed in the future, for observation of its sound changes as well as for reconstruction of its earlier sound system, and for the preservation of the nature of its present sound structure in view of the fact that it is fast losing speakers. As pointed out in Chapter Two, in section 2.3 ff. very few descriptions have been made of the language, and virtually none has been objective, quantitative analysis.

On a more general level, this study hopes to make a contribution to the study of consonant voicing, which, after many decades, still remains an intriguing area of study for a number of reasons. One of such reasons relates to the representation of voicing across linguistic systems. In physiological terms, voicing may be characterised by the behaviour of the vocal folds, with 'voice' being defined as the presence of vocal fold vibration and 'voiceless' as the absence of vibration. The delicate adjustment of the vocal folds in executing voice and voicelessness produces different acoustic characteristics (see § 3.2).
Phonologically, voicing is defined more abstractly, with the (distinctive) feature [± voice] which has acoustic and articulatory correlates, the presence or absence of vocal fold vibration being only one of the (articulatory) correlates. The mapping between the phonetic articulatory and acoustic properties of stop consonants and the phonological representation of these stops using the feature [± voice] is not a straightforward one-to-one relationship across languages. For example, the 'voiced' stops [b d g] are always physically produced with vocal fold vibration throughout the articulatory closure in all contexts in languages like Ibibio and French, but this may not always be the case in a language like English, where in utterance initial position, the vibration of the vocal folds may not occur throughout the entire period of stop closure. Similarly, the 'voiceless' stops [p t k] are normally produced without vocal fold vibration in French and Polish. In English, however, these sounds may occur as a realisation of the 'voiced' [b d g] series in utterance initial position. Inspite of these apparent not straightforward relationships, it is still common to broadly represent 'voiced' stops with the symbols /b d g/ and 'voiceless' stops with the symbols /p t k/ in various languages, although the physical activities involved in their execution are different in different languages. A study in the voicing structure of stops in Shekgalagari is, therefore, a fruitful area of research in that it contributes towards the understanding of the voicing structure of stops in the world's languages.

1.2 Scope and structure of the present study

As indicated above, the scope of this investigation is to examine voicing contrasts of oral stops only, as well as their articulatory characteristics in Shekgalagari. The thesis consists of three parts and has the following structure.

Part One consists of three chapters providing general background information on a number of areas. This part attempts to establish a context for the experiments to be performed in this study and the theoretical issues to be discussed afterwards.

Chapter One is an introduction to the thesis. It states the aim of the thesis, outlines the main objectives and presents the scope of the study. It also discusses the study's contribution to general linguistics.

Chapter Two focuses on Shekgalagari, a Bantu language spoken in Botswana, Southern Africa. Detailed discussion of: its controversial classification within South-eastern Bantu languages, its phonemic inventories and its prosodic features is provided.
Chapter Three focuses on stop consonants, the types of sounds to be analysed in this study. Articulatory and acoustic properties of stop consonants are discussed. A number of different theoretical concepts on the representations of different voice types of stops are presented, as well as theories accounting for co-articulation of stops as a function of flanking vowels. It also mentions work that has already been done on stop consonants in different languages.

Part Two presents the experiments performed in this study in the following way.

Chapter Four provides qualitative description and classification of the waveforms obtained for Shekgalagari stops based on visual analysis of the properties of Gx, Lx, Sp and spectrographic signals for the different voice types, and the classification on the general similarities and differences emerging from the visual inspection of the signals.

Chapter Five presents an experiment on variation in the duration parameter of VOT for the stops.

Chapter six presents an experiment on the articulatory properties of Shekgalagari stops by means of EPG, as well as addressing issues of co-articulation.

Each of the above chapters in Part Two reports on the findings of each stated parameter and proceeds to discuss the results in the light of the hypothesis postulated there, theories of stop voicing and what has been reported in the literature.

Part Three reports on the implications of the findings of the experiments performed in Part Two for theories of stop representation and accounts of co-articulation (cf. Chapter 3), and it also concludes the thesis. The chapters on this sections are as follows.

Chapter Seven discusses the language-specific properties of voicing in Shekgalagari stops in the light of: the findings of the experiments performed in Chapters Four and Five. Cross-language investigation is also addressed.

Chapter Eight is the conclusion of the thesis in the form of a summary of the study, focusing primarily on the findings and achievements of the thesis and pointing out suggestions for further research.
1.3 Statistical analysis

The statistical analysis applied to the data in this study was performed using the Statistical Package for the Social Sciences (SPSS). This package was used for the production of bar charts, the application of descriptive statistics, e.g. mean, standard deviation etc., parametric tests: $t$-test for independent or correlated samples, analysis of variance (ANOVA).
2. Shekgalagari\(^1\): Genetic classification and linguistic issues

2.1 Introduction

This chapter mainly gives information about Shekgalagari, a Bantu language spoken in Botswana, Southern Africa. The rest of the chapter proceeds as follows. Section 2.2 reports on the classification of Shekgalagari within the Sotho group of languages with particular focus on its relationship with Sesotho and Setswana.\(^2\) Section 2.2.2 deals with geographical considerations, showing areas where the language is currently spoken in Botswana. Section 2.3 focuses on phonemic inventories of Shekgalagari and Section 2.4 with its prosodic features. Syllable structure is addressed in section 2.5. Section 2.6 summarizes the chapter.

2.2 Shekgalagari

2.2.1 Classification

Shekgalagari is genetically affiliated to the south-eastern zone of Bantu languages, and specifically to the Sotho group. There is, however, a controversy regarding its position in relation to Sesotho and/or Setswana. It seems that the main problem is whether Shekgalagari should be described as an independent Sotho cluster or as a dialect of Sesotho or Setswana clusters. This is clearly reflected in Doke (1954)'s discrepancies when making reference to Shekgalagari, as shown in the following quotations.

\(^1\) In this study, we follow Andersson and Janson (1997) in adopting the method of attaching prefixes to the names of groups of people and their languages, for example, Bakgalagari and Shekgalagari (pronounced [\(\text{\textipa{\text{\texttimes}}\text{\textipa{\text{\texttimes}}}}\text{alaxari}\)])). The prefixes used here are proper to the language concerned (i.e. as used by native speakers). These prefixes have often been left out in the past, particularly in studies conducted by Europeans, and hence the common use of Tswana, Kgalagadi, Kalanga etc. in the literature. Prefixes will not be used for the discussion of language clusters and groups, e.g. Sotho cluster, Nguni group, etc.

\(^2\) Information regarding the classification of Southern Bantu languages, Southeastern group of languages and the Sotho language cluster of languages, as well as some phonetic and phonological aspects of South-eastern Bantu languages may be found in Appendix 2B.
Chapter Two. Shekgalagadi: Genetic classification and linguistic issues

Strangely enough the Lozi and Kgalagadi dialects of Sotho differ from the main groups in employing only five-vowel phonemes (Doke 1954: 28).

Other languages, though definitely belonging to certain groups or clusters, have developed or influenced away from the norm in such a way that they tend to stand apart from other members of their grouping; some may even be survivors of a much earlier form; among these are such a dialect of Karanga as Mhari ... and perhaps the Kgalagadi dialect of Tswana (Doke 1954:20).

Kgalagadi is sufficiently distinct from Tswana to warrant a separate classification (Doke 1954:24).

Kgalagadi ... has been described as 'one of the mixed dialects of Sotho.' It holds certain striking characteristics contrasting with the known literary forms in the group (i.e. the Sotho group). These with what are found in other extreme Sotho types ... such as Lobedu and Pai, seem to point to Kgalagadi as containing archaic elements of what might be called Proto-Sotho (Doke 1954 243). (All emphasis mine).

This discrepancy persists in other studies also. Schapera (1938:161) writes with regards to Shekgalagari: 'Basically it belongs to the same language as Sekwena and other dialects’ (i.e. other dialects of Setswana). He thus classifies Shekgalagari as a dialect of Setswana. In later work, however, his position on this classification changes. Schapera (1943:3), notes that Shekgalagari 'differs as much from [Setswana] as it does from either Pedi or Southern Sotho ... There is, therefore, no justification for classifying Sekgalagadi as a dialect of Tswana.' Cole (1955) also appears to advocate the autonomy of Shekgalagari when he points out that:

... there are probably some thousands among non-Tswana peoples of Bechuanaland Protectorate, who use Tswana as a second language. These include the Kalanga, Herero and Bandieru (Damara) Yei (Koba), Gova (Mbukushu), Ikwahani (Subia), Kgalagadi, Sarwa Bushmen, etc. (Cole 1955:xv)(emphasis mine).

Campbell (1979) and Neumann (1990) also support a separate classification of Shekgalagari as an independent language. Neumann (1990) conducted a pairwise comparison of word lists obtained from Bakgalagari informants and the Batawana

---

3 For the work of Neumann (1990), which originally written in Afrikaans, I am indebted to Andersson and Janson (1997).
(who speak Setawana, a dialect of Setswana), and came to the conclusion that Shekgalagari and Setswana could best be regarded as two distinct languages. Neumann (1990) follows Hansford et al. (1976)'s criterion of distinguishing between Nigerian languages and dialects.

According to Hansford et al. (1976), one way of making a distinction between a language and a dialect within a language group or cluster is based on analysing the cognacy rates between the relevant linguistic systems by way of comparing word lists. Based on the percentage of the cognacy rate, the relevant linguistic systems could be regarded either as dialects of a language or as distinct languages. In general, if the cognacy rate between the two systems is 80% and above, they constitute dialects of a language, and a score in the region of 70% to 80% classifies the systems as independent languages. In Neumann's analysis of Shekgalagari and Setawana (and hence Setswana), 'the facts clearly point to these being two different languages rather than dialects' (Andersson & Janson 1997:50).

Mutual intelligibility, another linguistic criterion thought to be correlated to that of cognacy rate (both attempt to determine the extent of mutual connection between linguistic systems (Hansford et al. 1976, Connell 1991)), has often been proposed in dealing with the issue of language/dialect distinction. But there are many problems inherent in the mutual intelligibility criterion. One such problem relates to the issue of determining 'whether or not two lexical items should be considered cognate, or what constitutes mutual intelligibility' (Connell 1991:9). An off-shoot of this problem is the failure of this approach to adequately address the issue of why some languages are generally considered to be distinct from each other despite having a good degree of mutual intelligibility; examples of such linguistic systems being Setswana and Sesotho; and Swedish and Norwegian (Andersson & Janson 1997). Connell (1991:9) argues that because of these problems, the technique of mutual intelligibility could best be regarded as arbitrary as far as distinguishing a language from a dialect is concerned.

Other problems are those associated with the political and social factors, the historical background of the language(s) and the people, and the perception(s) of the language speakers themselves.

Thus, in order to reasonably answer the rather complex issue of the classification of Shekgalagari in relation to (in particular) Setswana, both the linguistic considerations and non-linguistic aspects mentioned above have to be addressed. This would no doubt go beyond the scope of the present study, and remains a topic of future research. The classification of Shekgalagari in many ways thus still remains largely controversial. But, on the whole, most of the above
discussed works (Schapera (1943:3); Cole (1955:xv); Campbell (1979); Neumann (1990) appear to advocate the autonomy of Shekgalagari — as an independent cluster of the Sotho group of languages, quite distinct from Setswana and Sesotho, and with its own dialect cluster. Some of the dialects of Shekgalagari often named include Shelala, Sheshaga, Shebolaongwe, Shedjegwana, Shengologa, etc. (Andersson & Janson 1997, Cole 1955). This is the position I also adopt.

2.2.2 Geographical considerations

In Botswana, Shekgalagari is spoken in some parts of the Kweneng, Ngamiland and Southern districts, with the majority of speakers concentrated in the Kgalagadi district and its outskirts. Figure 2.1 shows the distribution of Shekgalagari in Botswana. Shengologa, which is the main dialect analysed in this study, is spoken in Hukuntsi village and in the nearby villages of Tshane, and Lokgwabe (not shown on the map).

---

4 In my opinion, most of Shekgalagari speakers are permanently resident in Maun, Ghanzi, Hukuntsi, Molepolole, Kanye and Lobatse regions shown on Figure 2.1 from Andersson & Janson (1997). More often than not, those who may be found in Gaborone, Mochudi and the area between Mochudi and Francistown may have possibly migrated there in search of work opportunities.
Chapter Two. Shekgalagari: Genetic classification and linguistic issues

2.3 Phonemic inventories

2.3.1 The vowels

Description of Shekgalagari vowel inventories in the meagre literature on the language has been both inconsistent and inconclusive. The earliest studies on Shekgalagari vowels are those of van der Merwe and Schapera (1943), and du Plessis and Kruger (1968). In these works the vowels are treated phonemically only. Dickens (1986) describes van der Merwe and Schapera’s treatment of Shekgalagari vowels in the following way:

5 The researchers discussed in this study analysed various dialects of Shekgalagari as follows: du Plessis and Kruger (1968): Shebolaongwe, van der Merwe and Schapera (1943): Shebolaongwe, Dickens (1986): Sheshaga, Shekoma and Shebolaongwe. It is not indicated which dialect Doke (1954) was looking into.
... the Qhalaxarzi vowels are merely listed without phonetic labelling at all, and are recorded as corresponding exactly with Tswana vowels ... The treatment of vowels is phonemic rather than phonetic, giving only an idea of the actual pronunciation (Dickens 1986:32).

With regard to du Plessis and Kruger (1968)'s treatment of Shekgalagari vowels, Dickens (1986) gives the following points:

(i) there are no high mid vowels (the symbols /e/ and /ø/ are used, but in the chart they are placed in lowered high position);
(ii) the low mid vowels are given as corresponding exactly to their cardinal equivalents ...;
(iii) the low vowel is shown to be precisely central instead of somewhat fronted;
(iv) the high back vowel is not shown as somewhat centralised.

(Dickens 1986:32-33)

A possible vowel chart for du Plessis and Kruger’s Shekgalagari would be as shown in figure 2.2.

![Figure 2.2: Shekgalagari vowel chart - Shebolaongwe dialect (du Plessis & Kruger (1968))](image)

From figure 2.2 it can be seen that although the symbols /e/ and /ø/ are used for the mid-high vowels, they are somewhat lower than their Cardinal counterparts. The mid-low vowels /e/ and /ø/ correspond exactly to their Cardinal cognates. The low vowel /a/ is central, and the high back vowel is slightly lower than Cardinal 8.
All the vowels are also reported to be nasalised irrespective of context (cf. Dickens 1986:33). In this thesis, however, this is considered to be an inaccurate description of the vowels.

Doke (1954)'s Shekgalagari vowel inventory is not clear, since he postulates two different inventories. In the first one, Shekgalagari, unlike other language clusters in the Sotho group which employ seven vowel phonemes, has a five vowel system and seven allophones like Nguni: a, i, u, e(e) and o(o) (Doke, 1954: 26). However, following van der Merwe and Schapera (1943), he gives the following seven vowel inventory with nine allophones which is typical of Sotho languages: i, i̯, e(e), a, o(o), u and u̯ (where i̯ and u̯ are not allophones of i and u respectively, but more closed front and back vowels. i̯ is an allophone of e and o respectively).

Dickens (1986)\(^6\) gives the following vowel chart in figure 2.3 for Shekgalagari.

\[\begin{array}{c|c|c|c|c|c|c|c}
\hline
 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\
\hline
[i] & i \\
[e] & e \\
[e] & e \\
\hline
[\alpha] & a \\
\hline
[u] & u \\
\hline
\end{array}\]

From figure 2.3 above, it can be observed that in Dickens (1986)'s Shekgalagari vowel chart, the mid-high vowel [e] and [o] correspond exactly to their Cardinal cognates, the mid-low vowels [ε] and [ɔ] are slightly raised with respect to their Cardinal cognates, the open vowel [a] is slightly fronted (compare with that of du Plessis and Kruger [P&K](1968) above, which is more centralised) and the close back vowel [u] is slightly forward than Cardinal 8 (contrast with that of P&K in figure 4.2 above).

Dickens maintains this seven vowel phoneme system for Shekgalagari in a later work (Dickens 1987) and gives the vowel system in the following way.

\^6 Dickens (1986) points out that his analysis of Shekgalagari vowels (and presumably those of van der Merwe and Schapera (1943) and du Plessis and Kruger (1968)) is not based on acoustic or any instrumental investigation. The analysis of vowel quality is subjectively evaluated against the Cardinal cognates, and the articulatory terms used do not necessarily represent exact articulatory configurations (Dickens 1986:18).
It appears that Shekgalagari (at least the Shengologa dialect) does have a seven vowel system like other Sotho languages as indicated by Dickens (1987, see also Andersson & Janson 1997:52) above. These are exemplified in the words below.

\[
\begin{align*}
/m&-\text{[pi:na]} & & 'song' \\
/e&-\text{[leima]} & & 'plough' \\
/e&-\text{[leana]} & & 'to spoil - a child for instance' \\
/a&-\text{[laila]} & & 'sleep' \\
/o&-\text{[noma]} & & 'gain weight' \\
/o&-\text{[noma]} & & 'a large number of men' \\
/u&-\text{[buita]} & & 'to break'
\end{align*}
\]

However, unlike other Sotho languages, the mid-front and mid-back vowels /e/ and /o/ do not constitute single phonemes with two varieties of allophones [e], [ɛ] and [o], [ɔ] respectively. Both seem to function distinctively and can form minimal pairs. The low vowel /a/ sounds slightly fronted.

A possible vowel chart of the position of Shekgalagari vowels relative to the Cardinal vowel system could be sketched as in figure 2.4.
In conclusion, Shekgalagari appears to exhibit a seven vowel system, with the low mid-front and mid-back vowels, /e/ and /o/ respectively, constituting individual, single vowels capable of operating contrastively and forming minimal pairs. This analysis, however, is not correlated with acoustic investigation, and this aspect still remains a subject for future research.

2.3.2 The consonants: stops

The earliest works on Shekgalagari stop consonants are those of van der Merwe and Schapera (1943) and du Plessis and Kruger (1968). Their respective inventories are presented in tables 2.1 and 2.2 respectively.

<table>
<thead>
<tr>
<th>Stops</th>
<th>Voiced</th>
<th>b</th>
<th>d</th>
<th>j</th>
<th>g</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vls.unas</td>
<td>p</td>
<td>t</td>
<td>c</td>
<td>k</td>
<td></td>
</tr>
</tbody>
</table>
| Vls.eject | kʔ seventh sound, according to Dickens (1986, 1987) has been incorrectly recorded by van der Merwe and Schapera (1943:7 ff) for [q']. For uvular
In these studies, it can be observed that there is an inconsistency with respect to ejectives. For instance, van der Merwe and Schapera (1943), on the one hand, record only one ejective for the stops, *[k?]*, which Dickens (1986:27,30; 1987:302,304) argues has been incorrectly recorded for *[q’]* (this point is discussed briefly below), du Plessis and Kruger (1968), on the other hand, record all the voiceless unaspirated forms (including affricates which are not indicated here) as ejectives, (also recording the ‘velar ejective’ stop — their *[kx?]* (not shown on the table) which Dickens argues is also *[q’]* (Dickens 1986:26, 1987:302)).

For most of the articulatory places: bilabial, alveolar/dental and palatal, the authors appear to postulate a three-way contrast, which, in the case of Van der Merwe and Schapera (1943) comprise the voiced stops, radicals (i.e. the voiceless unaspirated stops) and aspirated stops. For these authors, the velar place manifests a four way contrast, incorporating the so-called ‘ejective’ *[k?]*. Du Plessis and Kruger (1968)’s three-way contrast comprise the voiced and aspirated stops also, but the third distinctive group consists of ejectives rather than radicals. Thus van der Merwe and Schapera (1943), on the one hand, appear to suggest that ejectives are not a major feature of Shekgalagari languages, whilst for du Plessis and Kruger (1968), on the other, ejectives are a major feature of the language’s stop (and affricates) systems, and even have a distinctive function.


---

See page 41.
According to Table 2.3, ejectives in Shekgalagari, unlike in other south-eastern group of Bantu languages (ejective stops have been described as 'a typical south-eastern feature' (Doke 1954:31,33)), are indeed a marginal feature of stops, displaying only one ejective: namely, the velar plosive [k']. Thus, Doke (1954:243), like Van der Merwe and Schapera (1943) above, seems to suggest a three-way contrast for most of the places of articulation: bilabial, alveolar and palatal; viz. the radical stops, the aspirated stops and the voiced stops. Only the velar place shows a four way contrast, consisting of the so-called 'ejective' [k']. That Doke (1954)'s taxonomy so obviously takes for granted the presence of radicals in this language rather than 'ejectives' is somewhat surprising, since he has not been certain whether Shekgalagari has these sounds. In Doke (1954:31) it is clearly pointed out that 'Radicals ... are found in Shona, but not in South-eastern Bantu, except in the insufficiently investigated Kgalagadi'.

Dickens' (1986, 1987) consonantal table, presented in Table 2.4 does not record any ejectives for the Sheshaga dialect. According to him, these sounds were 'always produced pulmonically except on a few occasions when 'ejectives' were used meta-linguistically' — for instance, emphasising with an ejection when a voiceless unaspirated sound was incorrectly imitated as aspirated (Dickens 1986:26, 1987:302). He goes on to say that a hypothesis with regard to 'ejectives' is that they may be stored mentally as 'ejectives' but only realized as such when particular attention is paid to the sound during articulation.
In table 2.4, Sheshaga is displayed as a three-way category system with radical sounds, aspirated sounds and voiced sounds, for stops.

Apart from the fact the Dickens records no ejectives at all, it is worth noting the discrepancy with regard to the places of articulation. Dickens (1986) records dentals where Doke (1954) records alveolars for the same sounds. Dickens (1986) also records the uvular stop [q] which does not appear in Doke’s taxonomy.

Another point of discrepancy relates to the presence or absence of uvulars in Shekgalagari. Apart from Dickens (1986, 1987), no other researcher so far records uvulars in their inventory. Dickens (1986, 1987), however, argues that ‘although no uvulars are recorded in the two earlier studies (i.e. those of van der Merwe and Schapera (M&S) and du Plessis and Kruger (P&K) - Doke also falls in this category), the phonemic distinctions between ... velars and uvulars have been made’ (Dickens 1986:31, 1987:304). Consider the following illustration made by Dickens (D) (1986:31, 1987:304).

(3)

<table>
<thead>
<tr>
<th>D&amp;K</th>
<th>M&amp;S</th>
<th>D</th>
<th>Intended sound</th>
</tr>
</thead>
<tbody>
<tr>
<td>[k]</td>
<td>[k]</td>
<td>[k]</td>
<td>voiceless unaspirated velar stop</td>
</tr>
<tr>
<td>[kʰ]</td>
<td>[kʰ]</td>
<td>[kʰ]</td>
<td>voiceless aspirated velar stop</td>
</tr>
<tr>
<td>[ŋ]</td>
<td>[ŋ]</td>
<td>[ŋ] or [ŋ]</td>
<td>velar nasal</td>
</tr>
<tr>
<td>[*kʔ]</td>
<td>[kxʔ]</td>
<td>[q]</td>
<td>voiceless uvular stop</td>
</tr>
<tr>
<td>[kg]</td>
<td>[kxʰ]</td>
<td>[qʰ] or [qxʰ]</td>
<td>voiceless asp. uvular</td>
</tr>
<tr>
<td>stop/affricate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[ɡ]</td>
<td>[x]</td>
<td>[x]</td>
<td>voiceless uvular fricative</td>
</tr>
<tr>
<td>[ŋɡ]</td>
<td>[ŋ]</td>
<td>[N]</td>
<td>uvular nasal</td>
</tr>
</tbody>
</table>

Dickens, using palatography and visual observation of the places of articulation, points out the following phonetic information: (1) Dentals — the tongue tip articulates with the upper teeth or may protrude between them, and the front part of the tongue touches the alveolar ridge. (2) Alveolar — the tongue tip articulates with the alveolar ridge (Dickens 1986:13, 1987:298-299).

“Intended Sounds” notes have been added by me.

An aspirated uvular stop [qʰ] tends to neutralize to the aspirated affricate [qxʰ].
The latest work on Shekgalagari stop inventory is that of Andersson and Janson (1997:52). This is shown in table 2.5.

<table>
<thead>
<tr>
<th>Stops</th>
<th>Voiced</th>
<th>b</th>
<th>d</th>
<th>j</th>
<th>g</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Vls.unas</td>
<td>p</td>
<td>l</td>
<td>c</td>
<td>k</td>
</tr>
<tr>
<td></td>
<td>Asp.</td>
<td>pʰ</td>
<td>lʰ</td>
<td>cʰ</td>
<td>kʰ</td>
</tr>
</tbody>
</table>

Table 2.5 Shekgalagari stop inventory (Andersson and Janson (1997:52).)

Like Dickens (1986, 1987), Andersson and Janson (1997) record dentals instead of alveolars. There are no ejectives recorded at all. Their taxonomy also includes uvulars. The presence of uvulars in Shekgalagari may perhaps be further buttressed by the fact that they can form minimal pairs with velars as shown in (4).

(4)

[kɔːna] ‘to bend something’
[qɔːna] ‘to not want any more, especially of food’
[kɔːba] ‘to traditionally knit logs on top of the walls of a hut in order to thatch it.’
[qoːba] ‘to bend, often touching one’s toes’

It is hoped that the present study, using combined electropalatography and electrolaryngography, will make a contribution to clarifying the situation of ejectives in Shekgalagari (with particular reference to the Shengologa dialect), and to contribute also to the issue of place of articulation for some stop sounds in the language.

In summary, that Shekgalagari has voiced and aspirated stops does not appear to be disputable in the literature. ‘Ejectives’, it appears, still remain controversial, although later studies appear to suggest that Shekgalagari does not have ejective stops. There have also been some inconsistencies in some of the previous studies with respect to uvular stops. But their presence in Shekgalagari has been pointed out convincingly (Dickens 1986, 1987; also the subsequent

---

13 There is no specific dialect mentioned to which particular attention is being paid. It appears that this analysis is general and not necessarily phonetic.
studies Andersson & Janson 1997), and also by the illustration in (4) above, which indicates that they can form minimal pairs with velar stops.

2.4 Prosodic features.

2.4.1 Tone.

Shekgalagari has two phonemic tones: namely; high (H) and low (L). In linguistic writing, a high tone is normally marked with an acute sign (') and a low tone is usually left unmarked. Tone in Shekgalagari is semantically significant. A change in the tone pattern of the following words produces a complete difference in meaning in each case.

(5)

\[
\begin{align*}
[\text{napna}] & \quad \text{‘stretch your legs’} \\
[\text{nâma}] & \quad \text{‘meat’} \\
[3^\text{w}:\text{lâ}] & \quad \text{‘to get out’} \\
[3^\text{w}:\text{la}] & \quad \text{‘to tell’} \\
[\text{mabe:le}] & \quad \text{‘sorghum’} \\
[\text{mabê:le}] & \quad \text{‘breasts’} \\
[\text{rait:sâ}] & \quad \text{‘wash e.g. clothes or dishes’} \\
[\text{rait:s}a] & \quad \text{‘vomit’}
\end{align*}
\]

2.4.2 Downshift

Downshift is the slight lowering of each successive high tone in an utterance. This tonal phenomenon is observable in Shekgalagari as illustrated below, where the symbol (——) indicates downshift.

(6)

\[
\begin{align*}
[\text{o ra:xôr wêwâ dôxâ?}] & \quad \text{‘Do you mean you are going/leaving?’} \\
[\text{ntsô: bâcô mõbo-n-mõtu:n}] & \quad \text{‘Remove those people from the house’}
\end{align*}
\]

2.4.3 Length

Like most Southern Bantu languages, words produced in isolation in Shekgalagari, and final words in sentences exhibit characteristic length on the vowels in the penultimate syllable. This is illustrated below.
Chapter Two. Shekgalagari: Genetic classification and linguistic issues

(7) a. **Isolated words**

- [léqa] 'try/test'
- [ba:tā] 'look for/want'
- [mó:cʰo] 'person/human being'
- [qále:jʷɛ] 'maybe'

b. **In sentences**

- [ke leqa gá:tsʰi] 'I am trying/testing again'
- [ba bata pʊri] 'they are looking for a goat'
- [le mocʰo jo mo:jʷɛ] 'And the other person'
- [mbé qalejʷe o hɔjɔ] 'Maybe s/he s there/present/in'

(7) also shows that in non sentence final position, lengthening on the penultimate syllable is reduced considerably. The italic words in 7 (b), which were lengthened in isolation in 7 (a), are no longer lengthened when they occur within a sentence.

2.5 **Syllable structure**

As is typical of Bantu languages, Shekgalagari displays an open syllable structure consisting of a vowel which may be preceded, but never followed by a consonant — (C)V. In addition to this, there are syllabic consonants, the majority of which are nasals [m, n, n, and ŋ]. The occurrence of syllabic liquids is very marginal, with only the alveolar trill functioning in this position, and even then, only one word in the language exploits this sound as a syllabic consonant. Most of the words given below for illustration are disyllabic.

(8) a. **VV**

- [é] 'yes'
- [áo] 'to indicate disagreement, despise, irritation, being fed up, etc.'

b. **CV**

- [bɔ] 'see/look'
- [mɛ] 'take'
- [lɔ] 'this one - e.g. this stick'
c. CVCV

[bɔ:na] ‘see/look’
[qá:la] ‘ranch’
[ga:ba] ‘carry on the shoulder/back’

d. Syllabic nasals

[ŋːmɛ] ‘mother’
[ŋːná] ‘me/myself’
[ŋːá:ja] ‘no’
[ŋːá:ca] ‘hit me’

e. Liquids

[r̥a]/[r̥ɛ] ‘mister/man’

The syllable structure is maintained even with respect to affricates; a simplex dual cluster which consists of a stop and a homorganic fricative — and observing sonority sequencing and distance. In orthography, Shekgalagari, like Setswana which has the same syllable structure, can have what looks like a maximum of four consonants word (syllable) initially. This is by no means a contravention of the syllable structure, since the last two consonants always represent aspiration and labialisation respectively. (9) illustrates this aspect with examples from Setswana.

(9) Orthography | Phonetic | Meaning
--- | --- | ---

| tséna | [ts‘ɛ:na] | ‘come in’ |
| tshába | [ts’há:ba] | ‘run away’ |
| tshwára | [ts’hwá:ra] | ‘hold/touch’ |

Shekgalagari has no orthography of its own, but the phonetic transcription of the Setswana words used in (9) for Shekgalagari would be as follows:

(10) | Phonetic | Meaning
--- | --- | ---

| [tʃ ‘ɛ:na] | ‘come in’ |
| [tʃ’há:ba] | ‘run away’ |
| [tʃ’hwá:ra] | ‘hold/touch’ |

In conclusion, although four consonants seem to occur syllable initially in a word, Shekgalagari, like most Bantu languages, is a (C)V language. Nasals and liquids
may assume a vocalic function, and (when they operate in this position) may even acquire tone, which may either be high (H) or low (L). Shekgalagari also exhibits downstep in utterances which are longer than a word. Penultimate syllables in words exhibit lengthening. In longer utterances, however, this penultimate lengthening is altered — often occurring in the penultimate syllable of the final word only.

2.6 Summary

This chapter has sought to address the issue of the classification of Shekgalagari, a minority Bantu language spoken in Botswana, revealing in the literature strong arguments for the autonomy of Shekgalagari as an independent Sotho language cluster, with its own dialects. Its vowel inventory suggests that it is a seven vowel system: /i,e,e,a,o,o,u/. That the language has voiced and aspirated stops and affricates does not appear arguable. The controversy seems to be whether Shekgalagari has the so-called 'ejectives' or the voiceless unaspirated stops. The language has a (C)V syllable structure typical of Bantu languages, as are the prosodic features of tone (H and L), downstep and penultimate length manifested in the language.
3 Stops: VOT and F0 perturbation, phonological representation and issues of co-articulation

3.1 Introduction

The aim of this chapter is to present various aspects of stop consonants, specifically voice onset time (VOT) and F0 perturbation, and also to present theories of stop representation and matters of co-articulation. The rest of the chapter proceeds as follows. Section 3.2 deals with VOT and F0 perturbation. Section 3.3 presents theories on the representation of the voicing structure of stops, and is followed by studies on stop voicing in section 3.4. Section 3.5 deals specifically with ejectives and section 3.6 with co-articulation and accounts of co-articulation. Co-articulatory resistance theory is presented in section 3.7. The chapter is summarised in section 3.8.

3.2 Voice onset time and F0 perturbation

3.2.1 Voice onset time

VOT involves the temporal relations between the release signalled by the burst and the onset of voice pulsing, and may be specified by a single value. It has been shown to be highly effective in delineating contrast in most stops with different voice types in many languages. (Lisker and Abramson, 1964,1967).

In initial CV context, commencement of phonation may happen prior to the release of the burst (voicing lead), it may coincide with the release (coincident phonation) and it may be considerably delayed after the release of the burst (delayed phonation). Stops with voicing lead are assigned negative VOT values. Those with (almost) coincident phonation are assigned (short positive) or zero VOT values and those with delayed phonation long positive values. This variation in VOT varies from language to language. Some languages have a two-way variation in their stop systems and others are three-way systems, and yet others manifest a four-way contrast.

Two category languages include English and Spanish. In English, the relevant phonemic distinctions are manifested by the delayed voicing (for example the sound [p], as in [pʰɪn]), and coincident phonation (for example the
sound [b], as in [brm]).\footnote{Actually, English [b] is produced with either coincident, advanced or even slightly delayed phonation (Lieberman & Blumstein 1988:197). Unlike French [b], this sound is not produced with energetic vibration of the vocal cords throughout its articulatory closure. Rather, there may be no voicing at all (hence coincident phonation), and if any, it is only produced for part of the closure (hence advanced phonation).} Thai is an example of a language with a three-way contrasting system: namely; voicing lead, coincident phonation and delayed phonation. Hindi and Gujarati are four-way category languages with voicing lead, coincident phonation, delayed phonation, and a fourth category exhibiting aspiration and voicing concurrently.

The fact that ‘speakers can exploit very complex co-ordination of laryngeal behaviour in relation to supraglottal articulations to achieve phonological distinctions’ (Clark & Yallop, 1990) means that phonological systems have a rich diversity of stop types — so much that the parameter of VOT alone may not be able to effectively distinguish between them in a system. For instance, although Korean belongs to the category of three-way VOT contrasting languages, VOT’s resolution of the distinction between the unaspirated and the so-called slightly aspirated stops in this language shows overlapping values and therefore is insufficient to distinguish the contrast (Abramson, 1977). Similarly, four-way contrasting systems have ‘murmur’, where phonation and turbulence occur simultaneously (Hirose \textit{et al.}, 1974). These languages have voiced aspiration contrasting with voiceless aspiration, and the timing dimension falls short in distinguishing these particular stops from the voiced stops and the aspirated stops in the relevant systems. Marathi, Hindi and Gujarati are some of the languages which have ‘murmured’ sounds.

\subsection*{3.2.2. F0 perturbation}

The other voicing cue relates to the aspect of F0 perturbation and contour at the onset of the following vowel. This deals with the fundamental frequency of vocal fold vibration during the initial portion of the vowel and the F0 contour from the start of the vowel to the steady-state portion of the vowel. Stops with different voicing structures affect the pitch perturbation of the subsequent vowel (Ohde 1984, Lea 1973). F0 can either rise or fall as a function of the VOT property of stops (Ohde 1984). Voiceless stops (especially voiceless aspirated stops) may manifest higher F0 ranges than voiced stops in word initial context. This variation in pitch ranges may serve as an acoustic cue to the voicing structure of the stops (Hombert \textit{et. al}, 1979, Ohde 1984, Lea 1973).
There are two different hypotheses: namely, the aerodynamic and the vocal fold tension hypothesis, which attempt to explain the phonetic bases for the various effects of voiced-voiceless stops on F0 perturbation. These are given by Hombert et al. (1979) as follows. According to the aerodynamic hypothesis:

During a voiced stop, oral pressure gradually builds up, thus decreasing the pressure across the vocal folds ... which in turn decrease the F0. Upon the release of the stop, the pressure drop returns to normal, producing an initially low and rising F0 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 vibration at vowel onset. As the airflow returns to normal, the F0 will too. Thus, after the voiceless stops, the F0 contour will be initially high and falling (Hombert et al., 1979: 42).

With regard to vocal fold tension, Hombert et al. (1979) gives the following account:

The basic assumption of the vocal fold tension hypothesis is that, in the course of making the voiced versus voiceless distinction on stops, vocal cord tension is changed so as to affect the F0 of adjacent vowels. Halle and 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 F0 (Hombert et al. 1979:42).²

² Hombert et al. (1979) do not, however, support any of these hypotheses. They argue that they are relevant only to prevocalic consonants and do not apply the same effects to sounds in the post-vocalic environment (cf. Hombert et al. 1979:43) According to them, the voiced-voiceless dichotomy is effected by the vertical height of the larynx. Basing on experiments conducted by Ewan and Krones (1974), they mention that 'the vertical height of the larynx differs for voiced and voiceless stops ... Their (i.e. Ewan and Krones') findings of higher larynx position for voiceless stops as opposed to voiced stops - coupled with the well-documented fact that in normal speech, other things being equal, F0 is positively correlated with larynx elevation - is compatible with the 'vertical tension' hypothesis'. They continue to say that 'additional evidence in favour 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).
Another acoustic property of stops which contributes to the voicing distinction is that of the intensity of the release of the burst. Voiced and voiceless stops exhibit variation in their articulatory force. In the production of voiceless stops, on the one hand, air pressure in the oral cavity is considered to be very high. Consequently, when they are released, strong airflow escapes through the mouth. This high-pressure airflow, accompanied by turbulence, produces high intensity in the burst. Voiced sounds, on the other hand, do not entail such high pressure in the oral cavity during their articulation. This yields lower burst intensity that in voiceless sounds.

In summary, the voiced-voiceless dichotomy in stops is cued by the presence or absence of vocal fold vibration during the stop occlusion, turbulence after the release of the stop, and the timing of the onset of phonation after the release burst (in prevocalic stops). Other parameters which contribute to the voiced-voiceless contrast in stops, which have not been discussed above, include the duration of the stop occlusion and that of the preceding vowel.

3.3 Theories on the representation of the voicing structure of stops

There have been many theories on the specification or representation of the voicing of segments in the world’s languages, and the scope and focus of these theories have been varied, none of which has been all sufficient to account for the sounds of the world’s languages. In this subsection, not all of these theories are discussed, and not even the ones discussed are dealt with in exhaustive detail. Most theories discuss the representation of the complete range of speech sounds in the languages of the world, but we will only focus on stop consonants.

3.3.1 The Sound Pattern of English model (SPE)

The SPE representation of speech sounds is feature based, where features ‘represent the phonetic capability of man’ (Chomsky & Halle 1968: 299). Segments are described in terms of inventories of features, often called feature matrices, each feature being binary-valued e.g. [± anterior]. There are different types of phonetic features used to specify segments: major class features, e.g. [± sonorant], [± vocalic], [± consonantal]; cavity features, e.g. [± coronal], [± anterior]; manner of articulation features, e.g. [± continuant], [± tense]; source features, e.g. [± voice], [± strident], and prosodic features, e.g. [± stress], [± length] (cf. Ibid. 299-300). The specification for the (say French) voiceless unaspirated stop [p], would resemble (1).
Chapter Three. Stops: VOT and F0 perturbation, phonological representation and issues of co-articulation

In a complete inventory of features, or representation of a word, the rows represent features and, in this example, the column represents the segment [p]. It is because of this manner of segment specification that this model has often been referred to as a *linear* representation of speech sounds.

The phonological component of this theory comprises two types of rules: phonetic rules and phonological rules. Phonetic rules, on the one hand, operate on the features in a language specific manner. They give detailed systematic phonetic description of sounds in such a way that may differentiate one language from another. For instance, they may specify quantitative values for a phonetic parameter (e.g. VOT) of a particular language. The terminal output of the phonetic rules in this model is the phonetic transcription which also encodes information relating to the pronunciation of the output as determined by the grammar of the language concerned. A further 'universal' phonetic component will then convert the output of the phonetic rules into their articulatory correlates. This 'universal' phonetic component, unlike the phonetic rules just discussed above, is 'not technically part of the grammar' (Keating 1984: 287) and, given its universal nature, its output is taken to be automatic and not language dependent. (Ibid.).

Phonological rules, on the other hand, may alter the values for the features, they may insert or remove segments, but they may not change the

---

3 Phonological processes are described in terms of rewrite rules which work on feature inventories such as the one shown in (1). These rules have the general complexion: $W \rightarrow X/Y \ldots Z$, where $W$ is the input which is transformed into $X$ when it is preceded by $Y$ and immediately followed by $Z$. Consider, for example, the
content of the feature matrices — which specify segments. Distinctive description between the natural classes of segments is done by this component of the grammar.

There have been a number of criticisms levelled against the SPE model of segment specification and representation. One has been the framework’s endless list of features in specifying segments. Also, as has been apparent from the discussion above, the features used here are meant to convey both unique phonetic categories of individual languages and the phonological representations of those categories at a cross-linguistic level. Another criticism (which is also relevant for the Halle and Stevens and the Lisker and Abramson feature systems discussed below) relates to the fact that the SPE feature system does not have a single phonological feature of ‘voicing’ which distinguishes, say, ‘voiced’ sounds from ‘voiceless’ sounds across languages. The framework also suffers from an over-generation of phonological processes, some of which are never utilised in the world’s languages.

3.3.2 Lisker & Abramson (L&A) (1964): voice onset time (VOT)

The L&A framework of laryngeal distinctions in stops is based on timing, and has been described in section 3.2.1 above.

3.3.3 Halle and Stevens (1971) (H&S): laryngeal features

With regard to consonant voicing, the H&S approach characterizes the voicing of stop consonants in terms of two independent parameters of vocal fold activity at the moment of stop release: the tension of the vocal folds (i.e. slack/stiff folds) and glottal aperture (i.e. aspirated, unaspirated, etc.). When these two parameters are adjusted in various ways distinctive acoustic consequences result, and hence distinctive phonetic characteristics. The various manipulations of the vocal cords thus give the following four features: [spread glottis], [constricted glottis], [stiff vocal cords] and [slack vocal cords], which, according to H&S appear to be sufficient to classify sounds in the languages of the world, and which H&S propose should therefore be incorporated into the universal phonetic feature

nasalisation process in English where a non-nasal vowel acquires nasality because of the adjacent nasal sound, as in a word like band [bænd]. The SPE model would represent this process in this way: [- nasal] → [+ nasal] / ______ [+ nasal].
framework (H&S:201). By means of combination, these four features produce
nine distinctive phonetic categories of segments. These phonetic categories and
their feature specification are summarised in Table 3.1 obtained from Halle and
Stevens (1971:201). We will focus on stops only, although the H&S feature
geometry is for all obstruents.

<table>
<thead>
<tr>
<th>Obstruents</th>
<th>b₁</th>
<th>b</th>
<th>p</th>
<th>pₖ</th>
<th>bʰ</th>
<th>pʰ</th>
<th>⁶</th>
<th>ᵃ</th>
<th>p'</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spread glottis</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Constric. glott</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Stiff voc. folds</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Slack voc. folds</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 3.1: Segment specification for obstruents under the H&S (1971) feature system.
Notes: b₁ represents a lax (plain) voiceless unaspirated stop; pₖ — the lightly aspirated stop,
e.g. the so-called ‘tense’ stop in Korean; bʰ — the voiced ‘murmur’ stops found in Hindi
and Marathi; p’ — ejectives, found in, for instance Xhosa; and ⁶ — an implosive
found in Xhosa; ᵃ — a laryngealized stop which is not truly implosive.

In this feature theory stops are divided into three broad groups by using the
features [spread glottis] and [constricted glottis]. They may be: plain [-spread,
−constricted] (cf. columns 1 to 3), aspirated [+spread, −constricted] (cf. columns 4
to 6) and glottalized [-spread, +constricted] (cf. columns 7 to 9). These three
groups are further subdivided into three groups by using the features [stiff vocal
cords] and [slack vocal cords]. These subdivisions are the voiceless stops, voiced
stops and a third group which incorporates implosives, lax stops such as the
Danish [b] and the Korean moderately aspirated stop. Voiceless stops are marked
by the features [−stiff, −slack], and examples include voiceless unaspirated stops,
voiceless aspirated stops and ejectives (cf. columns 3, 6, 9). Voiced stops are
marked by the features [−stiff, +slack] and they include traditionally voiced stops,
e.g. the French [b] (cf. column 2), ‘murmured’ stops, e.g. the Hindi [bʰ] (cf.
column 5), and a third class of stops which, as opposed to the other two stop
types, is glottalized (cf. column 8). As opposed to true implosives (represented

---

4 The combinations of these features in specifying a segment are of course
governed by physiological possibilities. Thus [+stiff, +slack] and [+spread, +constricted]
combinations may not be expected because they are not physiologically (and logically)
possible.
Chapter Three. Stops: VOT and F0 perturbation, phonological representation and issues of co-articulation

here in column 7 [6]) where both the lowering of the larynx and laryngealized voicing occur, these particular glottalized stop are reliably indicated by accompanying laryngealized voicing only, but the downward displacement of the larynx may not always happen. The third class of stop types has the configuration [-slack, -stiff] and includes the true implosives shown in column 7, the ‘lax’ voiced stops [b₁] such as the one found in, for example, Danish and ‘may occur in initial position for many speakers of English’ (H&S 1971:206) (cf. column 1), and the Korean slightly aspirated ‘tense’ stop [p₃].

Halle and Stevens also discuss (though to a limited extent) acoustic correlates of their features which have not been discussed here, but the generation of their feature system is largely based on articulatory information.

Some of the objections raised against this model of segment specification are, for example, Keating (1984) and Lombardi (1990). Keating (1984) argues that the H&S feature system is not adequate for the representation of voicing in stops apart from its relation to pitch (pg. 288). The idiosyncratic pattern of some English speakers in producing the utterance initial ‘voiced’ stop as ꞌvoicedꞌ and as ꞌvoiceless lax’ stop is one example of this. Keating argues that the difference between these two types of stops is not one of glottal configuration (i.e. vocal cord slackness) as H&S suppose, but that of the amount of oral air pressure. Similarly, the difference between English [p] and [b] as in rapid and rabid is not one of glottal stiffness as H&S presume, but rather, the [p] stop appears to be produced with an opening of the folds in word medial context, and at the moment of release the configuration of the folds may well be similar to that of the [b] stop. This, therefore, may require an addition of more features to the H&S feature system to accommodate other distinctive characteristics of stops which happen at moments other than release (Keating 1984:288-289). But this leads to the generation of more features, which would be an attempt to describe in accurate phonetic detail exactly how individual sounds in the languages of the world are articulated.

Another problem of the H&S feature is the representation of voicelessness. This is represented with the features [+stiff, -slack] although a spreading of the vocal folds appears to be more appropriate. But this spreading gesture, however, represents ‘aspiration’ in the H&S geometry (Lombardi 1990: 6). Another point is that segment specification is always done using a combination of features. No one feature seems adequate to distinguish one stop type (e.g. voiced) against another stop type (e.g. voiceless). Like the SPE model, the H&S feature geometry is plagued with the problem of over-generation of features capable of specifying segments which never code contrast in languages; e.g. voiced laryngealized stops vs. true implosives.
3.3.4 Keating (1984, 1990)

In the 1984 article, Keating argues that in deciding phonological feature systems for sounds in the languages of the world, the inclusion of detailed phonetic information regarding these sounds should be removed. She proposes three kinds of representation. First, there should be as many phonological feature e.g. [± voice] and feature values, e.g. {voiced}, {vls. unas} and {vls.asp} as are required to distinguish natural classes in a phonological system. Secondly, there should be as many phonetic categories as are needed to distinguish between categories in any given language. The third and last one deals with the pseudo-physical\(^5\) component which deals with as many parameters as are necessary to describe acoustic description of the segments. Two of these, the phonological feature and feature values and phonetic categories will be addressed shortly below.

Part of Keating's work builds on that of Lieberman (1970, 1977) which proposed the binary phonological feature [± voice] as a feature which could be implemented differently in different languages 'along the continuous dimension of V[voice] O[nset] T[ime]' (pg. 290). Modifying this work, Keating (1984) proposes that there is a fixed universal and specified set consisting of three phonetic categories coding possible contrasts in stop consonants: {fully voiced}, {voiceless unaspirated} and {voiceless aspirated} stop consonants (which have acoustic and articulatory correlates), and that the binary phonological feature values (i.e. [± voice]) may be implemented in the languages of the world as categories selected from this fixed phonetic universal set. These phonetic categories directly map onto VOT lead, short-lag and long-lag respectively for stops in word initial context. Keating (1984:290) goes on to say that 'these mappings will be part of the definition of the phonetic categories, and therefore universal; e.g., {voiced} will involve vocal-fold vibration and low periodicity during consonant closure. To some extent, however, they will be language specific....' Just how these phonetic categories are language specific is the subject of the following paragraph.

Consider, for instance, English and Polish, two-way category languages, which have the phonological contrast between [+ voice] i.e. /b, d, g/ and [- voice] /p, t, k/ stops. The implementation of the phonetic categories by the languages will be as follows. Depending on context, [+ voice] stops in English would be {vls. unaspirated}, while in Polish they would be {+ voice}. Similarly, English [-voice] stops would be {vls. aspirated} while in Polish they would be {vls.

\(^5\) That is, not purely physical.
Thus phonologically, the two languages are the same in that they are two-way contrasting languages, (and thus could be described using the phonological feature [± voice]), but they are phonetically distinct since they implement the phonetic categories differently. It should be pointed out that on this point, Keating (1984)'s analysis is limited to two-way contrasting languages. Three-way and four-way systems such as Thai and Hindi are excluded from her discussion 'largely because it is unclear whether such languages should be analysed as having a single non-binary feature [voice], or more than one binary feature' (pg. 191).

Going back to the three kinds of representation mentioned above, Keating pointed out the need to use a phonological feature as well as feature values which would be adequate to distinguish contrast in natural classes in a language. We will now consider this point.

*The phonological feature.* Keating (1984) argues that a more accurate account of segment representation (for two-way contrasting languages) could be done by way of levels of representation. There is the phonological level of representation and the phonetic category level of representation (i.e. level of implementation). At the phonological level of representation, phonological rules are applied to binary feature values (e.g. [± voice]), and their output is the phonetic category values (e.g. {voice}, {vls. aspirated}, {vls unaspirated}) which are different in different languages. These phonological rules cannot anticipate their phonetic output, in which case they can work with whatever output a language allows. Thus two level representation helps to keep phonetic details separate from phonological considerations, and also allows for different implementations of the binary phonological feature across linguistic systems.

Keating points out experimental data from other researchers which supports her theory. Some of these studies investigated vowel duration before 'voiced' and 'voiceless' stops in several languages, mostly two-category languages, and observed that vowels tended to be longer before 'voiced' stops than before 'voiceless' stops. But similarly, vowels also tended to be longer before phonetically 'voiced' stops than before 'lax' stops. This, according to Keating (1984:291,292) appears to show that the relationship between vowel duration and voicing was conditioned by the underlying phonological feature [± voice], rather than being mechanically determined by the phonetic voicing during the occlusion of the stop. This further buttresses the need for separating phonetic and phonological levels of representation.

*Phonetic categories.* Another of the levels of representation proposed by Keating deals with phonetic categories. It is proposed that there should be as
Chapter Three. Stops: VOT and F0 perturbation, phonological representation and issues of co-articulation

many phonetic categories as are needed to distinguish between categories in any given language. The phonetic categories dealt with here are described in: word initial context, in terms of VOT 'and the voicing dimension', limited to three contrasting categories since it is thought that languages only have these categories. Keating also proposes that the three categories proposed here 'is the right number', and 'are the same three in various languages'. As already discussed above with respect to VOT, there could be voicing lead: when voicing precedes the stop release; short-lag: when voicing coincides with the release; and long-lag: when voicing is considerably delayed after the release of the stop. Stops with voicing lead are associated with the {voiced} category, those with short-lag are associated with the {vs. unaspirated} category and those with long-lag are associated with the {vs. aspirated} category.

Keating's proposition of three as 'the right number' for phonetic categories is based on a study by Lisker and Abramson (1984) which concluded that there the languages of the world exhibited no more than a three-way contrast in their stop systems. The study also pointed out that where a system appeared to display more than a three-way contrast, there would be overlap in VOT values between at least two of the categories, and that these could be distinguished through some other dimension, e.g. tension of the articulators. Thus the three phonetic categories: {voiced}, {vs. aspirated} and {vs. unaspirated} appear to be the basic ones implemented by the languages of the world, 'And in fact, they are also sufficient elsewhere, since no greater number of contrasts is found in any other position' (pg. 297).

Phonetic categories: the rules. In the theory, the implementation of phonetic categories is different in different languages. Keating, therefore, says that the phonetic category implementation rules are thus language-specific. She goes on to propose that these phonetic rules would not be essential if different implementation of phonetic categories by the languages of the world were derived by a general principle, which could be 'polarization of two adjacent categories along the voicing dimension.... According to this principle, within the limits of the implementation chosen — i.e. the phonetic categories — there is maximal separation of the distributions of values' (pg. 309). Consider, for example, the English and Polish stops discussed above. It was pointed out that, depending on context, [+ voice] stops in English would be {vs. unaspirated}, while in Polish they would be {+ voice}. Likewise, English [- voice] stops would be {vs. aspirated} while in Polish they would be {vs. unaspirated}. But, when the timing for the voiceless unaspirated stops for these languages are compared, Polish stops show 5 ms higher VOT values (and therefore slightly aspirated) than the English
ones. According to the polarization principle, the contrast between the stops in Polish can be explained by polarizing the 'slightly aspirated' short-lag stop away from the long-lead voiced stop. But the situation with English, however, is slightly complicated, since the so-called 'voiced' stops may have lead values (which are not as in truly voiced stops), and they may have lag values (which are not as in voiceless aspirated stops). Keating (1984:309) believes that this situation could be resolved by regarding English as having a bicategory distribution of VOT values: lead and short-lag. In this way, the English short-lagged 'voiced' stop can then be polarized away from the long-lagged voiceless stop.

Keating is careful to point out that the polarization principle may not always adequately address the contrast in all languages. She points out that, for instance, different VOT values may be obtained for the same phonetic category implementations in a language, making it difficult for the polarization principle to apply.

In the 1990 article, 'Phonetic representations in a generative grammar' the two levels of representation: phonological and phonetic, are still maintained. Phonological representations describe overall contrast between segments in a language. Phonetic representations express contrast in a given context in a language. But this time phonetic representation has three levels. The first level, categorical phonetic representation, is still the output of the phonology in that it is the implementation of the phonological contrast as discussed above. It is regarded as neutral with regard to articulation and acoustics/perception, which are the concerns of the other two levels of phonetic representation. Here it is defined as 'clusters of feature values aligned with elements of internal segment structure', where the feature may be unary or binary. The features adopted here are for voicing, aspiration and glottalization. These are [voice], [spread glottis] and [constricted glottis]. These features may be related to particular landmarks during the production of the stop. For example, [voice] expresses events during the closure phase and is therefore associated with the closure node. It distinguishes truly voiced closure periods from voiceless ones. The value [+voice] then indicates vocal fold vibration and low frequency periodicity during stop closure. [spread glottis] and [constricted glottis] do not necessarily relate to a particular point in the articulation of the stop, but rather to the configuration of the vocal folds at the moment of release. [spread glottis] describes whether aspiration is present or not. The feature value [+ spread glottis], indicates the fact that the vocal folds are open at the release of the stop, leading to the presence of aspiration. The reverse is true with the value [- spread glottis], which explains that the vocal folds are closed at the moment of release, leading to lack of aspiration.
Keating (1990:326) goes on to describe the traditional VOT categories: VOT lead, short-lag and long-lag using the three features mentioned above as follows

(2)

(a) Voiced

Stop

Closure  Release

[+ voice]  [- spr gl ]

(b) Voiceless unaspirated

Stop

Closure  Release

[- voice]  [- spr gl ]
The other two levels of phonetic representation deal with the physical dimension, where segments are described within a specific domain — in continuous time and space. One of these levels is the output of articulatory rules, and is called articulatory parametric representation, and the other one is the output of acoustic rules, called acoustic parametric representation. Both of these parametric representations are derived from the categorical phonetic representation, which as we have seen, is also derived from the phonological representation. The relationship between the levels is summarized by Keating as shown in (3).
Chapter Three. Stops: VOT and F0 perturbation, phonological representation and issues of co-articulation

3.3.5. Kohler (1984) fortis and lenis stops

Kohler (1984) proposes a single feature [± fortis], (where [- fortis] is also known as lenis) as a feature adequate to account for phonological contrasts between obstruents such as /p, t, k/ and /b, d, g/ in the languages of the world. This feature, according to Kohler, is not abstract, but is rather a "power feature ... realized in articulatory timing and/or phonatory power/tension ... thus providing a phonetic basis for the fortis/lenis dichotomy" (pg. 150). This 'power' refers to the intensifying of movements by the organs of speech, or of 'energy expenditure' during the production of sounds. It is also discussed in relation to the strength of the air stream and to glottal tension. Other factors which may be associated with the [± fortis] feature include articulatory timing and laryngeal power/tension. Articulatory timing deals with the speed of the formation and release of the stricture for obstruents and may be language universal, and laryngeal power/tension refers to properties such as aspiration, voicing and glottalization, and may be language specific. To what extent a component may contribute to the fortis/lenis distinction of sounds is also a function of the context of the sounds being studied. This is discussed below.

Kohler starts by arguing that 'all phonological theories have treated the /b, d, g etc./ versus /p, t, k etc./ opposition as an atemporal distinction at a static point in a segment chain' (pg. 152). That is, the approach of some accounts of the phonological representation for the /b, d, g etc./ versus /p, t, k etc./ distinctions has been to describe these abstract phonological entities in physical phonetic terms. He further argues that adding an intermediate level of 'possible phonetic category mappings' between the phonological features and their phonetic properties as Keating (1984) proposes does not alleviate inherent problems in this kind of approach. Kohler rather believes that 'the time dimension should be integrated into the phonology' (pg. 152).

Secondly, Kohler points out that the /b, d, g etc./ versus /p, t, k etc./ contrast is usually represented using the feature [± voice], even when both categories lack vocal fold vibration. This provides the basis for continuing confusion between phonetic and phonological voicing. If the feature [± voice] could be left to the phonetics, where it represents periodic vibration of the glottis, and if the feature [± fortis] could be assigned to phonological categories, this confusion could be resolved.

Thirdly, Kohler also proposes that the feature [± fortis] is sufficient to represent phonemic variation in properties such as aspiration, laryngealization, preceding vowel duration, gemination, etc. This feature is also proposed to be of a
Chapter Three. Stops: VOT and F0 perturbation, phonological representation and issues of co-articulation

gradient nature, so that different states on the scale can be assumed in various contexts and languages, particularly languages with more than a two-category distinction.

The main distinction between fortis and lenis across the world’s languages and across contexts in Kohler’s account is what he calls the ‘phonetic power.’ The fortis segments are produced with more intensity and are auditorily more salient than the lenis series. Contrast in the stops of the world’s languages and across different contexts is presented in the following way. For two-category languages, the difference between the [+ fortis] series, /p, t, k etc./ and the lenis, [- fortis] is described with reference to co-ordination between the oral, velopharyngeal and glottal valves. For the fortis series, a narrower or tighter constriction in the vocal tract has to be developed relatively faster than when the lenis stops are being produced. Also, the velopharyngeal closure has to be tighter for the fortis stops than for the lenis ones. Vocal fold vibration may occur for the lenis stop if appropriate aerodynamic factors are met. ‘The three valves form a coordinative structure ... for obstruents productions, characterized as a whole by the [+ fortis] feature’ (pg. 153). Kohler continues to say that:

In stops, aspiration and voicing are glottal reinforcements of the fortis and lenis actions at the oral valve to produce the necessary intensity differences in the acoustic signal for a clear category in perception. The timing of the whole coordinative structure is important for this purpose, not just the voice onset in relation to the release (Lisker and Abramson, (1964)), but the relative timings of all three valvular actions. They are readjusted in different utterance positions to meet their diverse signalling requirements. (pg. 153).

Thus aspiration and voicing are correlates of the [+fortis] feature, and distinguish between distinctive segments by means of relative and co-ordinative timing of the oral velopharyngeal and glottal states, VOT and intensity differences in the oral cavity. As pointed out earlier, the timing and the laryngeal components contribute towards the fortis/lenis distinction of sounds to varying extents depending on the context of the sounds being studied. This is addressed as follows.

Utterance-initial stops: In Kohler’s theory, the influence of the laryngeal element is most prominent in this position. Stops produced in this context are distinguished primarily by adjustments in the glottis: the presence versus absence of aspiration and vocal fold vibration. Fortis stops may be signalled by the aspiration element at the release phase, and lenis stops may have vocal fold vibration followed by a weak release. Two-way contrasting languages often
utilize only one of these feature-signalling mechanisms. If aspiration is present, then there is no need to emphasize the lenis signalling feature.

**Intervocalic stops:** Both the timing and the laryngeal components have about equal influence in distinguishing stops in this position. Closure for the stop as well as the speed of the occlusion in intervocalic position is what signals the fortis-lenis distinction. Fortis stops achieve closure faster and have longer closure periods than lenis stops.

**Utterance-final and before silence:** If phonemic distinctions are preserved in this position, they may be signalled by the articulatory power of the closure phase. In most cases, glottal adjustments, i.e. voicing and aspiration, are highly less reliable. For instance, in the pre-pausal position, the glottis may be anticipating subsequent breathing and may therefore open. This may neutralize aspiration, which is the basic correlate of the fortis feature. Similarly, it is more difficult to keep the vocal folds vibrating when silence is being anticipated.

Kohler gives a detailed description of the behaviour of the organs of speech during the articulation of the fortis and lenis sounds in different contexts and in different languages. This is summarized in table 3.2 (a) and (b).

Where VOT overlap occurs in a language, as for instance in Korean, and in four-way contrasting languages, as in Hindi for instance, this can well be accounted
### Individual properties

**Vls. asp. stop**

(a). wide glottal opening with maximum opening at burst release.

(b). greater posterior cricoarytenoid peak than for the voiceless unaspirated stops, occurring later in the stop closure, rather near the plosion, producing a wide opening at the moment of release.

(c). long delay in voice onset after the release burst.

(d). increased glottal airflow due to extensive movements of articulators.

(e). shorter closure duration

---

**Fortis 'Tense' stop**

(a). laryngeal tensing – i.e. body tensing of the vocal folds.

(b). greater posterior cricoarytenoid peak than for the voiceless aspirated stops, occurring at the release of the stop plosion.

---

**Vls. unaspl. stop**

(a). narrow glottal opening which is virtually closed after the release burst.

(b). posterior cricoarytenoid peak, less than for the voiceless aspirated and tense stops, occurring earlier than for the voiceless aspirated stops, and producing a small opening before the plosion.

### Shared properties

- (a). auditorily more salient than lenis stops.
- (b). tighter stricture formed rapidly in the vocal tract.
- (c). tighter velopharyngeal closure.
- (d). shorter vowel duration after stop.
- (e). shorter closure duration.
- (f). greater posterior cricoarytenoid peak than for the voiceless aspirated and tense stops.
- (g). narrow glottal which is dosed after release burst.
- (h). posterior cricoarytenoid peak less than for the voiceless aspirated and tense stops, occurring earlier than for the voiceless aspirated stops, and producing a small opening before the plosion.

---

**Table 3.2(a): Summary of the articulatory properties of fortis and lenis stops in the world’s languages (Kohler 1984).**

(i). More extensive movements as well as great velocities of the articulators producing the stricture, resulting in quick formation of occlusion and slow release – in non initial position.

(ii). Greater momentary impact between articulators at contact.
### Example Individual properties

- **Vd. asp.**

  - (a). vocal fold vibration under appropriate aerodynamic conditions, which happens concurrently with aspiration.
  - (b). Relatively shorter closure period than in the voiced cognates due to aspiration.

### Shared properties

- (1). auditorily less salient than the fortis stops.
- (2). less tight stricture formed in the vocal tract than for the fortis stops.
- (3). less tight velopharyngeal closure than for the fortis stops.
- (4). longer closure duration than for the fortis stops.
- (5). longer vowel duration after stop
- (6). vocal fold vibration under appropriate aerodynamic conditions (very essential)
- (7). low (and rising) F0 at vowel onset due to relatively relaxed vocal folds.
- (8). lesser intra-oral air pressure than for the fortis stops.

### Articulatory behaviour of the relevant organs of speech

- (i). Relatively slower articulatory movements than for the fortis stops, resulting in a slower formation of occlusion and fast release – in non initial position.
- (ii). Greatly reduced airstream power and tension associated with powerful articulatory movements.
- (iii). Less intra-oral air pressure
- (iv). Active expansion of the supra-glottal cavity to maintain voicing.

<table>
<thead>
<tr>
<th>Lenis</th>
<th>Voiced stop</th>
<th>Individual properties</th>
<th>Shared properties</th>
<th>Articulatory behaviour of the relevant organs of speech</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>(a).vocal fold vibration.</td>
<td>(1). auditorily less salient than the fortis stops.</td>
<td>(i). Relatively slower articulatory movements than for the fortis stops, resulting in a slower formation of occlusion and fast release – in non initial position.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(a).vocal fold vibration under appropriate aerodynamic conditions, which happens concurrently with aspiration.</td>
<td>(2). less tight stricture formed in the vocal tract than for the fortis stops.</td>
<td>(ii). Greatly reduced airstream power and tension associated with powerful articulatory movements.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(b). Relatively shorter closure period than in the voiced cognates due to aspiration.</td>
<td>(3). less tight velopharyngeal closure than for the fortis stops.</td>
<td>(iii). Less intra-oral air pressure</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(4). longer closure duration than for the fortis stops.</td>
<td>(iv). Active expansion of the supra-glottal cavity to maintain voicing.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(5). longer vowel duration after stop</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(6). vocal fold vibration under appropriate aerodynamic conditions</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(very essential)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(7). low (and rising) F0 at vowel onset due to relatively relaxed vocal folds.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(8). lesser intra-oral air pressure than for the fortis stops.</td>
<td></td>
</tr>
</tbody>
</table>

Table 3.2(b): Summary of the articulatory properties of fortis and lenis stops in the world’s languages, (continuation of (a)) (Kohler 1984).
for in terms of the activities of the organs of speech as follows. Korean has a three-way category stop system, all falling on the positive half of the VOT continuum. These are the aspirated stop, the weakly aspirated stop and the lax stop. According to Kohler (1984:162), both the aspirated and the weakly aspirated stop are fortis stops, and the lax stop is the lenis one. In order to solve the overlap between the two fortis stops, another aspect of the fortis/lenis distinction, 'laryngeal tensing,' is introduced. The weakly aspirated stops are distinguished from the aspirated stops in that 'they are accompanied by a strong and sharp activation of the vocalis muscle immediately before the stop release .... This body tensing of the vocal folds, combined with a decrease in stiffness of the vocal fold cover ... is absent from the ... aspirated ones' (Kohler 1984:161). The aspirated stops are realized by 'wide glottal opening with its maximum at the moment of release, resulting in a substantial increase in airflow' (pg. 161).

Hindi is another language with a similar overlap to that of Korean, with voiced and voiceless stops and their aspirated cognates. Kohler accounts for these four types of stops in terms of the size of the glottal width and also the timing of the glottal opening. These are described as follows. The voiceless unaspirated stops have a narrow glottal width opening before the release burst whereas their aspirated counterparts have a wide opening. The voiced aspirated stops have a narrow glottal width after the release burst, whereas their voiced cognates do not have any glottal opening after the release. According to Kohler's theory, the voiceless unaspirated stops and their aspirated counterparts belong to the fortis category, whereas the voiced stops and their aspirated cognates are lenis stops.

In summary, Kohler (1984) argued for the single feature [± fortis] as adequate for distinguishing phonological segments in the world's languages. This feature, according to Kohler, is not abstract, but has a phonetic basis. It is discussed in relation to the strength of the air stream and to glottal tension. Other factors which may be associated with the [± fortis] feature are articulatory timing and laryngeal power/tension. Articulatory timing deals with the speed of the formation and release of the stricture for obstruents and may be language universal. It is most prominent in initial context. Laryngeal power/tension refers to properties such as aspiration, voicing and glottalization. The contribution of articulatory timing and laryngeal power/tension components towards the fortis/lenis contrast may be affected by the context of the sounds being studied.
3.3.6 The Element based theory

In the Element Theory approach, contrast in consonants is represented in terms of elements. Adopting some of the terminology used by Halle & Stevens and Lisker & Abramson, the phonetic exponents and acoustic signals for the stop categories are as follows. Truly voiced stops are characterised by slack vocal cords, long VOT lead and lowered fundamental frequency (voice bar). Voiceless aspirated stops are characterised by stiff vocal cords, raised fundamental frequency and long VOT lag. Voiceless unaspirated stops have short VOT lag. Voiced aspirated stops are characterised by both the slack vocal cords found in truly voiced plosives and the raised fundamental frequency found in aspirated stops. The element specification for stops with lowered fundamental frequency is L. Those with raised fundamental frequency possess the element H. And those with both the slack vocal cords and raised fundamental frequency have the specification LH. Voiceless unaspirated stops are regarded as having no element specification. Element specification for the various linguistic systems has been summarised in table 3.3, adapted from (Monaka, Abberton & Harris 1997).

<table>
<thead>
<tr>
<th>Type</th>
<th>Language</th>
<th>short lag</th>
<th>long lead</th>
<th>long lag</th>
<th>breathy</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>S. German</td>
<td>/p/</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>II A</td>
<td>French</td>
<td>/p/</td>
<td>/b/</td>
<td></td>
<td>/p b/</td>
</tr>
<tr>
<td>IB</td>
<td>English</td>
<td>/b/</td>
<td></td>
<td>/p b/</td>
<td></td>
</tr>
<tr>
<td>III</td>
<td>Thai</td>
<td>/p/</td>
<td>/b/</td>
<td>/p b/</td>
<td></td>
</tr>
<tr>
<td>IV</td>
<td>Gujarati</td>
<td>/p/</td>
<td>/b/</td>
<td>/p b/</td>
<td>/b b/</td>
</tr>
</tbody>
</table>

Table 3.3: Element specification for some linguistic systems. Note: the shaded areas show that the system does not have the relevant element specification.

3.4 Studies on stop voicing

Examination of the characteristics of stop voicing has been conducted from three different fields: physiology, acoustics and auditory science, and there have been language specific studies and cross-language studies.

Physiological investigations are concerned mainly with analysing the behaviour of the larynx during the production of stops in an attempt to find

---

5 That is, Southern German.
properties of stops in individual languages and the implications of the findings to universal phonetic and phonological properties of stops. These studies make use of techniques such as electromyography (EMG) and fiberoptic observations. Hutters (1985) investigated vocal fold behaviour in aspirated and unaspirated stops in Danish using EMG and photo-electric glottography techniques. She concluded that these type of stops are distinguished by different vocal fold adjustments: with aspirated stops executed with the greater glottal width at the time of release in relation to their unaspirated cognates in which the glottal width is spindle shaped; and that variation in the timing dimension is a mere consequence of these laryngeal gestures. Dixit (1987) also used photo-electric glottographic techniques to examine the various mechanisms of vocal fold configurations in Hindi; in particular the aspect of timing of laryngeal gestures at the moment of the burst in relation to events in the mouth, and the degree of glottal opening in the four types of Hindi stops: the voiced aspirated and unaspirated stops and the voiceless aspirated and unaspirated stops. He also conducted a comparative study with other similarly labelled stops in other languages, and concluded that laryngeal mechanisms used to generate the four types of stops in Hindi could also be relevant for the similarly labelled stops in other languages.

Yadav (1984) conducted a fiberoptic observation of Maithili stops. His results are somewhat similar to those of Hutters in that Maithili stops are also differentiated by various vocal fold adjustments, with VOT as a repercussion of this. Voiced stops in Maithili are described in terms of the abduction/adduction of the vocal folds and aspirated stops in terms of the amount of glottal adduction: with aspirated stops manifesting the greatest glottal opening at the moment of or immediately after release. Another physiological study is a cineradiographic filming of the larynx by Kim (1970), during the production of Korean stops the aspirated, slightly aspirated and glottalized. This study also concluded that these stops are differentiated by the degree of aspiration at the moment of release. Aspirated stops have the greatest degree of aspiration, followed by lax stops and lastly the so-called 'tense' (least aspirated) stops. But like Hutters, Kim argues that the crucial factor in distinguishing between the Korean stops is the glottal width at the time of release, with VOT being only a concomitant of this. Thus, the wider the glottal width, the heavier the aspiration which in turn means longer VOT, and vice versa.

Authors have often reported an overlap in VOT values between lax and tense stops or between aspirated and lax stops in Korean. See, for instance, Kim (1965) and Han & Weitzman (1970).
Chapter Three. Stops: VOT and F0 perturbation, phonological representation and issues of co-articulation

Abberton (1972) conducted a non-invasive laryngographic study of Korean stops and observed variation in vocal fold activity in the production of these stops for two native speakers of the Seoul dialect. The Lx signal showed gradual build-up of the amplitude of Lx traces (vocal fold contact waveform) for the slightly aspirated stops, ‘the so-called tense stops’ (for the male speaker), which is characteristically absent in the other stop types. There is inter-speaker variation though, because the female speaker differentiates between the stop types in terms of VOT only.

Another laryngographic study is that of Lindsey et al. (1992) (see also Haruna 1990) on Hausa, a Chadic (Afro-asiatic) language which contrasts ejectives with plain consonants. The main objectives of this study were to investigate, amongst other things, whether there was an Lx waveform type or types associated with laryngealized consonants in the language. The study also investigates waveform characteristics of ejective consonants vis-à-vis plain consonants. At consonant onset, the Lx waveform for the intervocalic ejective affricate [ts’] in the word [tsaats’aa] ‘rust’, was the type which the authors ‘associate with glottal closure, indicating that the glottal stricture lasts the full duration of the consonant...’ (Lindsey, Hayward & Haruna 1992:519). This is shown in figure 3.1 (adapted from Lindsey, Hayward & Haruna 1992:519).

At vowel onset, the ejective affricate [ts’] also showed ‘markedly greater low frequency perturbation (slow up and down deflection) in the Lx waveform...’
Chapter Three. Stops: VOT and F0 perturbation, phonological representation and issues of co-articulation

(Ibid. 1992:520). By contrast in the plain intervocalic fricative [s] in the word [taasáá] 'metal bowl' the open quotient at the start of the consonant increases, 'anticipating abduction rather than adduction of the vocal folds, and showed faster up and down deflections on the Lx traces' (Ibid. 519 - 520). Ejectives thus generally showed decreased open quotient than plain consonants. Other characteristics of ejectives at vowel onset observed in Hausa included irregular mode of vocal fold vibration with period-by-period fluctuation in amplitude, and this was observed for the ejective stop [k'] in the word [shaak'aa] 'smell, sniff' (Haruna 1990:505).

Acoustic examination of stops incorporates the major contribution of Lisker and Abramson (1964). This work examined how effective the aspect of timing is in distinguishing stops with different voicing configurations in eleven different languages. VOT is shown to be effective in distinguishing the voicing contrasts of stops in these languages and it manages to classify languages into different groups depending on the number of stops established per language: two-way contrasting languages, three-way contrasting languages and four-way contrasting languages (although the dimension of timing does not satisfactorily distinguish the stop types in this category of languages). Stops are effectively divided into those with voicing lead, those with short voicing lag (sometimes the release of the burst is simultaneous with the onset of glottal pulsing) and those with long voicing lag. These are roughly associated with the phonological categories of 'voiced stops', 'voiceless unaspirated stops' and 'voiceless aspirated stops'.

Other similar studies have since supported Lisker and Abramson's findings (e.g. Flege 1979 and Keating et al. 1983). Stops exhibiting breathy voice, and the so-called 'tense' stops in Korean are some of the few that VOT does not satisfactorily distinguish.

Haag (1979) investigated the relationship between VOT and place of articulation in German stops, and came to the conclusion that a correlation exists between the magnitude of VOT and place of articulation. Generally, VOT tended to increase as the place of articulation moved to the back of the mouth. Other works include those of Keating et al. (1983) who examined allophonic variation for voiced and voiceless stops in fifty-one languages. They observed that the languages under investigation tend to aspirate the /p,t,k/ series before stressed

---

7 The 'voiced aspirated stops' also fall within this category, but VOT does not effectively distinguish these stops from the voiceless aspirated stops and from the voiced stops.

8 I was dependent on the literature for the information in this article since it was written in German.
Chapter Three. Stops: VOT and F0 perturbation, phonological representation and issues of co-articulation

vowels, maintain voicing for /b,d,g/ in intervocalic environments and have some type of voiceless unaspirated stop in all contexts.

With respect to auditory observation, an example could be that of Williams (1977) who conducted a perceptual study on initial stops in some Spanish dialects with an attempt to identify acoustic cues which distinguish between the voicing categories of Spanish stops in this context. VOT is identified as the main cue in identifying contrast between the dialects. Lisker & Abramson (1970) conducted a speech perception experiment using synthesised speech to study the identification modes of speakers of English, Spanish and Thai by systematically varying the time range (VOT) for the stops. They observed that the informants tended abruptly to shift from hearing a sound as 'voiced' to hearing it as 'voiceless', and that this was dependent on the timing (VOT) boundary point of each language. This is known as categorical perception. Shimizu (1977) also performed a perception experiment on the identification and distinguishing of initial stops in Japanese by native speakers using synthetic speech. He also observed results which were similar to those of Lisker & Abramson discussed above. The Japanese informants tended to abruptly stop hearing a sound 'voiced' to hearing it as 'voiceless' at a certain VOT perceptual boundary.

3.5 Ejectives

3.5.1 The production of ejectives

Ejectives, sometimes known as glottalized sounds (Clark & Yallop 1990:57) are produced on the egressive glottic mechanism initiated at the larynx. In this type of initiation, the glottis is closed so that only the air above the glottis is used to generate sounds. A constriction is simultaneously formed in the chambers above the glottis. The velum is raised to block airflow through the nasal cavity. And the relevant sounds are produced on the air trapped between the glottal and the supraglottal constrictions. Initiation involves abrupt vertical displacement of the whole larynx, initiating airflow; upwards in the egressive mechanism, and since pulmonic airflow is not involved, the whole impression is that of the larynx acting as a piston in the pharynx (see figure 3.2 for illustration).

---

9 The constriction in the supraglottal cavities can either be complete - as in the articulation of a stop, or partial - as in the production of a fricative. When the complete closure is followed by a fricative portion during the release phase, ejective affricates are produced.
In the egressive mechanism, the glottis is always closed, so that the sounds produced are always voiceless (although VOT variations are in principle possible). Ejectives are articulated on the egressive laryngeal pressure initiation which entails simultaneous constriction in the glottis and oral cavity. The whole larynx is then rapidly raised up, compressing the air trapped in the supralaryngeal cavities, and forcing it out of the mouth. In the production of plosives, this results in a characteristic explosive effect upon release.

Ejectives sounds are reported to be occurring in 18% of the world’s languages (Maddieson 1984a). In K’ekchi, a Quichean language spoken in Guatemala (Pinkerton 1986:130), contrast between voiceless egressive pulmonic plosives and their glottalic counterparts is illustrated in the following words.

\[(t’oqok)\] ‘to throw’  \[(toqok)\] ‘to break’
\[(fα:t’ok)\] ‘you threw it’  \[(fα:tok)\] ‘you broke it’

Ejectives have been reported in a number of North American Indian languages as well as in Caucasian languages (cf. Flemming et al. (1994) on Montana Salish, Gordon (1996) on Hupa, Maddieson, Bessell & Smith (1996) on Tlingit, and Maddieson, Rajabov & Sonnenschein (1996) on Tsez). Ejectives have also been reported in African languages, e.g. Hausa; (Lindau 1984), Haruna (1990), (Lindsey et al. (1992)), Tigrinya; (Fre–Wuldo (1985)), and Xhosa, (Finlayson et al. (1989))

\[\text{According to Ladefoged and Maddieson (1996: 78), the degree of pressure generated behind the occlusion is two times that of the normal pulmonic pressure.}\]
Chapter Three. Stops: VOT and F0 perturbation, phonological representation and issues of co-articulation

3.5.2 Studies on ejectives

As with other speech sounds, studies on ejectives have addressed various aspects of the sounds: e.g. VOT, stop duration, vowel characteristics (Lx waveform and amplitude of the Lx traces) prior to the stop and after the stop, burst spectra, F0 at vowel onset, etc. Various experimental techniques have been used: electrolaryngographic, e.g. Lindsey et al. (1992) on Hausa, Abberton (1972) on Korean so-called 'tense' stops, acoustic speech recording and spectra, e.g. Lindau (1984) on Hausa, Jessen (1999) on Xhosa, speech recording, spectra and perception e.g. Fre-Wuldo on Tigrinya (1985). (cf. Flemming et al. (1994) on Montana Salish, Gordon (1996) on Hupa, Maddieson, Bessell & Smith (1996) on Tlingit, and Maddieson, Rajabov & Sonnenschein (1996) on Tsez. And some studies, particularly early studies, have been subjective (cf. Chapter Two).

3.5.3 Features of ejectives

3.5.3.1 Features of ejectives: The problem. There is considerable variation on what features could be consistently associated with ejectives in the languages of the world. This is partly due to a number of reasons. Firstly, it has often been difficult to establish whether some stops in some languages were ejectives or plain voiceless unaspirated stops. Consider, for instance section 2.3.2 in Chapter Two, where various authors showed considerable inconsistencies with respect to whether Shekgalagari has plain voiceless unaspirated stops or ejectives. Some authors thought the language had plain voiceless unaspirated stop; e.g. van der Merwe and Schapera (1943); Dickens (1986, 1987); Andersson and Janson (1997), and others thought it rather had ejectives e.g. du Plessis and Kruger (1968). Doke (1954:243) records only one 'ejective' for Shekgalagari, the velar plosive [k']. Still, plain stops were thought to be realized as ejectives due to (emphatic) speech style (Dickens 1986:26, 1987:302). Whether ejectives exist at all in this language is still largely controversial. All of the above mentioned studies were, however, based on subjective description of the sounds in Shekgalagari.

Sometimes experiments on ejective stops in some languages have often given rather inconsistent and inconclusive results, partly due to idiosyncratic tendencies between informants and partly due to token-to-token variability between and within speakers. For instance, Xhosa has always been assumed to have ejectives in its stop system, Pahl (1989), Finlayson et al. (1989). But an
acoustic study conducted by Jessen (1999:29) observed that 'even in the relatively prominent conditions of the present study ejection was not a stable feature in the production of \( p, t, k \). There are many tokens that bear no audible characteristics of acoustical signs of ejection. Overall, the percentage of ejective productions turned out to be quite speaker specific."

Also, in languages which are thought to have true ejectives, what has been identified as cues for ejectives in one language has often not been a reliable cue in another language. An example could be Hausa and Navaho ejectives. In a cross-linguistic investigation of these languages (Lindau 1984), Hausa ejectives showed regular voicing at vowel onset whereas Navaho ejectives showed creaky voice after the release of the stop which continued into at least the first portion of the vowel. These observations make the determination of cues for ejective stops in the languages of the world rather difficult, the different experimental techniques used by the researchers also being a contributing factor.

3.5.3.2 Features of ejectives: General cues. Bearing in mind, as discussed above, that: different languages show different signal characteristics; sometimes speakers also show considerable variation in their realisation of ejective stops; and that considerable variation may also exist when the same ejective token is produced several times; with a token being sometimes produced in a simple voiceless unaspirated stop and sometimes as a glottalized stop; and also the fact that sometimes ejection of the voiceless unaspirated stops in a language may not necessarily be distinctive; the following may be considered to be general cues for ejective stops. Decreased open quotient and/or perturbation of some form (e.g. irregular vibration, one or two long cycles, alternating high and low amplitude of the Lx traces) prior to the target stop and/or at vowel onset (for the voiceless stops), cf. Lindsey et al. (1992) on Hausa, Lindau (1984) on Hausa, Abberton (1972) on Korean so-called 'tense' stops, Jessen (1990) on Xhosa and Henton et al. (1992). The rate of vocal fold vibration 'is typically, though not necessarily slower' (e.g. Ladefoged 1973, et. al., 1988). Long consonant duration has also been identified as a distinguishing cue in some languages e.g. Navaho (e.g. Lindau 1984). Long VOT duration has been reported in many languages of North America and some Caucasian languages where ejection in the voiceless unaspirated stops is reported to be distinctive (e.g. Flemming et al. (1994), Gordon (1996), Maddieson, Bessell & Smith (1996), and Maddieson, Rajabov & Sonnenschein (1996), see also Jessen (1999). High burst amplitude has also been
Chapter Three. Stops: VOT and F0 perturbation, phonological representation and issues of co-articulation

reported in languages like Tigrinya (Fre Wuldo 1985) and Xhosa—the stops were produced with some ejection (Jessen 1999).

It was pointed out in the previous chapter that the classification of Shekgalagari voiceless unaspirated stops is still largely controversial. Some authors have classified these stops as plain voiceless unaspirated stops while others thought they were ejectives. Chapter Four and Five of this thesis examine qualitative and quantitative aspects of these stops in the language respectively. In Chapter Four, these stops are examined using the Lx and Gx laryngograph, and spectrograms as well as the speech pressure waveform. The Gx signal maps out conductance change as the larynx is displaced. This displacement of the larynx may be caused by a number of factors, the glottalic airstream mechanism on which ejectives are produced being one of them. This gross displacement of the larynx may be non-invasively studied by means of ‘a direct-current connected version of the laryngograph providing a gross movement waveform, Gx’ (Fourcin & Abberton 1977:313). The Lx signal maps out conductance change across the vocal folds and provides rather detailed information about the nature of vocal fold activity during speech production in a non-invasive manner. The Sp signal gives information about activity in the vocal tract as a response to the action in the larynx. Some information on the production of the sound can also be obtained from this signal. The spectrogram also gives acoustic information, and has an excellent depiction of the various dynamic changes in the resonances of the vocal tract. Chapter Five focuses on VOT, and amongst other things, seeks to examine whether the VOT values obtained for Shekgalagari voiceless unaspirated stops could be comparable with those obtained for true ejectives in, for instance, native North American and Caucasian languages. It is hoped observations made in Chapter Four and Five will contribute to the understanding of the nature of production of Shekgalagari voiceless unaspirated stops specifically, and more generally to the understanding of ejective production in the languages of the world.

3.6 Co-articulation

The definition of co-articulation is still largely controversial because of its overlapping nature with a similar process—assimilation. Coarticulation may refer to a point in time when the configuration of the vocal tract is influenced by more than one sound (Farnetani 1997:371). This happens because during speech production the organs of speech interact with each other and their movements
overlap in time during the articulation of juxtaposed sounds. It is similar to assimilation which may refer to variation in speech sounds where one or more of a particular sound’s phonetic properties are altered and become similar to those of the neighbouring segments. But this definition could also be relevant to co-articulation. In fact, assimilation is sometimes referred to as optional co-articulation since it is not compulsory and inevitable. In reality, these two processes overlap with each other, and in practice, it is often not easy to determine whether a process is co-articulation or assimilation. Nevertheless, whether and how co-articulation is distinctive from assimilation is still a matter of controversy, and could perhaps also just be a matter of definition. However, it and will not be further dealt with here. Co-articulation is not necessarily always audible. It has been observed in most, in not all of the languages so far investigated, and it seems to differ from language to language (Farnetani 1997:376).

The influence of one segment on another in the production of speech could be of an anticipatory nature: where the segment on the right influences the one on the left, or it could be of a carry-over nature: where the preceding segment influences the one following it. Consider (5) adapted from Shin (1997: 13).

(5)

\[ \text{Carry-over co-articulation} \]

\[ \begin{array}{c}
\text{VI} \\
1 \\
2 \\
3 \\
4 \\
5 \\
6 \\
\text{V2} \\
\text{C} \\
\text{V1} \\
\end{array} \]

\[ \text{Anticipatory co-articulation} \]

From (5) it could be observed that six possible areas of articulation could be analysed: V1 to C (3), C to V2 (2), V1 to V2 (1), V2 to C (4), C to V1 (5) and V2 to V1 (6). When a vowel influences another vowel (V1 to V2 and V2 to V1), the process is referred to as vowel-to-vowel co-articulation. When a vowel influences a consonant, (V1 to C and V2 to C), the process is referred to as vowel-to-consonant co-articulation, and when a consonant influences a vowel (C to V2 and C to V1), the process is referred to as consonant-to-vowel co-articulation.
In the past three decades many models have been suggested to explain and predict co-articulation. Some of these are presented below.

3.6.1 Accounts of co-articulation

All accounts of co-articulation acknowledge the effects of co-articulation, i.e. modification of segments by neighbouring ones. The problem has been how this modification comes about. There have been a number of different theoretical accounts proposed for the presence of co-articulation in speech, and these have been addressed from a wide variety of angles. Some of the theories are based on the adaptation of speech organs to the communicative situation ('speech economy'), some on the notion that co-articulation is a result of the spreading and modification of the features of segments (feature spreading model), and yet others on the co-ordinative behaviour of articulatory organs during speech (the co-production or gestural model). These three approaches will be briefly discussed in this subsection.

According to the speech economy view, co-articulation in speech is a function of the communicative situation, which is almost always different. Speech organs therefore have to continuously adapt to these varied communication situations, resulting in variation in the production of speech sounds. Sounds may be over articulated (hyper-speech), leading to more perceptual contrast and less co-articulation, or they may be under articulated (hypo-speech), leading to less perceptual contrast and more co-articulation. Thus the acoustic properties of the same speech sounds may vary, depending on the communicative context. This is the basis of Lindblom's Hyper-Hypo Speech theory.

In the hypo pole of the theory, speech organs economically reduce their movements in producing sounds as they respond to context, resulting in more variation in the production of sounds (more co-articulation). In other words, because of this principle of economy, speech organs never really reach their targeted position for any particular sound, since they have to respond to motor commands for the adjacent sound in the course of producing another one.

Experiments in support of this view include that of Lindblom (1963), who analysed vowel reduction in English CVc utterances. He reasoned that in an ideal context free environment, acoustic targets for the vowels would be realised. His study, however, showed that acoustic targets for the vowels were never really reached. The targeted formant frequency values for these vowels often changed at mid-vowel point depending on the duration of the vowel. Target values were
Chapter Three. Stops: VOT and F0 perturbation, phonological representation and issues of co-articulation

reached when the vowels were long, but there was target undershoot when the vowels were short, with formant frequencies tending to shift at mid-vowel point from their movements towards target values for the vowels towards their values for the neighbouring consonant. This different realisation of vowels according to their duration indicated that vowel reduction was more of an articulatory process than a phonological one. In other words, vowel reduction happens because speech organs are instructed to produce the adjacent sound when they are still in the course of producing another, and when a single articulator is given consecutive commands within a short time distance, it has to economise, resulting in it not reaching its targeted position for sounds.

Lindblom, Pauli and Sundberg (1975) performed another experiment to show that economy in speech production was the norm. They experimented with apical consonants in VCV contexts and found that, whilst the tongue maintained closure at the front part of the mouth, it adopted a shape more compatible with least displacement from flanking vowels. This supports Lindblom’s view that coarticulation results when speech moves more in the direction of the hypo-pole of speech production, where under-articulation of speech sounds results in more co-articulation (Lindblom 1990, see also Farnetani 1997:383).

In Moon and Lindblom (1989)”s revised study of under-articulation, other factors are taken into consideration in the study of vowel undershoot. These include the rate of formant frequency change and speech rate. Moon and Lindblom performed an experiment on stressed vowels in American English produced in isolation and within a frame sentence. They observed results which appear to be opposed to those observed in the 1963 study. For instance, they observed that target formant frequency values for the vowels tended to be realised when they were produced in normal speech rather than in citation form where there was vowel undershoot, although normal speech style tended to exhibit larger formant velocity values than citation style — a situation which would be thought to reduce vowel duration and facilitate vowel undershoot.11 Realising formant frequency targets for the vowels in normal speech rather than in citation meant ‘...that for a given duration, the degree of context-dependent undershoot depends on speech style and tends to decrease with an increase in the speed of the articulatory movements’ (Farnetani 1997:383). Thus, it now appears that vowel undershoot seems to be a function of speech style also. In addition, another factor which may contribute to vowel reduction, apart from vowel duration and speech style, is force of utterance.

11 It is worth mentioning that other studies (e.g. Nord, 1986) have shown that vowel reduction does occur at slow speaking rate too.
In the feature spreading account, promoted by Daniloff and Hammarberg (1973) and Hammarberg (1976), co-articulation, far from being a by-product of constraints imposed on the speech organs during articulation, rather belongs in the phonological component of a particular language's grammar (Hammarberg 1976), and is a phonological process. This view is based on the premise that phonetics (articulation/production of segments) comes second to phonology (abstract entities/segments), and speech segments are abstract phonological entities which cannot be altered in any way by vocal organs. They, however, have phonological features and derived properties, where phonological features are the inherent properties of the segments. It is the phonological features of segments which can be manipulated, altered or modified by the speech apparatus, giving rise to derived properties of segments. The process operates in this way. The phonological rules of the grammar dictate which phonological properties may be manipulated or altered. The output of these rules is a phonetic representation, which is then fed into the speech apparatus, and in this way dictate the appropriate information regarding details of articulation and co-articulation. Hammarberg (1976) writes:

Given this view of the relationship between phonology and phonetics, and given that the phonology is indispensable when it comes to accounting for coarticulation, the simplest view of coarticulatory phenomena would be that they are creatures of the phonology, i.e. of the programming phase. That is, the phonology would specify all (relevant) details of the articulation of a segment, including the coarticulatory details, and the phonetics would simply perform what the phonology has stipulated, without adding any further properties to the articulation of each segment. (Hammarberg 1976:359).

Experimental investigation used in support of this model starts with that of Henke (1966) on anticipatory labial co-articulation in Russian. The domain within which co-articulation occurred was the CnV-syllable string type. In Henke's model, at the input level, articulatory targets are assigned to speech segments by means of binary phonological features [±]. Segments which do not have an articulatory specification are assigned the value 0 since they do not have any articulatory target and are therefore considered to be neutral (Moll & Daniloff 1971). In the given domain of co-articulation, all preceding unspecified segments are then given a feature of a specified (following) segment in an anticipatory fashion. In the event that unspecified segment(s) also follow the specified one, the feature of a specified one spreads onto those segments too. The spread of a feature may only be blocked by a segment which has its own inherent articulatory specification.
Thus, co-articulation, particularly anticipatory co-articulation, 'is a deliberate phonological process' (Farnetani 1997:385).

Other experiments include those of Daniloff and Moll (1968) and Moll and Daniloff (1971) in American English. These researchers have shown that co-articulation movements for anticipatory lip protrusion could actually start approximately two segments prior to the segment specified [+ round] and velum lowering may be triggered at the start of the first vowel in CVVN sequences. Many other studies have been conducted to investigate the co-articulatory behaviour of different organs of speech (e.g. velum lowering, tongue body movement, lip protrusion, etc.) in a wide variety of languages. Many of these are mentioned and some discussed in Shin (1997:21-35).

A number of shortcomings have been pointed out against the feature-spreading model. The co-articulation resistance theory points out some of these shortcomings, and is discussed in detail below. It has also been observed that co-articulation can still modify segments with contradictory feature specification, rather than the specified segments blocking the spreading of a feature, as the feature spreading model assumes. For instance, Butcher and Weiher (1976), on German, showed that the displacement of the tongue towards a vowel can commence during the preceding transconsonantal vowel in the asymmetric sequence type V₁CV₂. Thus V₂ in the analysis of vowel-to-vowel co-articulation is still modified (by V₁), inspite of it having its own articulatory goal. Some unspecified segments in some languages or even within languages have also been found to resist co-articulation rather than categorically acquiring the feature of a specified segment, as the feature spreading model predicts. In a study of the movement of the velum in the production of nasal and juxtaposed non-nasal sounds in the contexts NV and NVN (where N represents nasal and V represents vowel) in Japanese, Ushijima and Sawashima (1972) (see also Farnetani (1997:386)) showed that the velum stayed as low for the vowel as it did for the nasals in the NVN contexts, indicating that the oral vowel had acquired a contextual feature [+nasal]. In the context NV, however, the velum was low for the nasal but rising during the vowel. This indicated that the oral vowel, specified as [-nasal], fails to categorically acquire the contextual feature [+nasal] in the context NV.

Farnetani (1997:388) also showed that Italian consonants, unspecified for tongue body features, co-articulated to different degrees with various neighbouring vowels: /i/ to /i/ and /a/ to /a/, where C was /p/, /t/, /d/, /z/, /s/ and /l/, rather than behaving in the same way according to context, as the feature-spreading model proposes. Consider figure 3.3. In the iCi context, most of the
Figure 12.7 Tongue body EPG contact at various points in time during symmetric (C)V(C)V isolated words in Italian. VI, V2 correspond to vowel mid-points; T1, T2 to offset of VI and onset of V2 respectively; C1, C2 correspond to onset and release of consonant closures/constrictions respectively; Max refers to the point of maximum contact during the consonant (from Farnetani, 1991).
consonants showed moderate deviations from the i-like configuration of the tongue body, with the exception of /z/, which manifested considerable deviation from this tongue position. In the aCa context, /p/ and /l/ strongly co-articulate with the vowels, with /l/ showing the most co-articulation, whilst the other consonants showed deviations of varying extent, with /ʃ/ showing the largest displacement of the tongue. The fricatives showed the most resistance to co-articulation, followed by stops and lastly the liquids. Thus although all of these consonants are unspecified for tongue body features, they do not behave in an identical way as the environment dictates, at least in Italian.

Figure 3.3: The co-articulatory behaviour of Italian consonants in the iCi and aCa environments. Adapted from Farnetani (1997:389).
Nevertheless, different languages seem to display language particular characteristic trends favouring or disfavouring co-articulation, and additionally, the influence of the same sound on the process of coarticulation seems to be different from language to language (see Farnetani 1997:386, and references cited therein).

The feature-spreading model could not account for all of the observations made above. Also, the feature-spreading model was unable to account for other variation observed across other contexts, e.g. the lack of aspiration in stops preceded by the sibilant /s/ in English is one of the shortcomings often mentioned. Additionally, whereas the model rejects a physiological explanation for anticipatory co-articulation in preference of a phonological one, it resorts to the physiological one for the carryover co-articulation, which is explained as occurring partially due to constraints imposed on the speech mechanism and partially due to ‘a feed-back assisted strategy that accommodates speech segments to each other’ (Ibid.). This model is also largely biased towards anticipatory co-articulation at the expense of many carryover co-articulation observed in a number of languages, e.g. Italian (Recasens 1989), and other accounts not based on the spreading of features were consequently developed, e.g. the window model of Keating.

Keating’s account of co-articulation attempts to account for intersegmental spatial and temporal changes in speech as a function of context (e.g. the different co-articulatory behaviour of the allophones of the phoneme /l/ in English [l], [l] and [I] in a VCV environment), where in one study (Bladon & Al-Bamemi (1976) clear [l] showed the most co-articulation, syllabic dark [l] showing the least co-articulation and dark [l] coming somewhat in the middle-this is discussed below, and for cross-language variation in co-articulation. Keating’s approach is discussed briefly here, and a detailed discussion may be found in Farnetani (1990:390-393, and Shin 1997:28-35). Regarding cross-language variation in co-articulation, Keating (1985, 1990b) points out that this is different in different languages, and may well be taken care of by the phonetic rules in the grammar of individual languages and not by the universal phonological rules of the feature-spreading model. An example could be the different vowel duration before voiced and voiceless consonants. In English vowels are systematically long before voiced stops and short before voiceless stops. In other languages, like Polish and Czech, a West Slavic language, this variation in vowel duration is not so systematic and may even be absent. This, according to Keating, is the reason why each language should specify such phonetic details in its own grammar, since these details are not always universal.
Keating (1990a) proposes a window model to account for co-articulation. This is shown in figure 3.4.

Here, windows are defined as representing all possible contextual variation in the production of a segment, and they have their own duration and widths. The amount of contextual variation of a segment dictates what size width a window may have. Segments exhibiting little contextual variability (as in the case of specified features) have a narrower width than those showing more contextual variation (as in the case of unspecified features). Adjacent windows are linked to each other by means of contour lines, where the contour lines give information about the articulatory (and/or acoustic) variation of a segment over time within a particular context. These contour lines are governed by 'smoothness and minimal articulatory effort.' To give examples, in a sequence like VCV, an intervening
Chapter Three. Stops: VOT and F0 perturbation, phonological representation and issues of co-articulation

window for the C (stop) with a wide width (indicating that it is unspecified for tongue body position) allows for direct interaction between the preceding and the following (vowel) segments. Thus vowel-to-vowel co-articulation occurs with very minimal intervention from the intervening stop. This may be illustrated by Figure 3.4(c). If, however, the intervening consonant is specified for tongue body position, then its associated window width will be narrow, and will inhibit quick interaction between flanking vowels. Consider Figure 3.4(d) and the left most part of (a).

As we saw earlier, in this model, phonetic variation across languages occurs when the phonetic feature of a segment is interpreted differently in different languages (and thus belongs to the grammars of individual languages). This is what causes the differences in window width across languages. Similarly, phonological rules may operate differently in different languages, leading to phonological variation across languages. Thus cross-language variations in co-articulation exist at both phonetic and phonological levels.

One of the criticisms levelled against the window model is the fact that the theory is only limited to co-articulation within the segmental phonetic context. Co-articulation due to other factors, e.g. speaking rate is not accounted for. Also, ‘windows have no internal temporal structure allowing them to stretch or compress in time,’ (Farnetani 1997:393). The widths are rather limited to representing all possible contextual variations only. At any rate, the model would become rather weak if window widths associated with phonological features could also stretch or compress due to factors other than the phonological features with which they are associated.

The third theory of co-articulation, the co-production model, is based on the gestures of articulatory organs during speech. In this model, it is argued that the physical execution of speech sounds should be separated from their abstract phonological properties (the basis for the feature spreading theory). The model proposes that the physical production of sounds uses units (spatial targets and duration) which are substantially different from abstract linguistic information about those sounds. These are phonetic gestures, where a gesture is defined as ‘a member of a family of functionally equivalent articulatory movement patterns that are actively controlled with reference to a given speech-related goal (Fowler & Saltzman 1993:192, see also, Saltzman & Munhall (1989:334).’

Fowler & Saltzman (1993:172) also define gestures in linguistic, and physical terms as follows. ‘... phonetic gestures are linguistically significant actions of structures of the vocal tract; alternatively, the term is restricted to the control structures that generate those actions. Linguistically, gestures have been “hypothesized to be the primitives of a
As far as co-articulation is concerned, it is argued that speech organs — far from being unable to reach their target positions for adjacent segments as the ‘speech economy’ theory proposes, or rather than being the result of a scanning ahead strategy as the feature spreading model suggests — simply overlap with one another in time, co-producing juxtaposed sounds in turn (e.g. Fowler 1977; Fowler & Saltzman 1993). As pointed out above, the feature spreading theory accounts for co-articulation as a result of one segment modifying another segment in a context-sensitive manner by way of underlying linguistic features. In the gesture model, it is argued that the modification of segments by others in the process of co-articulation does not happen in a linguistic sense. Rather, segments have underlying linguistic identities which are defined free of context. At the level of articulation, however, there is *temporal* overlap when juxtaposed segments are produced, with context-sensitivity being only a concomitant of the dynamics of the movements of the organs of speech during co-production.

In the gesture theory, the articulatory effects of co-production may differ depending on the degree of spatial overlap on the gestures to be produced. Where there is an incomplete spatial overlap, articulators are not completely shared, and there is minimal influence between phonetic gestures and therefore minimum co-articulation. In this case, ‘gestures are defined along different sets of tract variables,’ and the gestures have no, or some but not all, articulators in common’ (Fowler & Saltzman 1993:181). Where co-produced gestures are defined along the same sets of tract variables, they share model articulators, creating a relatively greater possibility for mutual influence between phonetic gestures and therefore for co-articulation. In other words, ‘the highest degree of spatial overlap occurs when two overlapping gestures share the articulators directly involved in the production of gestural goals, and impose competing demands on them’ (Farnetani language-user’s phonological system....Physically, they are ... co-ordinated movements of the vocal tract that achieve a phonetically significant goal.’

Fowler & Saltzman (1993:178) give the following tract variables used by the model, which they adapted from Saltzman & Munhall (1989).

<table>
<thead>
<tr>
<th>Tract variables</th>
<th>Model articulators</th>
</tr>
</thead>
<tbody>
<tr>
<td>LP lip protrusion</td>
<td>upper &amp; lower lips</td>
</tr>
<tr>
<td>LA lip aperture</td>
<td>upper &amp; lower lips, jaw</td>
</tr>
<tr>
<td>TDCL tongue dorsum constriction location</td>
<td>tongue body, jaw</td>
</tr>
<tr>
<td>TDCD tongue dorsum constriction degree</td>
<td>tongue body, jaw</td>
</tr>
<tr>
<td>LTH lower tooth height</td>
<td>jaw</td>
</tr>
<tr>
<td>TTCL tongue tip constriction location</td>
<td>tongue tip, body, jaw</td>
</tr>
<tr>
<td>VEL velic aperture</td>
<td>velum</td>
</tr>
<tr>
<td>GLO glottal aperture</td>
<td>glottis</td>
</tr>
</tbody>
</table>
Chapter Three. Stops: VOT and F0 perturbation, phonological representation and issues of co-articulation

1997:395). This competition is dealt with in the model by way of a blending of gestures. Phonetic gestures have their own inherent degree of blending strength (specified context independently). Generally, the degree of blending strength has an inverse relationship with the sonority characteristics of the segment in question. Vowels, which are the most sonorous sounds, have the weakest blending strength, and stops, which are the least sonorous sounds, have the strongest degree of blending strength. Gestures of equal blending strength show an averaging of both (Fowler & Saltzman 1993:182). Where a segment with a weak degree of blending is co-produced with one with a weak blending strength, the influence of a stronger gesture is more than that of a weaker gesture, and thus stronger gestures show some resistance to being contextually modified whilst they themselves exert some influence on other gestures. In this way the gesture model captures the issue of phonetic aggression of a segment on exerting influence over other gestures, and the theory of co-articulatory resistance (presented in section 3.7 below).

A phonetic gesture (or segment)'s contextual influence also happens within a time span or 'time courses,' and is known as a gesture's activation period. In the model, this is addressed in the domain of the 'speech plan,' which is defined as 'the set of activation waveforms for the gestures in the utterance' (pg. 183), where activation waveforms have a shape and duration. Activation waveforms are thought to be smoothly shaped, with an onset phase, maximal and waning phases. The onset phase represents the start of the waveform, and the waning phase the end. Both phases exert very minimal interference on the shape of the vocal tract, which may or may not be detected. Also, during the phases, the influences of neighbouring gestures are at their height. The maximal phase is the most significant one in that it represents the period during which the gesture's attempt to influence the shape of the vocal tract is very strong. Consider figure 3.5 (after Farnetani 1997:395, adapted from Fowler and Saltzman 1993).
Thus, both the blending strength and relative activation value of a segment contribute in influencing the shape of the vocal tract during co-production.

With regard to the duration of activation waveforms, co-articulation, particularly at the anticipatory field of gesture 2 (cf. fig. 3.5), is thought to be time locked to the segment in question, and does not extend far back in time to preceding segments as some theories, e.g. the feature spreading theory, have claimed. As discussed above, there have been a number of instrumental experiments in support of the look ahead theory, most of which have been on American English. One such study was conducted by Moll and Daniloff (1971). These researchers studied anticipatory velum lowering for a nasal consonant as well as velopharyngeal port opening in CVVN sequences by means of X-ray motions using four American male speakers. They observed that velum lowering may be triggered at the start of the first vowel in CVVN contexts.

However, Bell-Berti (1980)'s observations on velum lowering showed that vowels and oral consonants have their own inherent characteristics with respect to velum height: vowels characteristically have lower velum height than oral consonants. This meant that investigations of velum lowering in strings of vowels preceding a nasal consonant have to take this aspect into consideration. Bell-Berti (1980) also observed that anticipatory velum lowering for a nasal consonant
began at a point just before the start of the nasal. These observations contradicted the theory of the feature spreading model above (cf. Moll and DanilloF (1971)), which predicted that that velum lowering may be triggered at the start of the first vowel in CVVN contexts. Bell-Berti (1980) argued that the velum lowered for the vowels even in the absence of a nasal consonant, suggesting that the actual anticipatory lowering of the nasal consonant was in fact shorter in duration than the look ahead (feature spreading) model appears to predict.

A similar study was conducted by Bell-Berti and Krakow (1991) on anticipatory lowering of the larynx in CVn sequences, where the duration of the string was manipulated by changing the number of vowel segments and the rate of speaking. The CVn string was followed by a nasal consonant in some tokens and by an oral consonant in others, with the latter sequence functioning as the control, and the former as test utterances. Like Bell-Berti (1980), (see also Bell-Berti & Harris (1981)), they observed that the velum showed a two-stage lowering pattern in CVn strings followed by nasal consonants, but showed only a one-stage lowering behaviour in control sequences. The first stage observed in these test utterances was also observed in the control utterances, and could only be logically attributed to the vowels, since these were the only segments shared by the two types of utterances. The two-stage was only observed in CVn sequences followed by a nasal consonant, and the velum lowering movement could only be attributed to the nasal consonant. This strongly suggested that velum lowering for the nasal was time locked to the nasal consonant. In other words, it did not spread from the nasal onto the preceding vowels in a scan ahead fashion, as the feature-spreading model predicts.

They also noted that the lowering of the velum was strongly correlated with the duration of the vocalic string: velum lowering was biphasic for longer strings and slower speaking rate and monophasic for shorter strings and faster speaking rate. From this they predicted a co-ordinated overlap between the gestures in the production of the CVn + nasal strings: when the vocalic string was short, then the velum lowering associated with the nasal consonant will be predominant, completely overlapping that associated with the vowel(s), thereby rendering a one-phased lowering. When the string was long, however, the two stages of velum lowering are manifested. This further strengthens the gesture theory that anticipatory movements in co-articulation are shorter in duration, and not very extensive as the feature spreading theory proposes.

In summary, three accounts of co-articulation have been discussed above. The discussions focussed on the different approaches of the models in accounting for co-articulation. Due to the fact that the experiments performed to support these
Chapter Three. Stops: VOT and F0 perturbation, phonological representation and issues of co-articulation

theories were done on a limited number of languages, it is still not possible to determine whether the results obtained for these experiments reflect language-particular tendencies or characteristics common to the languages of the world. For this to be determined, more cross-linguistic analysis needs to be made. Also, the 'speech economy' and co-production theories do not address the segment-specific characteristic behaviour of sounds in relation to co-articulation. Keating addresses this by way of 'windows' — where the width of a window is determined by the possible extent to which a segment accommodates the influence of neighbouring sounds. Thus the exact size of a window width for a segment is left to thorough investigation of that particular segment. All the three theories reviewed above have not considered the different magnitude of co-articulation between segments or allophones of a segment in a language. This is the subject of co-articulatory resistance discussed below.

Aspects of co-articulation with respect to Shekgalagari stops are investigated in Chapter Six, which attempts to examine lingual articulatory characteristics of stops in the language by means of EPG. Both qualitative and quantitative analysis is made on the type and extent of articulatory movements for Shekgalagari stops. The findings of the study are discussed in relation to some of the accounts of co-articulation presented above, particularly the feature-spreading model and the window model of Keating, and also in relation to the theory of co-articulatory resistance discussed below.

3.7 Co-articulatory resistance

This theory seems to be based on two different assumptions: firstly; that the co-articulatory behaviour of any given segment may differ from language to language and/or dialect to dialect (Bladon & Al-Bamerni (1976); secondly, that the lingual palatal contact area of a segment may be inversely related to the degree of co-articulation it exhibits (Recasens, e.g. 1987, 1990, 1991). According to the first assumption, to what extent a given segment in a language or dialect may co-articulate with neighbouring sounds may be specified for that segment, and may even be represented in the grammar of the language concerned. According to the second hypothesis, the co-articulatory resistance behaviour of a segment in any language of the world may be predicted from its lingual-palatal contact area. This seems to suggest universality of this aspect, and that this co-articulatory resistance does not need to be specified in the grammar of any individual language.
Concerning the first hypothesis, a study by Bladon & Al-Bamemi (1976) on the allophones of the phoneme /I/ in English showed that these allophones behaved differently in a VCV environment, with clear [I] showing more accommodation to the influence of flanking sounds, syllabic dark [I] showing the least accommodation of influence from other neighbouring sounds and dark [I] somewhat in the middle. This graded nature of co-articulation resistance for these allophones is also observed in relation to co-articulated voicelessness, and is not addressed in both the 'speech economy' theory and the feature-spreading model. According to Bladon & Al-Bamemi, because they resisted influence from neighbouring segments to different degrees, these allophones may therefore be given different co-articulatory resistance values.

Recasens is also a proponent of the co-articulation resistance theory as mentioned above. But his version – based on EPG and acoustic studies on Spanish and Catalan, however, is based on the relationship between the lingual palatal contact area of a segment and the degree of co-articulation it exhibits (Recasens, e.g. 1984b, 1987, 1990, 1991). Recasens (1984b) studied vowel-to-consonant co-articulation in VCV sets by means of EPG in Catalan, where V = /i/, /a/, and /u/, and C = /j/, /ɲ/ /ʎ/ and /n/, and observed that tongue-to-palate contact area was different for the consonants. The consonants ranked in this order: /j/ > /ɲ/ > /ʎ/ > /n/, where > means vowel-to-consonant co-articulation is greater than. Recasens also observed that the lingual palatal contact area of the consonants and their susceptibility to co-articulation appeared to be inversely related to each other. Segments with larger contact area appeared to show less susceptibility to co-articulation and vice versa. Recasens therefore proposed that the larger contact area a segment has the more resistance to co-articulation it will show, and appeared to suggest that this could be applied to the languages of the world.

The gesture theory also attempts to explain co-articulatory resistance in the following way. At the level of the speech plan, a gesture’s attempt to influence the shape of the vocal tract may be relatively context free; contextual modification at the level of articulation and acoustics would principally be determined by how a gesture blends its influence on the one being produced (Fowler & Saltzman 1993:180). In other words, a phonetic gesture’s ability to exert influence on the shape of the vocal depends on its own degree of co-articulatory aggression and on the extent of co-articulatory resistance shown by the gesture in progress.
3.8 Summary

There have been many theories on the specification and representation of stop voicing in the world's languages and the focus of these theories have been varied; none of them has been all sufficient to account for the sounds of the world's languages. Studies on stops have similarly been conducted from different angles: physiology, acoustics and auditory, and there have been language specific studies and cross-language studies. Stops may be affected by phonetic environment, and may therefore co-articulate with flanking sounds. The definition of co-articulation is still largely controversial because of its similarity to assimilation. Coarticulation may refer to a point in time when the configuration of the vocal tract is influenced by more than one sound, a definition which may also be relevant to assimilation. Coarticulation has been observed in most, in not all of the languages so far investigated, and seems to differ from language to language. Theoretical accounts proposed for the presence of co-articulation in speech have also been addressed from a wide variety of angles. These include the speech economy theory, the feature spreading theory, gesture model and the co-articulation resistance theory.
Part two:
Experimental study of Shekgalagari stops
4 Acoustics: Description and classification of speech and electrolaryngograph waveforms

4.1 Introduction

The aim of the present chapter is to provide a qualitative description and classification of the waveforms obtained for Shekgalagari stops. The qualitative description will be based on visual inspection of the properties of Gx, Lx, Sp signals and the spectrogram for each voice type; and the classification based on the general similarities and differences emerging from the analysis of the signals. The main objectives of the investigation are as follows. To find out: on the basis of the waveform types, the category of stop voicing contrast that Shekgalagari belongs to; whether Shekgalagari has 'ejectives' or plain voiceless unaspirated stops; what the language-specific phonetic characteristics of its stop system are and how they differ from, or are similar to, the phonetic realisations of stops in other languages reported in the literature — e.g. English, Xhosa, Korean Hausa and Thai.

The structure of the rest of this chapter is as follows. Section 4.2 discusses the speech signals to be described for the stops in this chapter. Section 4.3 deals specifically with the Gx's depiction of larynx movement in the production of stops in Xhosa, a South-eastern Bantu language of the Nguni group which is thought to have ejectives, implosives and pulmonic sounds in its stop system. This was considered necessary for the assessment of the level of confidence one may have in interpreting the Gx waveforms in the Shekgalagari data. Characteristics of the Lx signals are also discussed in this subsection. Section 4.4 deals with the procedure of the recording and the processing of the data. Section 4.5 deals with the processing of the waveforms for Shekgalagari stops. Section 4.6 discusses the results of the experiment whilst sections 4.7 and 4.8 respectively provide a discussion of the results of the study and a summary of the whole chapter.

4.2 Properties of the Gx, Lx, Sp signals and of the spectrogram: background information

The Gx signal maps out conductance change, essentially at the level of the thyroid cartilage, as the larynx is displaced. In speech, this displacement of the larynx may be caused by a number of factors: the glottalic airstream mechanism; the voicing structure of the segment being investigated (e.g. voiced/voiceless); the quality of the adjacent vowel and the suprasegmental dimensions of the speech being
analysed, e.g. high vs. low tone. The last three of these are less well understood, and are still the subject of further research.

The glottalic airstream mechanism is initiated at the larynx and involves abrupt displacement of the larynx. This displacement can either be upward, as in the production of ejectives, or downward, as in the production of implosives. Since pulmonic airflow is not involved in this mechanism, the whole impression is that of the larynx acting as a piston in the trachea – pharyngeal conduit (cf. Appendix 3C for detailed discussion). This gross displacement of the larynx may be non-invasively studied by means of ‘a direct-current connected version of the laryngograph providing a gross movement waveform, Gx’ (Fourcin & Abberton 1977:313).

There are a number of theories which have been proposed to account for the interaction between the voicing of a stop and the displacement of the larynx. According to one theory which is based on aerodynamic reasons, the larynx is often lowered during the occlusion for the voiced pulmonic egressive stops. The reason for this being that during the stop closure, the larynx is lowered to increases the volume of the vocal tract, which helps to sustain adequate transglottal air pressure required for voicing. The lowering of the larynx in turn leads to low pitch in the following vowel, as was discussed in section 3.2.2 in Chapter Three. For the voiceless stops, the larynx is thought to stay in a raised position (Hombert 1978, Clark & Yallop 1990:282). Another theory, based on vocal fold tension, proposes that vocal fold stiffness is different during the closures for voiced and voiceless stops: slack for voiced stops — leading to pitch drop during the stop closure (and the initial portions of the adjacent vowel), and stiff for voiceless stops — leading to pitch rise during the stop occlusion (and the initial portions of the adjacent vowel). It has been observed that, in most languages, it is the voiceless stops which have a notable effect on the pitch of at least the initial portion of the following vowel compared to the voiced stops. These stops are thought to raise the pitch of the adjacent vowel and therefore the position of the larynx (Clark & Yallop 1990:282). But, as said earlier, the effect of the voicing structure of stops on the following vowel and therefore on the position of the larynx is still a matter of investigation. It nevertheless has potential to influence the Gx signal.

With respect to vowel quality, it is thought that there is an interaction between vowel height and pitch: high vowels have intrinsic high pitch, and therefore a raised larynx, and low vowels have intrinsic low pitch, and therefore a lowered larynx. A number of hypotheses have been formulated to account for this. Two are discussed here. According to hypothesis one, high vowels narrow the vocal tract. This in turn ‘causes an acoustic loading of the vocal folds, which
Chapter 4: Acoustics: Description and classification of speech and 95 electrolaryngograph waveforms

means that $F_o$ tends to be pulled towards the $F_i$ of the vowel, and this may have an influence of raising the larynx. According to hypothesis two, as (especially the back of) the tongue is pulled up towards the roof of the mouth it also raises the larynx as well as affecting any on-going phonation accordingly (Clark & Yallop 1990:284 and references therein). However, other studies (e.g. Ladd & Silverman 1984, and Silverman 1984) have suggested that the effect of vowel identity on larynx height is quite minimal, especially in running speech.

In a similar vein, for languages with different tone types, the larynx may be thought to move upwards for high tones and downwards for low tones, and this may affect the Gx signal accordingly. Although to study the effect of segment voicing, vowel identity, and tone type would be desirable, these will not be the main focus of this study since they constitute objects of study in their own right, although, where necessary, they will also be discussed. They are, however, recommended for future research. In this analysis we will focus mainly on the effect of the airstream mechanisms (the pulmonic and the glottalic) on larynx movement.

There are other activities which are not related to speech production, but which might affect the form of the Gx. Activities like swallowing or moving the head or neck can contribute to the Gx waveform (cf. Appendix 1A). Unfortunately, such things contribute to the difficulty of interpreting the waveform.

As mentioned earlier, the Gx signal records conductance change as the larynx is displaced. However, the direction of the signal and the direction of larynx movement have not yet been shown to be directly and reliably correlated. Nevertheless, we may still consider upward movement of the Gx signal for upward displacement of the larynx and downward movement of the signal for downward displacement of the larynx as a working hypothesis (Fourcin 1999, personal communication, see figures 4.2 through 4.4 also).

The Lx signal maps out conductance change across the vocal folds and provides useful, rather detailed information about the nature of vocal fold activity during speech production in a non-invasive manner (see Appendix 1A). The shapes of the Lx traces can give valuable information about the nature of vocal fold contact and the regularity of vocal fold vibration during phonation. The relative size of the traces can also provide information about the amplitude of vocal fold vibration. The Lx waveform can also provide information about the frequency of

---

1 It must be borne in mind that when the larynx is displaced, all the associated muscles, which are in fact responsible for displacing the larynx, are also displaced. The Gx signal therefore may be considered to reflect the displacement of the larynx AND the associated muscles.
the vibration of the folds, and can be processed to provide fundamental frequency contours. For more detailed discussion, see Appendix 1A.

The Sp signal gives information about activity in the vocal tract as a response to the action in the larynx. This signal is a result of larynx activity and vocal tract activity. It can record a burst for the release of a stop, aspiration for aspirated stops and even show some aspiration on the Sp resonances in the initial part of the vowel. The spectrogram also gives acoustic information, and is an excellent depiction of the various dynamic changes in the resonances of the vocal tract. It can effectively record a voice bar for stops which are produced with voicing and gaps for stops without voicing; a spike for the release burst of a stop; aperiodic noise associated with aspiration with the accompanying apparent attenuation of the first formant being clearly visible; prominent and lengthy bands of energy associated with vowel formants as well as formant transitions (see Appendix 1A for detailed discussion).

With the suite of programs employed in this study, the Gx, Lx and Sp signals and the spectrogram are usually displayed on the monitor with the Gx, Lx and Sp waveforms displayed as time and amplitude representations of the corresponding signals — with amplitude displayed as a function of time. The spectrogram is displayed as a time and frequency representation of the speech signal — with the frequency displayed as a function of time. The energy distribution is shown on the spectrogram as the intensity of the colour of the pattern.

In the waveforms described in this chapter and shown in figure 4.1, the topmost plot is the speech signal (Sp). The spectrogram (Spect.) is shown directly below the Sp signal, and is followed by the Lx waveform, which is the third signal on the figure. At the very bottom of the display is the Gx signal — on figure 4.1 shown with large gain. Since the Gx signal is a low pass filtered version of Lx and therefore additionally shows Lx activity, it is shown in this study as Gx/Lx.
Chapter 4: Acoustics: Description and classification of speech and electrolaryngograph waveforms

4.3 The Gx: An exploration with Xhosa stops.

In order to test the reliability of and the confidence we may have in the relationship between the direction of Gx and larynx displacement in our data, and before examining the possible ejectives in Shekgalagari, it was thought worthwhile to examine its behaviour in stops whose production certainly involved the displacement of the larynx in different directions. Therefore, with the help of an informant, a short list of Xhosa words (21 words in all) was generated and recorded for investigation. The Xhosa speaker was a student of law and had been studying in London for one academic year. She came from South Africa. Data recorded by this speaker was collected from her, but she was kept ignorant as to the purpose of the experiment and encouraged to produce the words in as natural a way as possible during the recording. The words were produced in isolation and within a frame sentence [ngoko ndit'i X] ‘Now I say X’. They were produced at the bilabial, alveolar and velar places of articulation. The data may be found in Appendix 4D.

Xhosa is thought to have ejectives, as [p’t] in for instance, “pená-pená” [p’enáp’:rä] ‘throb with pain’; and implosives, as [ɓ] in for instance “bôna” [ɓô:na] ‘see’; as well as simple, plain pulmonic consonants, as the egressive voiced [b] in for example “bhembêtha” [bebêtha] ‘struggle/suffer’; in its stop system. These three stop types use different airstream mechanisms which displace the larynx in different directions. Ejectives and implosives both use the glottalic airstream mechanism in which the larynx is displaced — upward in the case of the ejective [p’t] in [p’enáp’:rä], and downward in the case of the implosive [ɓ] in
Chapter 4: Acoustics: Description and classification of speech and electrolaryngograph waveforms

[60:na]. In the word [bebe:tʰa], the plosive [b] is produced on the pulmonic airstream mechanism which does not involve the displacement of the larynx.

Figures 4.2 through 4.4, show Gx, Lx, Sp signals and the spectrogram for the ejective, implosive and pulmonic stops respectively from different Xhosa words. The description of the signals here is based on the words produced in citation form and within a frame sentence. Figure 4.2 (a) shows the word [p’enáp’e:ná] produced in isolation.

It can be observed that in both instances when the ejective [p’] is produced, the Gx (and Lx) signal deflect upwards. We may attribute this to the upward movement of the larynx as it initiates the airstream required to produce [p’]. The Lx signal deflects upwards accordingly, since the vocal folds are also displaced as the larynx moves up.² The Gx signal stays raised for a relatively longer period of time (i.e. compared with the Lx signal). This may be correlated with the manner of articulation for ejectives where two simultaneous constrictions made at the glottis and in the oral cavity as well as the blockage of the nasal cavity means that air pressure builds up in the vocal tract. And the larynx therefore stays raised for as long as the pressure is unreleased. According to Ladefoged and Maddieson

² Lx is a high pass filtered version of the Gx. In fact as will be observed in the following sections, the Gx signal was not filtered at all in this study. This means that Lx therefore lacks the very low and steady information contained in the Gx signal, but only shares information obtained in higher frequencies (Fourcin 2000: personal communication).
(1996:57), the degree of pressure generated in the vocal tract is twice that of the normal pulmonic pressure. The downward movement of the signal(s) may be due to the fact that when the stop is released, the pressure in the larynx, along with the muscular tension in the glottis, is released and the larynx is subsequently lowered. It is interesting to observe that this behaviour of $G_x$ and $L_x$ is still observable when the word is produced within a carrier sentence. This is shown on figure 4.2 (b).

Figure 4.2 (b) shows the word [p’ěnáp’ěná] produced within a frame sentence. Again the $G_x$ waveform deflects upwards and remains raised for an appreciable period of time. We consider that this may be due to the movement of the larynx as it is suddenly raised upwards, and that it stays raised relatively longer (compared to the $L_x$ signal) because of the air pressure built up between the constrictions in the vocal tract. The $L_x$ signal shows the same upward deflection since the vocal folds are also displaced as the larynx is moved.

Characteristics of vowel onset are shown on figure 4.2 (c), for the word [p’ě.p’ě] ‘porridge’.
Chapter 4: Acoustics: Description and classification of speech and 100 electrolaryngograph waveforms

Three things can be observed on the Lx signal: irregular mode of vocal fold vibration, the decreased open quotient and the amplitude of vibration (see Appendix 1A for detailed discussion). Irregularity in the vibration of the vocal folds is shown by the shapes of the Lx traces at the start of the vowel. The first few cycles show steep rising deflections, indicating rapid closure of the vocal folds. The downward slope is by contrast more gradual, with the first, second, third and fifth, and to some extent the sixth and seventh showing somewhat flat tops. This indicates slower opening of the vocal folds in relation to the closed phase. This means that the closed phase is relatively longer than the open phase in each of these cycles, and consequently there is a decreased open quotient. This may be associated with creaky voice type of vocal fold vibration. The shapes of the subsequent traces change to a more regular mode of vibration, with the rising steepness of the traces being more rapid than the opening downward slope, and the cycles becoming periodic (although there are still some traces of irregularity in that the sizes of the traces differ, with two being noticeably smaller than the others). These traces are also noticeably bigger in size than the first six cycles, suggesting increase in amplitude of vibration. The frequency of vocal fold vibration may also be seen to increase in the subsequent cycles compared to that in the first cycles.

This irregular mode of vibration was observed for tokens produced in isolation and within a carrier sentence for most of the tokens. In the frame sentence, this state of the vocal folds could be seen in the final portion of the vowel prior to the articulation of the stop as well as after the stop had been articulated. In the word [p’ɛnáp’ɛná], these distinctive attributes were observed for the second plosive also. These characteristics were not always present all the time for all the
tokens. Sometimes only the decrease (or increase) in the amplitude of vibration was observable, or two or three long cycles were present.

The behaviour of Gx with respect to implosives was not as consistent or reliable as in the case of ejective stops because of the similarity with the voiced pulmonic stops. For the implosives vowel onset characteristics were the most reliable cue. Figure 4.3 (a) shows the behaviour of the Gx signal for the implosive [ɓ] in the word [ɓõ:na] produced in isolation.

![Figure 4.3 (a): The behaviour of the Gx signal (indicated by the arrow) in the Xhosa implosive [ɓ] in the word [ɓõ:na] 'see', produced in isolation.](image)

For the production of [ɓ] in [ɓõ:na], the distinctive movement of the Gx signal appears to happen prior to the occlusion for the stop, where direction of change in the signal is **downwards** and only moves **upwards** for and during the production of the implosive stop and its release burst. It is not easy to give a correct interpretation of the downward and upward movement of the Gx waveform in relation to the larynx. It is nevertheless interesting to note that voiced implosives, as the Xhosa [ɓ], are produced on two different airstream mechanisms: the ingressive glottalic mechanism and the pulmonic airstream mechanism (see Appendix 3C). In the ingressive airstream mechanism, the larynx is rapidly lowered (albeit only a little, see section 4.7 for discussion on the extent of larynx movement) to increase the volume of the cavity for voicing as well as creating room for an inrush of airflow. It is plausible to attribute downward Gx for [ɓõ:na] in figure 4.3 (a) to this lowering of the larynx. During the downward movement of the larynx, the vocal folds are often slightly separated, allowing pulmonic air to escape and causing phonation in the process. This counteracts the lowering of the
larynx, and upward movements are initiated, and plausibly, Gx deflects upwards.
According to Clark and Yallop (1990:57), lung air escaping through the approximated vocal folds in the production of voiced implosives can be sufficient to ‘offset the suction action of the downward larynx movement so that there is little or no inward airflow through the mouth, even to such an extent that the net airflow is actually egressive’. Similar Gx movements were observed for the same stop in the word [ɓalɛ:ka] ‘run/running’ produced within a carrier sentence. Consider figure 4.3 (b).

Figure 4.3 (b) shows the behaviour of the Gx signal for the implosive stop [ɓ] in the word [ɓalɛ:ka] ‘run/running’ produced in the frame sentence [ŋgʊkʊ ndi:tɪ bi X] ‘Now I say ḃale:ka’. Turbulence on the Sp signal and on the spectrogram is for the aspirated stop [tʰ] on the carrier sentence and this was confirmed by using the play back facilities. There is some indication of downward movement of the Gx (and Lx) prior to the articulation of the stop, followed by upward movement for the production of the stop and its release. Irregularity of vocal fold vibration at the start of the vowel can also be seen on the Lx (and Gx) waveform. But, for the other two tokens on the data, vowel onset showed modal vocal fold vibration. This can be seen on figure 4.3 (a) for [ɓo:na] and is shown clearly on figure 4.3 (c) for
Chapter 4: Acoustics: Description and classification of speech and electrolaryngograph waveforms

[6ɛtʰa] ‘hit/hitting’ where irregularity is rather seen prior to and on the initial portion of the closure for the implosive.

With respect to the pulmonic stops, for some of the tokens the Gx (and Lx) signal seemed to stay level prior to or during the articulation of the stop. Consider figure 4.4 (a) for [b] in the word [bɔːka] ‘praise’, produced in citation form. (Note that this stop, as well as the other voiced bilabial pulmonic stops recorded for analysis in this study, was produced without prevoicing).
Similar characteristics were observed when the word was produced within a frame sentence. This is shown of figure 4.4 (b) for the stop [b] in the token [bebɛ:tʰa] ‘suffer/struggle’. But, as mentioned earlier, there were some similarities in Gx behaviour between the implosives and the pulmonic voiced stops. This can be clearly seen in [bebɛ:tʰa], where Gx stays in a relatively level position for the first [b], but shows evidence of downward movement when the stop is produced in the second syllable again.
It is not easy to give an accurate correlation of the Gx waveform and the articulation of the pulmonic voiced stops. For some of the tokens, the Gx appeared to stay in a relatively level position for [b], but for other tokens, it showed some indication of downward movement. For tokens where Gx stayed in a comparatively level position, this seems to correspond with the fact that this plosive is produced with an airstream mechanism which does not involve the gross displacement of the larynx. It is also plausible to relate downward Gx with the lowering of the larynx in voiced stops to create room in the supraglottal cavities in order to sustain the transglottal pressure drop necessary for voicing (see § 4.2). This interaction of the factors which may contribute to the lowering of the larynx also makes the interpretation of the Gx trace very difficult.

Pulmonic voiced stops also showed considerable regularity in vocal fold vibration at vowel onset. Figure 4.4 (c) shows vowel characteristics after the production of the stop [b] in the word [bô:ka] ‘praise’.
It can be observed in figure 4.4 (c) that the shapes of the Lx traces from the start of the vowel show a regular mode of vibration which seems to persist through the rest of the vowel. The traces are of the same shape and size and the time between the cycles is approximately the same. For most of the voiced pulmonic stops, regularity in vocal fold vibration could also be observed prior to the stop in the carrier sentence.

On the whole, it appears from the observations above that there is some element of distinctiveness in the pattern of Gx baseline movement for the stops which are produced by displacing the larynx in different directions. For the ejectives, Gx showed upward movement for most of the tokens in the data studied here. Although implosives showed evidence of downward movement, the most reliable cue for these was irregularity of vocal fold vibration and/or gradual build-up of intensity, either prior to the stop or at vowel onset (see figures 4.3 (b) and 4.3 (c) respectively). The pulmonic voiced stops showed both a somewhat level Gx and downward Gx, but they were most reliably distinguished by regularity of vibration prior to the stop in the carrier sentence and at vowel onset. This observation may perhaps give us some confidence in interpreting the Gx signals and help in distinguishing between the stop contrasts in the Shekgalagari data.
Chapter 4: Acoustics: Description and classification of speech and electrolaryngograph waveforms

4.4 Procedure

4.4.1 Data, subjects and recording conditions

4.4.1.1 Data

The data used in this study consisted of real meaningful disyllabic words of the structure CVCV (see Appendix 4D, where the first consonant of the first syllable was the only stop under investigation (the second consonant, in the second syllable, could be any consonant in the language (including stops) and was not meant to participate in the study). Words of the form CV are very limited in the language and would not have generated enough data for study. The corpus consisting of real meaningful words was preferred since they can be produced more naturally, with correct tone patterns and vowel lengths (Lindsey et al. 1987, Haruna 1990).

Word samples consisted of all the stops in word initial position produced at five different places of articulation: bilabial, dental, palatal, velar and uvular. They were homorganic stops with different voice types followed by various vowels where the vowel could be [i, e, ə, a, o, u] since one of the concerns of the study is to investigate the effect of the following vocalic segment on the start of voicing for the vowels after the stop. The words were produced at a normal speaking rate in citation form and within a frame sentence [qare X xaitsʰi] 'I say X again'.

For words produced in isolation, the vowel following the stop under investigation is lengthened as a consequence of the penultimate lengthening phenomenon typical of Bantu languages (cf. § 2.4.3). Within a carrier sentence, however, this lengthening is reduced considerably (cf. § 2.4.3). Apart from allowing for the production of the words at a more natural conversation rate and manner and smoothing out the vocalic weight brought about when words are produced in isolation, the frame sentence was also meant to provide a phonetic environment which is voiced, i.e. where the stop concerned is surrounded by voiced sounds (vowels in this case). My 'intervocalic' stops are positioned word initially and are separated from the preceding vowel by a word boundary. Segmentally, the stop concerned is in intervocalic position but is prosodically separated from the preceding vowel. This context has been referred to as 'word initial intervocalic' position (Jessen 1997:48,55).

Each subject recorded the material (61 real words) once, yielding a total of 244 words. And the material was stored in a computer for analysis.

---

3 All stop and vowel combinations were not possible for all of these vowels.
Chapter 4: Acoustics: Description and classification of speech and electrolaryngograph waveforms

4.4.1.2 Subjects and recording conditions

Four subjects (two males and two females) were involved in this experiment and all are native speakers of Shekgalagari. They are all students in various fields of study, except one of the female speakers who is a student of phonetics at UCL and is the author of this thesis. Three of the speakers speak the Shengologa dialect of Shekgalagari which some have described as the standard dialect (Andersson & Janson 1997), although this is neither official or acknowledged by consensus. All of the speakers have no known history of speaking or hearing disorders.

The informants were seated comfortably in a chair and the electrolaryngograph electrodes were placed on either side of their thyroid cartilages and secured with an elastic band behind the neck, taking special care to ensure that they were not choked and could speak comfortably and naturally. They were then advised to try to keep their head against the head restraint in order to maintain as much as possible the 500 millimetres distance between their lips and the microphone head in front of them. The subjects were then asked to read a few words from the list presented to them prior to the recording process as the level of recording was adjusted to avoid overloading the signals or producing signals which could be too low and rather less useful during analysis. Each subject recorded the data individually in an acoustically sound-treated anechoic room at the Department of Phonetics, UCL, and an experienced technician monitored and controlled the recording process throughout.

The Gx signal from the Gx laryngograph was adjusted to 0 volts (V) (which was at the centre of the range), using the oscilloscope and monitor in the chamber. Special care was taken to try to maintain this level during the recording since this signal was meant to measure the relative displacement of the larynx in the production of different sounds. It was crucial therefore that, should the signal drift during the recording process, the recording be stopped and the signal appropriately adjusted to 0V level. In our recordings, however, the Gx signal remained relatively stable to allow a recording of approximately 1-2 minutes to be made without interruption.

This was a three-channel acquisition, using a Data translation acquisition PC card for storage on the hard disc of a PC. The signals were speech (Sp) and two laryngograph signals: the Lx; which monitors vocal fold vibration, and the Gx; which additionally monitors the gross displacement of the larynx and the associated muscles. The speech signal was recorded onto Channel 1 using a Bruel and Kjaer half-inch free field condenser microphone (cartridge 4165) in conjunction with a sound level meter (No. 2231). The Lx signal was recorded onto Channel 2. Both
of these signals were first passed through variable gain amplifiers and analogue anti-aliasing filters set to 8.2 KHz, before being applied to the data translation card. The sample rate was set to 20 KHz. The Gx signal was not filtered, but was rather fed directly to Channel 3 of the data translation card from the laryngograph. This maintained direct current (dc) coupling of the signal, which ensured that slow changes of the signal due to gross larynx movement were preserved. A block diagram of the arrangement of the recording equipment is shown in figure 4.5.
The data was processed in an acoustic laboratory using a computer and the Speech Filling System (SFS) software programs written by Mark Huckvale, UCL.

4.5 Processing

Visual displays of the Sp, Lx, Gx waveforms and the corresponding spectrograms were generated. Displaying the speech data in this way enabled the identification of the relevant portions of the signals for analysis.

4.5.1 Method of description for the waveforms.

Hard copies of the various waveforms were generated for all the data and visually inspected to determine the characteristics of the Gx, Lx, Sp and spectrographic signals. The characteristics for the signals were described for all the tokens in the following order: Lx, and Sp waveforms, the spectrogram and the Gx signal, and in the following way.

For the Lx signal, the mode and intensity of vocal fold vibration before, during and after the stop occlusion were noted, and the build up of Lx traces at vowel onset was also observed. The shapes of the traces during the articulation of the stop and at vowel onset were also observed since this would give valuable information about the character of vocal fold vibration. The description of the Sp signal was based on the prominence of the burst, which at this point was subjectively determined visually from the appearance on the waveform and not in an objective quantitative manner. Acoustic events following the burst, e.g. aspiration and the characteristics of vowel onset were also studied. From the spectrogram, events during the stop closure, at the release burst and after the burst, and the presence of transitions after the burst were observed.

The direction of movement for the Gx signal was observed prior to, at the onset of and during the stop occlusion and at the moment of the release burst. In this study, the behaviour of Gx was studied separately, mainly as a function of airstream mechanism. The influence of other factors like the identity of the following vowel and the tone type on the vowel were also noted where possible, but keeping in mind that interaction between all these elements would also make the interpretation of Gx complicated (see § 4.2). It was also decided to study the behaviour of Gx separately because of the need to observe its behaviour in one group of stops in relation to its behaviour in another different stop type group. In this way the situation with the voiceless unaspirated stops would become clearer: whether the behaviour of Gx is sufficiently distinct in the voiceless unaspirated
stops from that seen in the other stop types to support their traditional classification as 'ejectives'. And it was hoped that the behaviour of Gx in the Xhosa stops might provide valuable information in this respect.

4.5.2 Criteria for classification of waveforms and determination of categories.

As mentioned in the introduction above, the classification of the waveforms was determined from consistent similarities and differences observable from the visual inspection of the signals. Tokens with common signal characteristics were classified as belonging to one category e.g. type 1, type 2 and type 3. For example, tokens sharing the following signal attributes: the presence of Lx activity and the accompanying Sp response during the stop occlusion, the presence of the voice bar on the spectrogram, were grouped together under the same category (say, Type 1 category).

Categories were to be distinguished from each other on the basis of their consistent significant distinctions between the signals with which they are associated. For instance, during the stop closure, Type 1 category displays activity in the Lx waveform and the consequent Sp response and the presence of the voice bar in the spectrogram, signal properties which are typically absent from the Type 2 and 3 categories. Type 3 category exhibits an intense turbulence in the Sp signal, which can also be observed as random noise on the spectrogram after the burst. This turbulent noise is lacking in the Sp waveforms and spectrograms of Type 1 and 2 categories. Type 2 category varies from Type 1 category in failing to register voicing during the occlusion and from Type 3 category in failing to manifest intense turbulence after the release burst. This is summarised in table 4.1 (a) ~ (c).
Waveform | Signal characteristics | Category
--- | --- | ---
(a) | | 
Stop closure | (1) Activity on the Lx waveform and Sp resonances indicating phonation during the stop occlusion. | 1
Voice bar | (2) Presence of voice bar on the spectrogram on the space corresponding to the stop closure. |
Stop closure | (3) Regular vibration of the vocal folds (cf. Lx traces) at the commencement of the vowel. |
Regular vocal fold vibration at vowel onset | |

Table 4.1 (a): The classification of the waveform of type 1 category. Compare and contrast waveform shape and signal characteristics of this category with those of the categories shown on Table 4.1 (b) and (c).
Waveform Signal characteristics Category

(b) Stop closure

(1) No activity on the Lx waveform and Sp resonances indicating lack of phonation during the stop occlusion.

(2) Absence of voice bar on the spectrogram on the space corresponding to the stop closure.

(3) No turbulence after the release of the stop.

(4) Regular vibration of the vocal folds (cf. Lx traces) at the commencement of the vowel.

No voice bar on spectrogram

Stop closure

Modal phonation at vowel onset

Table 4.1 (b): The classification of the waveform of type 2 category. Compare and contrast waveform shape and signal characteristics of this category with those of the categories shown on Table 4.1 (a) and (c).
Waveform | Signal characteristics | Category
--- | --- | ---
burst | (1) The presence of turbulence on the Sp signal after the release burst indicating aspiration. This turbulence persists onto the start of the vowel. Troughs observable on Lx traces and Sp traces rather ragged. | 3
Stop closure | (2) No Lx and corresponding Sp activity during the stop closure. | 3
Aspiration on Sp after the burst | (3) Absence of voice bar on the spectrogram on the space corresponding to the stop closure. | 3
Aspiration on Sp resonances at vowel onset | 3
Troughs on Lx at vowel onset | 3
Stop closure | 3
No voice bar for stop on spectrogram | 3

Table 4.1 (c): The classification of the waveform of type 3 category. Compare and contrast waveform shape and signal characteristics of this category with those of the categories shown on Table 4.1 (a) and (b).
Chapter 4: Acoustics: Description and classification of speech and electrolaryngograph waveforms

Inter-speaker and intra-speaker variations for each category were expected and were noted. Since these categories were regarded as designating each stop type, the traditional description of each category will be given. Waveforms for Xhosa ejective (and implosive where possible) stops were generated for comparison/contrastive purposes and are discussed where relevant.

4.6 Results

Three main categories of waveforms were identified from the data. The representations for these categories are shown in figures 4.6 through 4.8, and are considered to be acceptable examples of the stop type in each case.

Type 1 category: isolation

This category of stop types was produced with voicing during the closure period and with considerable regular mode of vocal fold vibration which either continued for the whole of the stop closure or showed voicing decay towards the end of the stop occlusion. Consider figures 4.6 (a) and (b) respectively.

Figure 4.6 (a): Prevoicing in the production of Shekgalagari voiced palatal stop, [j], in the word [jáˈxɑː] 'sojourn' produced in isolation (speaker EN). Prevoicing is maintained from the start of the stop and into the vowel. The burst is unidentifiable in the Sp waveform and may be determined only by playback.
Chapter 4: Acoustics: Description and classification of speech and electrolaryngograph waveforms

In figure 4.6 (a) voicing continues throughout the stop closure into the vowel. The burst on the Sp was in most cases very small but recognisable (otherwise its presence was determined using play the back facility), and was almost always superimposed on the Sp resonances. At the start of the vowel, the amplitude of the Lx traces and Sp resonances increased and appeared to remain so for the rest of the vowel. This type of stop in this category was common with two of the four speakers in this study, EN and MK. They produced only one or two tokens with voicing decay, with speaker MK producing these with creaky voice. Voicing decay is shown on figure 4.6 (b).

Figure 4.6 (b): Showing voice decay (indicated by dotted arrows) in the production of a voiced [b] in the word [bi:na] ‘dance’ produced in isolation. In this example the delay of voicing after the burst is very brief. This particular example was taken from speaker TM who tended to produce voiced bilabial and most of the voiced velar stops with creaky voice (i.e. when they were produced in citation form). This phonation is also observable in the figure during the stop, and is shown by one long cycle at the start of the stop and by different sizes of the other cycles. The start of the vowel is produced with modal voice.
In figure 4.6 (b) we can observe that voicing during the stop occlusion is not continued into the vowel. Glottal pulsing gradually declines and eventually dies out before the burst is released. The speech signal registered acoustic activity during the stop occlusion and a low amplitude burst, which in most cases was small but recognisable. After the burst, there is a short delay of voice onset reminiscent of the voiceless unaspirated stops (to be discussed below). Regular vibration of the vocal folds is maintained in the following vowel, with big sharp glottal pulses being observable straightaway. This type of stop was common to the other two speakers (TM and CM). Occasionally, these two speakers would produce irregular phonation during the stop, with TM producing some of the bilabial and the most of the velar stops with creaky phonation (see figure 4.5 (b)), and speaker CM producing some of the bilabials with creaky voice. Speaker TM also produced a few tokens without prevoicing.

Type 1 category: within carrier sentence

In the carrier sentence, again there were notable differences between the speakers. For speakers EN and MK, the closure period was characterised by continuous modal voicing into the vowel. This is shown on 4.6 (c) for the word [biːna] ‘dance’ produced by speaker MK.
Chapter 4: Acoustics: Description and classification of speech and electrolaryngograph waveforms

In figure 4.6 (c) it can be seen that the shapes of the Lx traces during the stop are considerably regular from one period to the other and do not appear to vary from one cycle to the other. The sizes of the traces can be seen to reduce in size towards the end of the stop. The speech signal registers acoustic activity during the stop occlusion and a small but recognisable burst.

For speakers TM and CM, modal voicing was exhibited from the vowel in the carrier sentence into the stop. However, voicing was not maintained throughout the stop. The sizes of the Lx traces gradually reduced and died out eventually before the stop was released. After the burst, a short period of voicing delay was observable. Consider figure 4.6 (d) for the token [bɪːna] produced by speaker TM.

Modal voicing is resumed in the vowel, where the Lx traces and Sp resonances start with big sharp pulses.

For the stops in the type 1 category, a voice bar could be observed on the baseline of the spectrogram for the period corresponding to the closure period of the stop. This corresponds to the first formant (F₁), appearing as a single formant in low frequency regions during the period of the stop occlusion. The sounds that
Chapter 4: Acoustics: Description and classification of speech and electroglaryngograph waveforms

typically illustrate these types of Lx, Sp and spectrographic characteristics have traditionally been described as voiced stops.

Type 2 category

This type of category is represented on figure 4.7 (a), for the stop [t] in the word [tɔiqə] ‘find lost property’, produced in a carrier sentence by speaker MK.

As can be seen from the example on figure 4.7 (a), in this category the Lx signal exhibited considerable regularity in the vowel portion before the production of the stop. The Lx traces in the vowel portion gradually decreased in amplitude and finally died out at the commencement of the stop where Lx registered a line, which remained for the entire period of stop articulation and the burst. This is also reflected in the speech pressure waveform, which recorded no activity up to the point of the burst. The acoustic signal registered a burst, which in most cases was not prominent. At vowel onset the Lx began to record vocal fold activity, and the Lx traces for the vowels following these stops began with high amplitude which continued throughout the vowel. The shapes of the Lx traces show cycle-by-cycle
regularity. Lx activity at vowel onset is immediately followed by onset of maximum response from the vocal tract as shown in the Sp signal, and vibration is maintained in a regular manner right from the start of the vowel.

There is no activity on the spectrogram during the stop occlusion. The burst appears on the spectrogram as a vertical spike with a short duration. Formant transitions may be observable at the onset of the following vowel with a pattern which is in more or less steady-state condition. Often speakers would randomly produce very few tokens with one or two cycles being smaller in size than the others at vowel onset. But, in general, there was not much discernible variation between the speakers with respect to this category, except perhaps that, for some of the tokens, speaker TM showed some breathiness at vowel onset for this category of stops. Also, some of the palatal and velar stops tended to display aspiration after the burst, and therefore manifesting some slight affrication. Consider figure 4.7 (b) for the token [cʰaːca] ‘be inhospitable/ chase someone away’ produced in isolation by speaker TM. See figure 4.10 also for the token [kʰiːka] ‘mortar’, produced by speaker EN, where the velar stop [k] similarly showed aspiration after release, albeit to a lesser extent.

---

**Figure 4.7 (b): Aspiration on the Sp signal after the release burst of the word initial palatal stop [c] in the word [caːca] ‘chase someone away’ produced in isolation by speaker TM. (Notice also the first and third cycles at vowel onset are smaller in size than the other cycles).**
Chapter 4: Acoustics: Description and classification of speech and electrolaryngograph waveforms

Type 3 Category

This type of category is represented on figure 4.8 (a) and (b) for the stop [tʰ] in the word [tʰɑː.mə] "bind". Figure 4.8 (a) shows the characteristics of the signals for the stop produced by speaker MK in a carrier sentence.

Aspiration

As can be observed from figure 4.8 (a), this type of waveform exhibited considerable regularity in the vowel portion in the Lx waveform before the production of the stop. The beginning of closure is not followed by any activity in the acoustic signal until the release burst. Thus the speech display of this type of waveform at consonant articulation shows a silent period. Whilst Lx continues to register no activity, the Sp signal records a burst which in most cases was small — i.e. the burst had a very low amplitude. The Lx signal continued to register no activity after the burst, but the acoustic signal and the spectrogram recorded a period of random noise immediately following the burst which continued into the initial portions of the following vowel. Breathy vowel onset can be seen more clearly on figure 4.8 (b) for the same word produced by speaker EN.
Spectrographic characteristics show no activity during the stop occlusion, and register a spike for the burst and a period of random noise immediately following the burst which is continued into the initial portions of the following vowel. Attenuation of the first formant is also observable. The onset of periodicity for the vowel is delayed because of aspiration. No variations were observed between the speakers with respect to this stop type. This type of waveform activity characterises stops traditionally described as voiceless aspirated stops.

\textit{Gx per voice type of stops}

\textbf{Voiced stops:} The Gx movement for the voiced stops showed considerable consistency across all places of articulation and for all the four speakers in isolation and in sentence form. Figure 4.9 shows a representation of the behaviour of Gx in the voiced stop, in this case [d].
Chapter 4: Acoustics: Description and classification of speech and electrolaryngograph waveforms

Arytenoid approximation prior to voicing for the stop. Since this 'bump' appear to occur when the tokens are produced in word initial positions, it seems reasonable to conclude that it may be a general prevoicing adjustment may by the larynx.

The Gx signal descends for the voicing of the stop and rises again for the release burst and appears to descend again for the vowel.

Figure 4.9: The movement of the Gx in the production of the voiced stop [d] in the Shekgalagari word [d5:mi] 'stupid' produced in isolation (speaker EN).

From figure 4.9, we can observe that prior to the production of the stop, the Gx (and Lx) makes upward movement possibly as the arytenoid approximate for a prevoicing adjustment. The signal then descends during the production of the stop itself. This movement may possibly correlate with the fact that in the articulation of voiced stops, the larynx is lowered to increases the volume of the vocal tract, which in turn helps to sustain adequate transglottal air pressure necessary for voicing (Hombert 1978, Clark & Yallop 1990:282, see § 4.2). During voicing decay towards the release of the stop, the Gx signal moves up, possibly as a constriction is made in the larynx for the burst. It then moves downwards again as voicing for the vowel commences in the larynx. This downward movement may possibly correlate with the fact that the vowel following the stop is a non-high vowel, which we pointed out in section 5.2 may contribute in the lowering of the larynx, but possibly also with voicing in the vowel. It is not clear what the larynx is doing prior to the events just discussed. It may be presumed that the Gx at this point maps out the condition of the larynx in a rested or relaxed position. There were instances of inconsistencies, but the movements of Gx just given were
generally fairly consistent for the voiced tokens in the data and across the four speakers, in isolation and within a carrier sentence. Consider figure 4.6 (b) discussed above for similarities.

**Voiceless stops:** There was a lot of inconsistency in the movement of the Gx for the voiceless stops particularly when they were produced in isolation. This may possibly have been because the speakers could have been doing other things with their larynx during the period of silence which may not have been related to speech; like for instance, swallowing or moving the neck (see § 4.2). It is not possible to establish the cause of inconsistency reliably, but it nevertheless made the interpretation of Gx not only difficult but also rather unrevealing. For this reason Gx movement is described for the tokens produced within a carrier sentence. Figures 4.10 and 4.11 show representations of the Gx for these stop types.

**Voiceless unaspirated stops:** Figure 4.10 (a) shows a representation of the behaviour of the Gx signal in the production of the voiceless unaspirated stop. The figure shows the production of [p] in the word [piːna] ‘song’, produced within a carrier sentence by speaker EN.

![Figure 4.10 (a): The behaviour of the Gx in the production of the voiceless unaspirated stop [p] in the Shekgalagari word [piːna] ‘song’, produced in a carrier sentence (speaker EN). Contrast with figure 4.2 (a) and (b) for the Xhosa voiceless unaspirated ejective [p’] in [p’ɛnáːp’ɛːna] ‘throb with pain’.](image-url)
Chapter 4: Acoustics: Description and classification of speech and electrolaryngograph waveforms

In figure 4.10 (a), the Gx signal appears to remain in a relatively level position during the production of the voiceless unaspirated stop [p]. Two other shapes of the Gx trace were also observed, and these were common for both the voiceless unaspirated stops and the voiceless aspirated stops. One such shape resembled a line such as the one in figure 4.7 (a) for [t∫̂qa] ‘find lost property’ discussed above (speaker MK), where Gx seemed to remain level for the stop but appeared to rise after release. This is shown in figure 4.10 (b) for word initial [k] in the word [kiːka] ‘mortar’ produced by speaker EN in a frame sentence.

![Figure 4.10 (b): Gx movement for the voiceless unaspirated stop [k] in the word [kiːka] ‘mortar’, produced in a carrier sentence (speaker EN).](image)

The other shape of the Gx waveform resembled the one in figure 4.8 (a) above for the aspirated stops, which I show here again as figure 4.11. In this case the Gx seemed to remain relatively level for the stop, but, as opposed to the upward movement after the release burst observed above on figure 4.10, this time Gx appeared to move downwards after the release of the stop. Although the example given on figure 4.11 is an aspirated token, this behaviour of Gx was common with the unaspirated stops too.
It seems that the behaviour of Gx during the production of the voiceless stops was in general relatively level, and its direction of movement after the stop may possibly be a function of factors which may not be part of the stop itself, e.g. tone, vowel height and F0 (see § 4.2).

At an individual level, speaker CM tended to show similar Gx movement for both of the voiceless stops for most of the tokens. For this speaker, Gx tended to rise for the occlusion of the stop and stayed raised during the articulation of the stop itself, and then rise further up after the release of the stop and the onset of the following vowel. This seems to be in keeping with the theory that in the production of voiceless stops the larynx is raised (see § 4.2).

4.7 Discussion

The general pattern emerging from analysis of the waveforms has revealed three types of stops: voiced, voiceless unaspirated and voiceless aspirated stops, thus
conforming to their traditional description with regard to voicing. The voiced stops register modal phonation during the occlusion of the stop, which is maintained after the burst at the start of the vowel. Irregular vibration was observed for a very small number of tokens for some speakers, and may be attributed to random variability in the production of tokens, considering the fact that the majority of voiced stops were produced with modal voice.

For two of the speakers the voiced stops manifested decay of voicing towards the release burst and a brief delay of voicing after the burst for tokens produced in isolation and within a frame sentence (see figures 4.6 (b) and (d), and § 5.6 in Chapter Five). It is interesting to note that the cessation of voicing does not only remain for the rest of the stop up to the burst but continues for a brief period after the burst until the start of the subsequent vowel. Voicing does not actually start immediately after the release burst. Thus from the burst to the vowel these voiced stops appear to behave more or less like the voiceless unaspirated stops, and at least for the speakers in this study.

Decay of voicing during the occlusion for a voiced plosive was also observed in the pilot study, although in this report, which is based on a second recording, this was observed mostly for the female speakers. It has also been observed for other languages, e.g. Degema, a Niger-Congo language (Lindau 1984:148-149). For the [b] plosive in this language (at least in this study), ‘the amplitude of the vocal cords vibrations decreases gradually throughout the closure. As the supraglottal pressure increases, airflow through the glottis decreases, and eventually voicing dies out’ (Ibid. 148).

The situation with English and Xhosa is somewhat parallel to Shekgalagari and Degema, although the status of English ‘voiced’ stops as ‘voiced’ or voiceless’ in utterance initial, word initial context is not quite straightforward since they are ‘produced either with coincident or advanced phonation’ (Lieberman & Blumstein 1993:197), and in Xhosa ‘voiced’ stops are in most cases assumed by some researchers to be just that, ‘voiced’. With regard to English, what is of particular interest is when the ‘voiced’ stops are produced with advanced phonation. In this case English ‘voiced’ sounds manifest some phonation during the stop occlusion prior to the release burst. But this phonation is only partial, as it only occurs towards the end of the stop (and hence advanced), rather near the burst. Phonation may also be delayed after the release of the stop in English, as also observed for Shekgalagari voiced stops in this study. In intervocalic position, voicing may die out and start again during the occlusion for the stop. This makes the English situation rather complicated. In some languages with truly voiced stops, e.g. French and Ibibio, prevoicing may persist from the
start of the stop to the start of the vowel, although the amplitude of vocal fold vibration may decrease towards the burst during the stop occlusion.

The presence and degree of voicing in Xhosa ‘voiced’ stops appears to be controversial (cf. Jessen 1999:2 and references therein, Jessen & Roux 1999:5). It was observed earlier on that the Xhosa informant produced all of the three pulmonic voiced stops (in word initial context) in this study without prevoicing (see, for example, figure 4.4). This lack of or, if any, rather negligible amount of prevoicing in the language has been observed in other studies also, with perhaps the more interesting situation being lack of or near lack of prevoicing even in intervocalic contexts where most languages, including English, show the presence of voicing of some form and to some extent during the stop closure for the voiced stops. The particularly interesting thing with Xhosa ‘voiced’ stops in this respect is their relationship with the voiceless unaspirated stops in the language when or if they are produced without ejection. This, however, falls outside this study, and will not be addressed here.

The presence and cessation of phonation during the production of voiced stops observed in, for instance, Shekgalagari in this study, and in Degema, may be explained in terms of aerodynamic conditions. Voicing is produced by interaction between the airflow and the elasticity of the vocal folds. The airstream from the respiratory system builds up and gradually becomes pressurised in the subglottal region, behind the adducted vocal folds. When this pressure exceeds the pressure above the folds, it ultimately overcomes the muscular forces holding the vocal folds adducted, forcing the vocal folds apart slightly, starting with the bottom edges and moving upwards. The compressed air in the subglottal system now jets through the narrow slit created in the glottis into the pharynx.

Stops are produced with a complete closure at a site in the supraglottal cavities. In voiced stops, it is therefore necessary to maintain at least the minimum rate of airflow required to produce voicing whilst at the same time holding the closure in the vocal tract. This situation consequently means that for voicing to be sustained, the vocal tract has to adopt a compliance mechanism to absorb the air from the lungs, and this could be done in three different ways. There may be a passive expansion of the walls of the cavities above the larynx caused by accumulating supraglottal pressure; the supraglottal cavities may also be actively enlarged by the muscles of the larynx, or there may be an escape of airflow through an incompletely closed velopharyngeal port.

---

This applies for the non pre-nasalised stops in the language. The pre-nasalised ones show voicing during the stop closure, at least for a considerable portion of the occlusion, both when they occur word initially and in post vocalic context.
Chapter 4: Acoustics: Description and classification of speech and electrolaryngograph waveforms

As rise in air pressure above the larynx may cause the walls of the supraglottal cavities to expand involuntarily when they are in a relaxed position. In the case of bilabial and alveolar stops, the cheeks and lips most obviously comply by bulging outwards. This passive extension of the walls of the supraglottal cavities may of course be offset by a number of things, e.g. if for any reason these walls are already tensed. This may happen when the closure for the stop is made in the vocal tract when the vocal folds are still open and if, at the same time, the pressure in the subglottal system is high. This causes the air pressure in the subglottal and supraglottal systems to be balanced quickly and the expansion of the walls to happen before vocal fold adjustments for the voiced stop are made. In this case additional passive expansion of the walls to sustain voicing for the stop may not be achieved since the walls would be tensed already. Another contributing factor to this may be the fact that articulatory closure in itself reduces the size of the supraglottal cavities. Another contributing point could be high amplitude of vocal fold vibration (strong phonation), and/or long articulatory closure for the stop. The consequence of failure of the supraglottal walls to expand is that even the minimum rate of airflow required to sustain phonation is not met, and voicing during the occlusion for the stop dies out eventually.

Supraglottal cavities may also be actively enlarged, and this may be done purposely or non-purposely. Purposeful enlargement occurs when the movements of the articulators involved in the production of the stop are not used in the production of the other features of the stop or other surrounding sounds. Non-purposeful enlargement occur when the movements of the articulators involved in the production of the stop are also used in the production of the other features of the stop, e.g. the pre-release movement of the stop, or the following sound. Distinguishing between these types of enlargement may not be easy.

Enlargement may be archived through contracting the muscles of the pharynx the tongue and the neck (Rothenberg 1968:95).

The pharyngeal cavity constitutes the posterior portion of the vocal tract, and is around 12 cm in length (Clark & Yallop 1990:42). It contributes to speech production mainly by forming part of the length of the supraglottal cavities, but its geometry can also be altered since it can be narrowed or widened by contracting or relaxing the constrictor muscles. These are the inferior constrictor muscles, the

---

5 The other cavities; the laryngeal cavity, the nasal cavity and the oral cavity all open into the pharyngeal cavity. The pharyngeal cavity is therefore normally divided into three parts: namely, the laryngo-pharynx, the naso-pharynx and the oro-pharynx (Borden, Harris & Raphael 1994-92, Clark & Yallop 1990).
middle constrictor muscles and the superior constrictor muscles. Relaxing these muscles may increase the volume of the pharynx a little and help to sustain the transglottal air flow necessary for voicing during the closure for the stop, and contracting them achieves the opposite, and may contribute to voicing decay during the stop occlusion.

Another point which contributes to the volume of the pharynx relates to the movement made by the tongue in the oral cavity. Whilst, on the one hand, fronting and elevating the tongue increases the volume of the pharyngeal cavity and may therefore contribute to sustained voicing, lowering and backing the tongue, on the other, reduces the area of the pharyngeal cavity, which is not supportive of sustained phonation.

Larynx lowering and raising also affects the volume of the pharynx. The volume of the pharynx (as well as of the vocal tract as a whole) is increased when the larynx descends, and this helps to maintain voicing. When the larynx rises the opposite is achieved, and voicing may die out. It was pointed out in section 3.2 in Chapter Three that for some researchers the voiced-voiceless dichotomy between stops was described in relation to the vertical height of the larynx. Ewan and Krones 1974 found that the larynx was positioned low for voiced stops (as opposed to raised for voiceless stops). This seems to support the hypothesis that lowering the larynx increases the volume of the pharynx and contributes to sustaining voicing.

As well as affecting the volume of the pharyngeal cavity the movement of the tongue also affects the size (and shape) of the oral cavity. Lowering and backing the tongue increases the volume of the oral cavity, and this may contribute to sustained voicing during the stop closure. But fronting and elevating the tongue reduces its volume and this may contribute to a drop in transglottal air pressure required to produce phonation, a situation which may lead to voicing dying out during the stop occlusion.

An incomplete closure of the velopharyngeal opening passage also causes air to escape through the nose. There has been exploratory work which has shown that partial opening of the velopharyngeal port appears to help absorb

---

6 The inferior constrictor muscles are situated at the larynx, the middle constrictor muscles are positioned at the top-most back part of the neck and run down to reach the hyoid bone, and the superior constrictor muscles which comprise the back part of the pharynx and run from the area level with the palate to the lower jaw (Borden, Harris and Raphael 1994:92).

7 Some quantitative studies have however showed that vertical displacement of the larynx was less than one centimetre (1 cm), in fact it was even less than 0.2 cm (Rothenberg 1968:99).
pressure in the production of voiced non-nasalised stops, and least in some languages and/or dialects of some languages (Rothenberg 1968:99-106 and references therein). This, however, needs to be studied more.

The contribution of the lips to the volume of the supraglottal cavities is made mainly by their protrusion. The vertical movement of the lower jaw seems to be the only dimension which is significant in speech. This movement changes the size and volume of the oral cavity.

Decay of voicing during the occlusion for the stop as observed in this study does happen in other languages plausibly as an aerodynamic factor, e.g. English (especially in intervocalic position) although it is often not reported. We may still regard these types of stops in Shekgalagari as truly voiced since active voicing occurs for most of the stop production, and it may therefore be considered to be a sufficient distinguishing cue — though this remains to be investigated further through perceptual experiments. The long duration of (active) prevoicing in these stops (see § 4.3) conforms to the universal property of truly voiced stops of large negative prevoicing.

In terms of cross-linguistic analysis, as noted before, (initial) voicing for ‘voiced’ stops is realised differently in the languages discussed above. It seems that in some languages, like Shekgalagari, the physical and aerodynamic conditions for prevoicing are met early on, at the start of the stop occlusion, but are not maintained for the rest of the stop and a short time after the burst. This results in different voicing status: decay during the stop closure and delay after the burst. In English, adjustments for voiced stops require voicing to happen rather near the burst, and there may be voicing delay after the release burst or voicing may be maintained for the rest of the syllable. In French and Ibibio voicing adjustments for voiced stops may be maintained from the start of the stop and throughout the syllable. Cross-language investigation has shown that sounds that are alike in some sense, e.g. ‘voiced’, may often be produced differently in different languages and may even be produced differently by different speakers in a given language (Shimizu, 1990). The voiced stops with decay and delayed onset of voicing in Shekgalagari may be considered to be an example of this.

The voiceless unaspirated stops are the category of stops in Shekgalagari that have traditionally described as 'ejectives', which suggests some glottalization in their production (Clark & Yallop 1990:57). Glottalization is related to both area and duration of vocal fold contact; the greater the degree of glottalization, the greater the area and duration of contact and, consequently, the greater the closed quotient (CQ), which is the percentage of the amount of time the glottis is closed in the period of a single vibratory cycle. This information can be derived from the
Lx signal since, as we noted earlier on, the shapes of the Lx traces provide valuable information about the nature of vocal fold vibration (§ 4.2).

Laryngographic investigation of ejective sounds includes that of Lindsey, Hayward and Haruna (1992) (see also Haruna 1990) on Hausa, which contrasts ejectives with plain consonants. For the intervocalic ejective affricate [ts’] in the word [tsaat’saa] ‘rust’, Lindsey, Hayward and Haruna observed, at consonant onset, the Lx waveform ‘which we associate with glottal closure, indicating that the glottal stricture lasts the full duration of the consonant...’ (Lindsey, Hayward & Haruna 1992:519). This is shown in figure 4.12 (adapted from Lindsey, Hayward & Haruna 1992:519).

At vowel onset, the ejective affricate [ts’] also showed ‘markedly greater low frequency perturbation (slow up and down deflection) in the Lx waveform...’ (Ibid. 1992:520). By contrast, for the plain intervocalic fricative [s] in the word [taasáa] ‘metal bowl’ the open quotient at the start of the consonant increases, ‘anticipating abduction rather than adduction of the vocal folds, and showed faster up and down deflections on the Lx traces’ (Ibid. 519 - 520). Ejectives thus generally show decreased open quotient than plain consonants. Other characteristics of ejectives at vowel onset observed in Hausa included irregular mode of vocal fold vibration with period-by-period fluctuation in amplitude, and
this was observed for the ejective stop [k'] in the word [sháak'aa] ‘smell, sniff’ (Haruna 1990:505).

Abberton (1972) studied the Korean so-called ‘tense’ stops. For the male speaker of the study, these so-called ‘tense’ stops showed gradual build-up of intensity on the Lx traces at vowel onset, typical of creaky voice, and the Gx waveform appeared to suggest some indication of whole larynx movement (Abberton 1972:73, Fourcin & Abberton 1977:313). Consider figure 4.13 (adapted from Abberton 1972:74).

Different languages show different signal characteristics, and sometimes speakers also show considerable variation in their realisation of ejective stops. Considerable variation may also exist when the same ejective token is produced several times, with a token being sometimes produced in a plain manner and sometimes as a glottalized stop. But the general trend seems to suggest decreased open quotient and/or perturbation of some form (e.g. irregular vibration, one or two long cycles) prior to the target stop and/or at vowel onset (for the voiceless stops). Another observation which has been made with regard to glottalized consonants is that the rate of vocal fold vibration ‘is typically, though not necessarily slower’ (Ladefoged, 1973, et. al., 1988, cf. figure 4.12 and the subsequent discussion). We have already observed similar Lx signal characteristics for the Xhosa ejectives and implosives, where vowel onset showed
irregular vocal fold vibration and/or decreased open phase (see figures 4.2 (c) and 4.3(c)) both prior to the target stop or at vowel onset.\(^8\)

Other studies on Xhosa ejectives, have, however, not made the same or similar observations. A recent acoustic based study by Jessen (1999:13), for instance found out that

Creaky voice during the early part of the vowel (as indicated by low F0, irregularity of period durations, and alternations between high- and low- amplitude periods) was very uncommon in the entire corpus. When it did occur it was only vaguely related to ejectives and implosives and also occurred in other cases. The occurrence of creaky voice seemed to be more subject to inter-speaker and pure token-to-token variability than to be associated particularly with glottalic consonants. (Jessen (1999:13).

This appears to support the observations made by other researchers that irregular phonation at vowel onset may not be a reliable cue for ejectives in the languages of the world (Henton et. al 1992). For example, Hausa ejectives were observed to exhibit regular voicing at vowel onset whereas Navaho ejectives showed creaky voice after the release of the stop which continued into at least the first portion of the vowel (Lindau (1984). For the speakers of these languages, the long glottal closure shown by Navaho ejectives was the main distinctive cue. This shows the fact that often what would be a significant distinctive cue for ejectives in one language may not necessarily be a cue for the same types on stops in another language.

With respect to Shekgalagari voiceless unaspirated stops, although a few tokens were produced with irregularity of vibration in the initial portion of the vowel for one or two speakers, modal phonation on the Lx signal prior to the stop and at vowel onset for a large proportion of the data for these stops in this study does not seem to suggest any glottalization in these stops (see figure 4.7 (a)). For most of the tokens, the Lx traces did not show cycle-by-cycle fluctuation in period and intensity, and no gradual build-up of intensity was observed. Also, a small, rather insignificant burst was a feature in most of the waveforms for these stops, with the burst not even observable in other waveforms. This appears to be in contrast to how ejectives are thought to be produced: with compressed air behind

---

\(^8\) Jessen (1999) conducted an acoustic study of Xhosa ejective stops [p,t,k] and found that 'ejective productions could be identified auditorily by their distinct bursts and ... were manifested acoustically with relatively long VOT's and high burst amplitudes (Ibid. 1999:28). This study did not involve the Gx laryngograph.
Chapter 4: Acoustics: Description and classification of speech and electrolaryngograph waveforms

the articulatory closure doubling that of the normal pulmonic pressure, which, upon release, gives greater amplitude to the burst (Ladefoged & Maddieson 1996:78). Note that in the study conducted by Jessen (1999) for Xhosa, where ejectives were produced, they were identified by high amplitude of the burst. In Shekgalagari, however, this remains to be investigated through acoustic experiments. But all the observations made for Shekgalagari voiceless unaspirated stops seem to suggest that this may be considered to be representative of the way they are produced in the language, as simple plain voiceless unaspirated stops, and not ejectives.

Nevertheless, more experimental work needs to be done on these stops, and more cross-linguistic analysis made, in order to determine whether these stops may possibly be ejectives of the weaker type. Investigations that may be made on these stops include, for example, quantitative analysis of the burst amplitude, measurement of closure duration, of intra-oral air pressure during the hold phase amongst other things. Compared with other stops, ejectives have often manifested long VOT values. For some North American languages and some Caucasian languages, VOT values of around 80+ ms in duration are not uncommon. Relatively long VOT was also found to be a feature of the Xhosa ejectives, albeit a less prominent characteristic, and VOT values were far less than those observed for the above mentioned languages (Jessen 1999:28). With respect to glottal closure, ejectives in some languages are identified by long closure duration. It was just noted above that the relatively long closure duration in Navaho ejectives is a significant factor in distinguishing them from Hausa ejectives. Ejectives have also been found to exhibit high burst amplitudes in some languages (Jessen 1999:28 (Xhosa), Fre-Wuldo 1985 (Tigrinya)). Ejectives have also exhibited higher oral air pressure than other voiceless stops in some languages (Fre Wuldo 1985:126 ff).

In addition to significant variation in the phonetic realisation of ejectives across languages, there could also be significant speaker variation and/or token-to-token variation as observed earlier on with Xhosa. This may in turn contribute to complications and associated inconclusiveness relating to whether 'ejection' is a phonological feature of a language (Jessen 1999:29, Jessen & Roux 1999:4-5 (on Xhosa). Quantitative analysis of Shekgalagari (voiceless unaspirated) stops, (VOT), is the subject of the next chapter.

As regards the voiceless aspirated stops, the results of the investigation in this study accord with what is reported in the literature for voiceless aspirated sounds in other languages, where vowel onset displays breathiness (Abberton 1972, on Korean aspirated stops). This seems to conform to, as well as confirm, their traditional description as voiceless aspirated stops.
The Gx signal appears to show some different movement between the stops according to their voicing structure, in that distinctive downward movement is made for the voiced stops in comparison to the movement it makes for the voiceless stops. In the present data, the shapes of the Gx trace observed for the voiceless unaspirated stops was also common for the voiceless aspirated stop (see figures 4.10 (b) and 4.11) — most tokens showed a relatively level Gx waveform for the voiceless stops (see figure 4.10 (a) ~ (b) and 4.11). Gx behaviour after the release of the burst for the aspirated and unaspirated voiceless stops appears to be affected by properties of the subsequent segments e.g. tone type, vowel height and F0, and possibly by a combination of all the factors. Nevertheless, it is interesting to note that these shapes were observed for the voiceless stops only and not the voiced stops. This seems to suggest that Gx movement may be more in keeping with the voicing status of the stops than in the displacement of the larynx due to the glottalic airstream mechanism (as seemed to be the case with the behaviour of Gx for the Xhosa ejective in [p'énap'énå] ‘throb with pain’ (see figures 4.3 (a) and (b)). It has already been noted in section 4.2 that the larynx tends to be displaced downwards for the voiced stops as the vocal tract complies to sustain transglottal air pressure for voicing. For the voiceless stops, however, the larynx remains raised, and this appears to be in association with the devoicing adjustment as opposed to the voiced one. The fairly consistent pattern of Gx movement for speaker CM seems also to support this. For this speaker, the Gx waveform tended to rise for the occlusion of the stop and stayed raised during the articulation of the stop itself, and then rise further up after the release of the stop and the following vowel for both types of voiceless stops. This seems to be in keeping with the theory that in the production of voiceless stops the larynx is raised (see § 4.2).

In fact, the behaviour of Gx in the production of voiced vs. voiceless stops in this study appears to support Hombert et al. (1979) and Ewan and Krones (1974)’s larynx height hypothesis. According to them, the voiced-voiceless dichotomy of stops is distinguished by the vertical height of the larynx. They, however, continue to say that ‘additional evidence in favour 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). This is particularly relevant for speaker CM for whom Gx rose for the occlusion of the stop and stayed raised during the rest of the pronunciation of the stop, and then rose further up after the release of the stop and during the following vowel (thus appearing to show correlation between F0 and larynx elevation) for both types of voiceless stops (cf. § 3.2.2). But for the rest of the speakers the shape of the Gx trace after
Chapter 4: Acoustics: Description and classification of speech and electrolaryngograph waveforms

the release of the voiceless stops was different as observed in figures 4.10 ~ 4.11. It was mentioned above that for the stops investigated in this study, there may be other factors influencing the behaviour of the Gx trace at vowel onset which may not necessarily be part of the properties of the stops, e.g. tone.

It was observed in figures 4.10 and 4.11 for the word initial velar and aspirated alveolar stops in [kʰiːka] 'mortar' and [tʰaːma] 'bind' respectively, that Gx appeared to rise at vowel onset. But it was not always clear what to attribute this to. In figure 4.10 for [kʰiːka] 'mortar', for instance, the rise of the Gx trace at vowel onset may be thought to be caused by or associated with F0 perturbation (F0 is thought to rise after the production of voiceless stops, in this case [k], see § 3.2.2), or by the high vowel [i] (see § 4.2), or by the high tone in the (first) syllable or possibly by a combination of all these three factors. In figure 4.11 for [tʰaːma] 'bind', the rise of the Gx trace at vowel onset may still be plausibly attributed to F0 perturbation and/or the high tone in the (first) syllable, and this inspite of the low vowel [a]. It is not possible (at least in this study) to determine which of the three factors is the most influential, or indeed the order of influence of these factors on the Gx, especially when they all occur on the sound or syllable under investigation. In addition, the influence of such factors on the Gx is an aspect less well studied in linguistics. All this makes the interpretation of the Gx less easy.

Observations made for the production of ejectives in Xhosa may also shed some light on the nature of production for Shekgalagari voiceless unaspirated stops, particularly with regard to the airstream mechanism employed. In figures 4.3 (a) and (b), it was observed, for this particular Xhosa informant, upward deflection of Gx in the production of the Xhosa ejective [pʰ] in the word [pʰɛnáːpʰɛnáː] 'throb with pain' produced in isolation and within a carrier sentence. It was also pointed out that this could plausibly be attributed to the abrupt displacement of the larynx as it initiates the airstream mechanism required for the production of these stops in Xhosa, at least for our speaker.

This distinctive, sharp upward deflection of Gx was however not observed for the voiceless unaspirated stops in Shekgalagari (contrast figure 4.2 (a) and (b)).

Only one person served as an informant for the investigation of Xhosa stops. For this reason, Gx observations made here with regard to the glottalic stops in this language may not be conclusive. Other (acoustic based) studies (e.g. Jessen 1999, Jessen & Roux 1999) have since shown that ejection in Xhosa seems to be speaker dependent, and that there was also variation on this aspect from token-to-token. Other features which have been associated with ejectives, e.g. creaky voice at vowel onset were uncommon for most speakers in these studies, and occurred in other stop types which were not necessarily glottalic. Jessen (1999:29) concludes that due to this instability, it is difficult to consider ejection as a phonological aspect of Xhosa, although it does seem to be a characteristic of Xhosa stop system which has to be learned.
Chapter 4: Acoustics: Description and classification of speech and electrolaryngograph waveforms

with 4.10 (a) and (b) and 4.11) for all the four speakers of this language who served as informants. This may possibly be suggestive of the probability that Shekgalagari voiceless unaspirated stops may not be produced on the glottalic airstream mechanism, and may therefore not be ejectives as they have traditionally been thought to be. But, as mentioned earlier, further systematic experiments are nevertheless necessary to try to ascertain whether Shekgalagari voiceless unaspirated stops may possibly be ejectives of the weaker type. And certainly more informants have to be recorded to help reach a more reliable conclusion.

4.8 Summary

The aim of this chapter was to provide a qualitative description and classification of the waveforms obtained for Shekgalagari stops. The qualitative description of the waveforms has been discussed based on visual inspection of the properties of Gx, Lx, Sp signals and the spectrogram for each voice type, and the classification on the shared characteristics and on general distinctions rising from the examination of the signals. The main objectives of the investigation were as follows. To find out: on the basis of the waveform types, the category of stop voicing contrast that Shekgalagari belongs to; whether Shekgalagari has 'ejectives' or plain voiceless unaspirated stops; what the language-specific phonetic characteristics of its stop system are and how they differ from, or are similar to, the phonetic realisations of stops in other languages reported in the literature — e.g. English, Xhosa, Korean Hausa and Thai. The findings of the experimental investigation have been discussed in the light of these objectives as well as within the framework of speech production.

From the findings of this study and the discussion above it seems reasonable to conclude that Shekgalagari is a three-way category language, with voiced, voiceless unaspirated and voiceless aspirated stops. We may also conclude, at least on the basis of the findings of the present chapter, that the voiceless unaspirated stops in Shekgalagari may be simple plain voiceless unaspirated stops, and not ejectives. But this remains inconclusive pending further experimental analysis and cross-linguistic examination to determine whether these stops may possibly be ejectives of the weaker type.
5 Acoustics: Voice Onset Time

5.1 Introduction

The aim of this chapter is to determine the effectiveness of voice onset time (VOT) in distinguishing a series of word initial homorganic stops in Shekgalagari. Issues pertinent to VOT will also be addressed: namely, VOT as a function of place of articulation, the effect of phonetic context on VOT — specifically, the influence of the following vowel, and to consider inter- and intra-individual variation.

The previous chapter focussed on qualitative analysis of the different voice contrasts of Shekgalagari stops, and was based on the visual inspection of the Sp, Lx and Gx waveforms and of the spectrogram. It was mentioned that more experimental work needs to be done on the voiceless unaspirated stops (and more cross-linguistic analysis made), in order to determine whether these stops may possibly be ejectives of the weaker type. Such study may include, for example, measurement of the burst amplitude, of closure duration, of intra-oral air pressure during the hold phase amongst other things. Compared with other stops, ejectives have often manifested long VOT values in some languages. For some North American languages and some Caucasian languages, VOT values of around 80+ ms in duration are not uncommon. Relatively long VOT was also found to be a feature of some speakers’ production of ejectives in Xhosa, albeit a less prominent characteristic, and the VOT values obtained here were far less than those observed for the above mentioned languages (Jessen 1999:28). With respect to glottal closure, Navaho could be an example of a language employs relatively long closure duration in ejectives in a distinctly significant manner from Hausa ejectives. Ejectives have also been found to exhibit high burst amplitudes in some languages (Jessen 1999:28 (Xhosa), Fre-Wuldo (1985) (Tigrinya)). Ejectives have also exhibited higher oral air pressure than other voiceless stops in some languages (Fre-Wuldo 1985:126 ff.,). In addition to significant variation in the phonetic realisation of ejectives across languages, we also noted, in the previous chapter, that there could also be significant speaker variation and/or token-to-token variation, and that this may in turn contribute to complications and associated inclusiveness relating to whether ‘ejection’ is a phonological feature of a language (Jessen 1999:29, Jessen & Roux 1999:4-5 (on Xhosa).

We also noted in the previous chapter that voicing decay was observed for the voiced stops in Shekgalagari, at least for some speakers, and that this was also
observed during the pilot study. It has been observed for other languages e.g. English and Xhosa, that the amount of voicing in certain contexts (word initial being the commonly reported context) was almost negligible and that the labelling of these so-called ‘voiced’ stops in these languages was not a straight-forward matter. Voiceless aspirated stops in general manifest long duration of aspiration, with duration values for these stops being rather larger in some languages than those which have been reported in the literature for other languages.

The present chapter focuses on quantitative analysis of Shekgalagari stop contrasts, and will deal specifically with the duration aspect of voice onset time (VOT). As mentioned above, some authors appear to suggest long VOT values for ejectives in some languages (almost as long as the voiceless aspirated stops), and longer VOT values for the voiceless aspirated stops than those which are reported for other languages. The extent of voicing during closure for the voiced stops, as well as that of voicing decay where applicable, are measured. The findings in this study will be compared with similar studies in the literature, and it would be interesting to note what the situation with the voiceless unaspirated stops in Shekgalagari is compared to similarly labelled stops in other languages.

The rest of this chapter proceeds as follows. Section 5.2 discusses the theoretical background in relation to VOT. Section 5.3 discusses the two measurement criteria used for VOT. In section 5.4 we discuss the conditions of recording, informants and data. This is followed by the discussion of how VOT and prevoicing were measured in section 5.5. Section 5.6 presents the results of the measurements. The discussion of the results is in section 5.7, followed by a summary of the chapter in section 5.8.

5.2 VOT: Theoretical background

VOT relates to the co-ordination of vocal fold vibration and articulatory events in the vocal tract and is exploited differently in different languages for contrastive purposes. It may be affected by a number of things: for instance; the quality of the following vowel and the place of articulation for the stop.

Voice onset time is defined as the time interval between the release of a stop and onset of vocal fold vibration in the following vowel, and may be specified as a single value in milliseconds. The beginning of phonation may happen during the stop closure and prior to the release of the burst, as in certain voiced stops, or it may occur after the burst as in voiceless stops. Where phonation follows the burst, it may coincide with the release (coincident phonation), or it may be considerably delayed after the release of the burst
(delayed phonation). These three divisions of voice onset time effectively divide stops into three main domains in term of voicing structure: voice lead, short lag and long lag respectively; which in turn correspond to the phonological categories of voiced, voiceless unaspirated and voiceless aspirated. Sounds with voicing lead, such as those in French, Italian, Dutch and Ilwana, where energetic vibration of the vocal fold is maintained throughout the articulatory interval corresponding to the stop occlusion are assigned negative VOT values. Voiceless stops, where onset of glottal pulsing follows the release burst for the stop, fall on the positive half of the VOT continuum; with variation in duration ranging from 0 ms to 20 ms for those with coincident phonation, and from above 20 ms upwards for those with delayed phonation (Lieberman & Blumstein 1988: 197, Kent & Read 1992:108).

Languages exploit the voicing manoeuvres of the larynx in different ways to distinguish between the phonemic categories of their stops. Some languages have a two-way variation, others a three-way, and yet others manifest a four way-contrast. Two category languages include English and Spanish. In English, the relevant phonemic distinctions are manifested by delayed voicing (for example the sound \( [p^\text{h}] \), as in \( [p^\text{h}m] \), and coincident phonation (for example the sound \( [\text{b}] \), as in \( [\text{b}m] \)).\(^1\) Thai is an example of a three-way contrasting language: namely, voicing lead, coincident phonation and delayed phonation. Hindi and Gujarati are four-way category languages with voicing lead, coincident phonation, delayed phonation and a fourth category exhibiting aspiration and voicing concurrently.

Korean belongs to the category of three-way contrasting languages, with all of its stops falling on the positive half of VOT. But, VOT values for the unaspirated and the so-called ‘tense’ stops in this language shows overlapping values and therefore is insufficient to distinguish the contrast (Han & Weitzman 1970:114, Hardcastle 1973:266, Shimizu 1990:78). Similarly, four-way contrasting systems have ‘murmur’, where phonation and turbulence occur simultaneously (Hirose et. al., 1974, Shimizu 1990:149). These languages have voiced aspiration contrasting with voiceless aspiration, and the timing dimension on its own falls short in distinguishing these particular stops from the voiced stops and the aspirated stops effectively. Marathi, Hindi and Gujarati are some of the languages which have ‘murmured’ sounds.

\(^1\) Actually, English [b] is produced with either coincident or advanced phonation (Lieberman & Blumstein 1988:197). Unlike French [b], this sound is not produced with energetic vibration of the vocal cords throughout its articulatory closure. Rather, there may be no voicing at all (hence coincident phonation), and if any, it is only produced for part of the closure (hence advanced phonation).
With regard to effects on VOT, it has been found that, for voiceless stops, VOT systematically varies as a function of the quality of the following vowel and of place of articulation for the stop. In a wide variety of languages, stops that are articulated in the context of high vowels have longer VOT values than stops which are followed by non-high vowels (Klatt 1975:694; 697, Port 1979). The hypothesis for this has been that the raising of the tongue for the high vowel narrows the oral cavity and this creates some resistance to the flow of air from the lungs. This in turn delays the development to transglottal air pressure necessary for vocal fold vibration. With respect to the relationship between VOT and place of articulation, it has also been observed that stops which are produced at the posterior most part of the vocal tract (the velar stop /k/ being the most commonly reported) have longer VOT values than stops produced at anterior parts of the tract. Two hypotheses have been proposed to account for this.

The first hypothesis relates to the aerodynamic factors involved due to the volume of the cavities above the larynx, and the second is related to an articulatory factor based on the size of the tongue body involved and its relative speed in the production of the stop. According to the first hypothesis, as the tongue moves further back in the mouth, it gradually reduces the size of the supraglottal cavity behind itself. As a result of this, there is a high level of air pressure behind the obstruction for articulations that are made further back in the mouth. When the obstruction is released, there is some delay in balancing air pressure below and above the glottis, causing delay in the commencement of voicing for the subsequent vocalic segment. The further back in the mouth the obstruction is, the longer it will take for the development of air pressure required for voicing (Klatt 1975:701-702, Hardcastle 1973:266). In most of the languages reported in the literature, there has been a tendency for velar stops to have longer VOT values, followed by alveolar stops and lastly the bilabial stops. Lisker and Abramson (1964) made an earlier observation of this. Many subsequent studies have since made similar observations, e.g. Shimizu (1990).

According to the second hypothesis, VOT is affected by the speed of the active articulator involved. It has long been observed that larger parts of the tongue body, e.g. the dorsum, move relatively more slowly than the smaller parts, e.g. tip and blade (Klatt 1975:695, Hardcastle 1973:266). This means that back articulations are released relatively more slowly than front articulations. And, according to Klatt (1975:695), an additional point for slower release of posterior constrictions relative to anterior ones is ‘the release vector of the tongue motion is usually not perpendicular, except perhaps before /r/’. A consequence of this in
terms of timing is that it takes longer for the aerodynamic conditions necessary for voicing to be met in back articulations than in front articulations.

These are some of the factors which may also be considered in relation to VOT, along with the effects of rate of speaking, suprasegmental features like tone, stress and lengthening in anticipation of a pause or as a language feature, such as penultimate lengthening in Bantu languages (see § 2.4.3).

5.3 Measuring VOT

There has been a variety of opinion with respect to measuring VOT, particularly for voiceless aspirated stops. The main problem hinges on defining the end of aspiration in the stop and the beginning of periodicity in the vowel. The first measurement criterion, used by some researchers, e.g. Kunzel (1977, mentioned in Jessen (1997)), Fisher-Jørgensen and Hutters (1981), Braun (1988, also mentioned in Jessen (1997)) and Jessen (1997), uses spectrograms and/or speech signals to measure the duration of aspiration from the burst to the onset of $F_2$. This thus includes breathy phonation during which turbulence occurs as well as vibration of the vocal folds for the vowel. It is argued that $F_2$-onset is more reliably associated with the end of the aspiration associated with the stop than positive VOT, which only marks the beginning of vocal fold vibration for the following vowel (Jessen 1997:63). By using $F_2$-onset as the terminal point for aspiration we could therefore make more reliable measurements for aspiration and voicing in languages like Hindi, Marathi and Gujarati which make contrasts in terms of both of these factors — a situation which therefore requires that voicing and aspiration be measured separately (Jessen 1997:63).

The second criterion simply uses positive VOT (Lisker & Abramson 1964), where the beginning of vocal fold vibration in the vowel is the point at which measurement of aspiration ends.

Whilst I regard the $F_2$-onset view as valid, at least for languages with 'murmured' stops, in this study, I follow Lisker and Abramson (1964) in measuring VOT from the burst release of the stop to the start of vocal fold vibration (determined from the Lx signal). The beginning of vocal fold vibration in the vowel is considered to signal the end of the stop, and the aspiration superimposed on the vowel is an inevitable consequence of vocal fold configuration at the start of voicing. Figures 5.1 through 5.3 show the measurement criteria for the different types of stops followed in this study.
5.4 Procedure

5.4.1 Recording conditions, subjects and data

The recording conditions for the data to be analysed in this chapter, the information regarding the informants involved in the experiment as well as the data are exactly the same as in the previous chapter, i.e. it is the same recording.

5.5 Processing

The SFS program was used for analysing the data time aligned the Lx and Sp waveforms so that there was no need for manual manipulation. After displaying the waveforms on the monitor, voice onset time was measured by placing one cursor (cursor 1) at the onset of the burst (which was marked by abrupt, aperiodic energy after the stop occlusion) on the acoustic waveform and the other cursor (cursor 2) at the commencement of the steep rise of the first recognisable Lx trace on the laryngograph waveform (which marks the beginning of voicing for the following vowel) for the voiceless stops. For the voiced stops, prevoicing was measured by placing one cursor at the beginning of voicing for the stop occlusion on the Lx signal and the other cursor at the start of the burst on the speech signal. This included the period of voicing decay where there was any, since, also, this is a common practise, although voicing decay is not often reported in other languages when and if it occurs. SFS programs then automatically computed the measurement between the cursors by subtracting the time at one location (e.g. the burst) from the time at the other location (the beginning of voicing in the case of voiceless stops. Where necessary, voicing decay was also measure for the voiced stops. Measurement criteria for the stops is illustrated in figures 5.1 through 5.3 for the three types of stops.
Figure 5.1: The determination of voice onset time for a voiceless unaspirated stop

Figure 5.2: The determination of voice onset time for a voiceless unaspirated stop
5.6 Results

The tables 5E(1) to 5E(4) in the appendix present detailed documentation of prevoicing and VOT values for the stops for all of the four informants individually. In table 5.2 the following are presented: the ranges (R), average values for prevoicing and VOT, and the standard deviations (SD) for the different stop types for each of the four speakers any, observable variation between the informants is discussed. First the results of the experiment for each voice type pooled across all the four informants are shown in table 5.1 (a).

Table 5.1 (a) shows average prevoicing and VOT values for the vowels per stop, the means for the stops per place of articulation, the means for the vowels across the places of articulation for the stops and the standard deviation of the means for data pooled across all of the four informants. The averages were calculated for the stops per vowel context in the following way. For each consonant type by place, the averages per vowel context were calculated by adding prevoicing and VOT values for the speakers and dividing the result by the
total number of the speakers. Then the means for the stop types were calculated for each place from the values obtained for the informants, and not from the averages calculated for the vowel contexts. Mean values for the vowels were also calculated from the values obtained for the informants and not from the averages. Thus the means recorded on the tables are not the means of the averages (means of the means). This criterion was adopted because calculating the means from actual values obtained for the data is more reflective of the time values and ranges than calculating it from the averages, which tend to narrow the time and time ranges down. These calculations: the averages, means for the averages and the standard deviations were calculated by means of the SPSS program and recorded.

It can be observed from table 5.1 (a) that the voiced stops in Shekgalagari manifest a rather long voicing lead. This is generally consistent with values reported for voiced stops in other languages; e.g. Thai (Shimizu 1990). The standard deviations indicate that the ranges for the stops in the tables are quite wide.

There also seems to be some interaction between the values for prevoicing and vowel identity. There is a non-consistent tendency for stops which are followed by close or half close front vowels to manifest higher prevoicing values than those which are followed by non-close vowel(s). The effect of the vowel on prevoicing appears to be in this order: / i > e > o > o > u/, with the non-close vowel /a/ showing the least prevoicing value. Consider table 5.1 (b).
Table 5.1 (a): Averages, ranges, means and standard (SD) for prevoicing and VOT for all the four speakers. The values recorded for the vowels represents the averages for each vowel, and was calculated from the values obtained for the four speakers. The ranges represent the range of values for each vowel per stop. Mean 1 represents the means for the stops per place of articulation calculated from the whole data, and not from the averages for the vowels, as well as SD 1. Mean 2 represents the means for the vowels across the place of articulation for the stops. It was also calculated from values for the whole data, and not from the averages for the vowels, as well as SD 2.
The effect of place of articulation on prevoicing, in this study, appears to give the following rank order: /j > b > g > d/.

There were, of course, certain inconsistencies between speakers in the production of these stops. Informants EN and MK, the male speakers, produced almost all of the voiced stops without voicing decay, both in isolation and within a carrier sentence. Speaker MK produced all the velar stops without prevoicing but with delayed voice onset after the burst, i.e. as voiceless unaspirated stops. Voicing decay was, however, a feature with informants CM and TM, the female speakers, both in isolation form and within the frame sentence, except for the bilabials in the case of CM (see table 5E(1) to 5E(4) in Appendix 5E). They also showed delayed phonation after the burst.

A number of things are observable for the voiceless unaspirated stops. These include the time values obtained for these stops in relation to those obtained for the voiced stops; the effect of the following vocalic segment on the VOT values, as well as the effect of place of articulation. It can be observed that timing values for the voiceless unaspirated stops fall on the positive half of the VOT continuum and manifest short positive voicing lag. This is distinctly different from the time values obtained for the voiced stops, which were on the negative half. There is clearly no overlap in the time values obtained for the voiced stops and the voiceless unaspirated stops. Mean time values for the two stop types were subjected to a two tailed t-test for statistical significance, and the t-value = 9.54, p < 002 (see table 5.1 (c)), indicating that the difference between the mean values is statistically significant.

2 The negative sign is not relevant in the determination of t-test. Also, although paired sample tests were performed here, it is thought that 2-sample t-tests would have been more appropriate.
Paired Samples Test

Paired Differences

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Std. Error Mean</th>
<th>95% Confidence Interval of the Difference</th>
<th>Sig. 2-tailed t-value</th>
<th>d.f</th>
<th>Sig.</th>
</tr>
</thead>
</table>

Table 5.1(e): T-test of the mean values for the voiced and the voiceless unaspirated stops. Note: the uvular place for the unaspirated stops was not included.

With respect to the effect of vowel identity, a similar observation is made for the voiceless unaspirated stops as was made for the voiced stops: that stops followed by high vowel tend to display longer VOT values than those which are followed by non-high vowels. The order of rank is as follows: /i > u > o/ o > e/e > a / (see table 5.1 (d)). As regards the effect of place of articulation, stops produced at the palatal place of articulation have lower VOT values than those of stops produced at other places of articulation. The order of the effect of place of articulation in this study is as follows: palatal > velar > uvular > bilabial > dental. We shall discuss the possible underlying reasons for the effect of vowel context and place of articulation in section 5.7. Intra- and inter-speaker variation will be discussed later on following table 5.2.

<table>
<thead>
<tr>
<th>Place of articulation</th>
<th>Vowel identity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Factor</td>
<td>VOT</td>
</tr>
<tr>
<td>[c]</td>
<td>40.4</td>
</tr>
<tr>
<td>[k]</td>
<td>31.8</td>
</tr>
<tr>
<td>[q]</td>
<td>15.7</td>
</tr>
<tr>
<td>[p]</td>
<td>14.1</td>
</tr>
<tr>
<td>[t]</td>
<td>12.6</td>
</tr>
<tr>
<td>[i]</td>
<td>30.2</td>
</tr>
<tr>
<td>[u]</td>
<td>27.1</td>
</tr>
<tr>
<td>[o/ɔ]</td>
<td>20.0</td>
</tr>
<tr>
<td>[e/e]</td>
<td>19.8</td>
</tr>
<tr>
<td>[a]</td>
<td>14.6</td>
</tr>
</tbody>
</table>

Table 5.1 (d): VOT as a function of place of articulation and vowel context for the voiceless unaspirated stops (cf. Table 5.2 (a)), showing the ranking order of factors on

The voiceless aspirated stops also produce some similar observations to those made for the voiceless unaspirated stops. These include the time values obtained for the voiceless aspirated stops in relation to the values obtained for the voiced stops and the voiceless unaspirated stops; the effect of the following vowel on
VOT as well as that of place of articulation. Like the voiceless unaspirated stops, VOT values for the voiceless aspirated stops fall on the positive half of the VOT continuum, but unlike the voiceless unaspirated stops, the aspirated ones show a long positive lag. This makes them quite distinct from the voiceless unaspirated stops. VOT mean values for the voiceless unaspirated and the voiceless aspirated stops were subjected to a two-tailed t-test for statistical significance, and the t-value = 32.4, p < 0.000 (see table 5.1 (e)), this indicates that the difference between the mean values is statistically significant.

<table>
<thead>
<tr>
<th>Paired Samples Test</th>
<th>Paired Differences</th>
<th>t</th>
<th>df</th>
<th>Sig. (2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>32.42</td>
<td>7</td>
<td>.000</td>
</tr>
<tr>
<td>Unasp. - asp.</td>
<td></td>
<td>3.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>51.90</td>
<td>1.6005</td>
<td>95% Confidence Interval of the Difference</td>
<td></td>
</tr>
<tr>
<td>Deviation</td>
<td>3.2010</td>
<td>3.2010</td>
<td>Lower Upper</td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>-</td>
<td>-56.9936</td>
<td>-46.8064</td>
<td></td>
</tr>
<tr>
<td>Error</td>
<td>0.00</td>
<td>1.6005</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>0.00</td>
<td>1.6005</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deviation</td>
<td>0.00</td>
<td>1.6005</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>0.00</td>
<td>1.6005</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Error</td>
<td>0.00</td>
<td>1.6005</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5.1 (e): T-test of the mean values for the voiceless aspirated and unaspirated stops. Note: the uvular place for the unaspirated stops was not included.

As regards the effect of the following vowel, again a similar observation to that made for the voiced and the voiceless unaspirated stops can be noted with the voiceless aspirated stops also: that stops followed by high vowels tend to display longer VOT values than those which are followed by non-high vowels. This seems to be more consistent with the voiceless stop in this study. The rank order is as follows: /u > i > e/e > o/o > a / (see table 5.1 (f)).

With regard to the effect of place of articulation, the palatal place exerts more influence on VOT than other places of articulation since stops produced there manifest higher VOT values than those produced at other places (see table 5.1 (f)). The ranking order of place of articulation is as follows: palatal > velar > bilabial > dental. We discuss possible explanations for the effect of vowel identity and place of articulation in section 5.7.
A comparison of the average prevoicing and VOT values presented in table 5.1 (a) indicate that voice onset time may well serve as an adequate basis for distinguishing the three voicing categories in Shekgalagari. VOT resolution for the three stop categories is very clear. Shekgalagari appears to locate its stops between, approximately, the means -108.0 and -80.7 ms for the voiced stops, +12.6 - +40.4 for the voiceless unaspirated stops and +63.7 - +97.0 for the voiceless aspirated stops.

The time values for the stop types tables 5.1 (a) are summarised in a graph of time distribution in graph IA.
Mean prevoicing and VOT for Shekgalagari stop contrasts

<table>
<thead>
<tr>
<th>stop categories</th>
<th>Voiced</th>
<th>Vls Unasp</th>
<th>Asp</th>
</tr>
</thead>
<tbody>
<tr>
<td>bil</td>
<td>-88.5</td>
<td>14.1</td>
<td>64.5</td>
</tr>
<tr>
<td>den</td>
<td>-80.7</td>
<td>12.6</td>
<td>63.7</td>
</tr>
<tr>
<td>pal</td>
<td>-108</td>
<td>40.1</td>
<td>97</td>
</tr>
<tr>
<td>vel</td>
<td>-83.6</td>
<td>31.8</td>
<td>81.3</td>
</tr>
<tr>
<td>uvu</td>
<td></td>
<td>15.7</td>
<td></td>
</tr>
</tbody>
</table>

Graph 5.A: Mean prevoicing and VOT for Shekgalagari stop contrasts.
Graph 5.A summarises the variation between Shekgalagari stop consonants by way of voice onset time. The stop categories are plotted on the X-axis and the time (ms) and the time distribution on the Y-axis. It is immediately clear that VOT values for the three categories are well distinguished from each other. Aspirated sounds display high positive VOT values, voiced sounds high negative values (and sometimes a small positive VOT value after the burst) and the unaspirated sounds short positive VOT values. Thus different stop types which are produced at the same place of articulation are produced at non-overlapping places along the timing dimension. For the stops produced at the bilabial place, for instance, [b] may be found on the long negative half of the timing parameter, and [p] and [pʰ] on the short and long positive half respectively.

This graph also shows the effect of place of articulation on the duration of prevoicing and VOT. For the voiceless stops, both unaspirated and aspirated, the palatal stops show longer mean VOT duration than the stops produced at other places of articulation. Graph 5.A indicates that the places of articulation are ranked in the following order: /c > k > q > p > t/ for the unaspirated stops and /cʰ,kʰ,pʰ,tʰ/ for the aspirated stops, (as we observed earlier). For the voiced stops the effect of place of articulation on VOT appears to be in the following order: /j > b > g > d > l/. In section 5.7, the interaction between values for prevoicing and VOT and place of articulation is discussed and a possible explanation of this relationship is pointed out in terms of vocal tract configuration, the speed of the body of the tongue and aerodynamic conditions which may be involved. Similar explanations are suggested for the effect of vowel identity (discussed below) on the values for prevoicing and VOT for the stops.

The three-way contrasting languages reported in similar studies include Thai, East Armenian and Korean. Of these, Thai and East Armenian display a similarity with Shekgalagari in that the distribution of their stops in the time dimension ranges falls on both halves of the timeline: from -115 ms to -96 ms for the voiced stops, +3 ms to 30 ms for the voiceless unaspirated stops, and +58 ms to +78 ms for the voiceless stops for East Armenian (Lisker & Abramson 1964); and in Thai -106 ms to -104 ms for the voiced stops, +5 ms to +23 ms for the voiceless unaspirated stops and +73 ms to +95 ms for the voiceless aspirated stops (Shimizu 1990). The other point relates to the rank order of the effect of place of articulation on the time values. In Shekgalagari, contrary to popular report (e.g. Lisker & Abramson 1964, Shimizu 1990), velar stops come second to palatals with respect to the magnitude of VOT, at least for the voiceless stops.³

³ Perhaps I should mention here that, as I have already discussed in Chapter Two, the palatals are sometimes classified with the velars as dorsal sounds. But this classification still remains controversial amongst phoneticians. However, if we go with
Korean stops are all voiceless and VOT can only adequately separate the heavily aspirated from the other two: the unaspirated and the slightly aspirated, for which VOT resolution is not quite clear.

The mean duration for Shekgalagari stops as a function of vowel identity are also plotted in graph 5.B.

---

the view that they are dorsals, then the VOT values for the palataals may well be consistent with the general view that dorsals exhibit longer VOT values than stops produced at other places of articulation. A second point to mention is that most of the languages reported in the literature do not have palatal stops. Hence a comparison of VOT values for the velar stops with those for the palatal stops seems to be non-existent in the literature.
The effect of vowel quality on prevoicing and VOT for Shekgalagari stops

Graph 5B: The effect of vowel quality on prevoicing and VOT for Shekgalagari stops
Graph 5.B clearly displays the observation made earlier about the effect of vowel identity on the mean duration of timing for the stops. Generally, stops which are followed by high vowels manifest longer prevoicing (if they are voiced) and longer lag (if they are voiceless) than stops which are followed by non-high vowels. The order of the vowels in affecting the duration values is in the following order. For the voiceless unaspirated stops: /i > u > o/o > e/e > a/; for the voiceless aspirated stops: /u > i > e/e > o/o > a/; and for the voiced stops: /i > e/e > u > o/o > a/. In section 6.7 the relationship is discussed between vowel context and time duration for the stops in the light of two hypotheses which have been postulated as possible explanations for this observation: vocal tract configuration and the speed of the body of the tongue.

As was mentioned earlier, the data was also analysed in order to observe inter-speaker variation. Table 5.2 presents the results of the analysis speaker by speaker. This table shows the ranges (R), average values for prevoicing and VOT and the standard deviations (SD) for the stops and vowels respectively for each subject.
<table>
<thead>
<tr>
<th></th>
<th>CM Range</th>
<th>MN</th>
<th>SD</th>
<th>EN Range</th>
<th>MN</th>
<th>SD</th>
<th>MK Range</th>
<th>MN</th>
<th>SD</th>
<th>TM Range</th>
<th>MN</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>[b]</td>
<td>-135 - -81.3</td>
<td>-105.6</td>
<td>19.6</td>
<td>-141.6 - -67.3</td>
<td>-95.0</td>
<td>28.2</td>
<td>-118.1 - -39.4</td>
<td>-76.6</td>
<td>29.9</td>
<td>-106.4 - -57.3</td>
<td>-73.8</td>
<td>22.2</td>
</tr>
<tr>
<td>[d]</td>
<td>-142 - -105.8</td>
<td>-116.7</td>
<td>14.7</td>
<td>-105.9 - -74.9</td>
<td>-91.7</td>
<td>14.7</td>
<td>-100.3 - -41.6</td>
<td>-71.5</td>
<td>24.2</td>
<td>-94.9 - -47.4</td>
<td>-78.8</td>
<td>21.5</td>
</tr>
<tr>
<td>[j]</td>
<td>-143.7 - -107.3</td>
<td>-120.8</td>
<td>14.2</td>
<td>-190 - -103.3</td>
<td>-128.4</td>
<td>35.2</td>
<td>-95.2 - -71.5</td>
<td>-84.3</td>
<td>9.1</td>
<td>-111.7 - -73.4</td>
<td>-96.0</td>
<td>16.1</td>
</tr>
<tr>
<td>[g]</td>
<td>-140.9 - -64.6</td>
<td>-96.1</td>
<td>36.0</td>
<td>-95.3 - -71.0</td>
<td>-80.6</td>
<td>10.5</td>
<td>(no prev. voices)</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
</tr>
<tr>
<td>[p]</td>
<td>4.1 - 12.3</td>
<td>9.0</td>
<td>3.2</td>
<td>9.9 - 22.3</td>
<td>16.8</td>
<td>5.5</td>
<td>4.3 - 24.9</td>
<td>11.5</td>
<td>7.2</td>
<td>6.0 - 33.9</td>
<td>18.4</td>
<td>10.4</td>
</tr>
<tr>
<td>[t]</td>
<td>4.1 - 12.1</td>
<td>7.2</td>
<td>3.6</td>
<td>9.0 - 19.4</td>
<td>15.0</td>
<td>5.2</td>
<td>5.4 - 16.2</td>
<td>9.9</td>
<td>4.5</td>
<td>9.0 - 31.0</td>
<td>18.1</td>
<td>9.4</td>
</tr>
<tr>
<td>[c]</td>
<td>18.5 - 37.3</td>
<td>28.5</td>
<td>9.2</td>
<td>33.2 - 70.7</td>
<td>54.4</td>
<td>18.7</td>
<td>17.9 - 44.6</td>
<td>28.4</td>
<td>11.4</td>
<td>33.2 - 69.4</td>
<td>50.3</td>
<td>15.4</td>
</tr>
<tr>
<td>[k]</td>
<td>19.2 - 34.4</td>
<td>24.8</td>
<td>7.4</td>
<td>26.8 - 53.5</td>
<td>40.6</td>
<td>9.7</td>
<td>12.6 - 36.6</td>
<td>21.2</td>
<td>9.1</td>
<td>26.8 - 53.3</td>
<td>40.6</td>
<td>9.7</td>
</tr>
<tr>
<td>[q]</td>
<td>7.1 - 13.3</td>
<td>9.3</td>
<td>2.7</td>
<td>12.1 - 24.7</td>
<td>18.9</td>
<td>6.6</td>
<td>7.0 - 24.8</td>
<td>15.8</td>
<td>7.1</td>
<td>12.1 - 26.5</td>
<td>18.9</td>
<td>6.6</td>
</tr>
<tr>
<td>[p']</td>
<td>57.7 - 79.4</td>
<td>66.4</td>
<td>9.0</td>
<td>43.2 - 87.2</td>
<td>60.9</td>
<td>16.8</td>
<td>66.0 - 90.3</td>
<td>70.5</td>
<td>14.1</td>
<td>44.8 - 91.1</td>
<td>60.6</td>
<td>18.9</td>
</tr>
<tr>
<td>[t']</td>
<td>55.0 - 77.7</td>
<td>64.1</td>
<td>8.8</td>
<td>30.3 - 71.3</td>
<td>62.0</td>
<td>18.9</td>
<td>55.3 - 75.5</td>
<td>63.2</td>
<td>7.9</td>
<td>36.7 - 93.0</td>
<td>65.3</td>
<td>22.0</td>
</tr>
<tr>
<td>[c']</td>
<td>68.0 - 98.3</td>
<td>78.9</td>
<td>12.3</td>
<td>71.4 - 116.9</td>
<td>95.3</td>
<td>17.5</td>
<td>80.8 - 123.3</td>
<td>102.1</td>
<td>17.9</td>
<td>80.8 - 145.6</td>
<td>111.5</td>
<td>30.8</td>
</tr>
<tr>
<td>[k']</td>
<td>61.1 - 82.0</td>
<td>71.1</td>
<td>9.0</td>
<td>75.5 - 95.3</td>
<td>87.9</td>
<td>7.8</td>
<td>64.2 - 98.2</td>
<td>80.8</td>
<td>14.8</td>
<td>61.3 - 128.0</td>
<td>86.2</td>
<td>26.7</td>
</tr>
</tbody>
</table>

Table 5.2: Means (MN) (ms). Ranges and standard deviations (SD) for the three stop types for the four speakers. The means represent the averages for each speaker per stop type and across vowel contexts. SDs were calculated from the values obtained for the data for each speaker. The ranges indicate each speaker's time range for each stop type across vowel contexts.
With respect to table 5.2, each of the individual speakers shows long positive VOT mean values for the voiceless aspirated stops, and short positive VOT values for the unaspirated stops. Speaker CM tends to manifest VOT values which are somewhat lower than those of the other speakers for the voiceless and unaspirated stops, and for the voiceless aspirated palatal stops. For the palatal stops, all the four speakers show VOT values which are longer than those of stops produced at other places of articulation. Speaker MK does not show any prevoicing for the voiced velar stops, but only a very small positive VOT.

From the presentation of results above, we have so far noted four factors which may have an effect on the time values (prevoicing and VOT) obtained for the stops: place of articulation (Place), vowel context (Vowel), voicing contrast (Voicing) and the identity of the speaker (Sex). For the place of articulation there were four places involved; bilabial, dental, palatal, and velar (and, of course, five places for the voiceless unaspirated stops which are also produced at the uvular place). Five vowel qualities were involved for vowel context: \[i, e/e, a, o/\sigma, u\], and four for the speakers: CM, EN, MK and TM. The means obtained for the stops were statistically explored and subjected to an ANOVA test in order to examine the level of significance of these factors and the interaction between them. The results are shown in the following tables and figures.

Voicing was explored with VOT and prevoicing values as dependent factors. The results are shown on table 5.3 as well as the associated graph, i.e. figure 5.4.
The horizontal line in the middle of the black bars represents the mean. The top edge of the bars represents the standard deviation (SD) above the mean, and the bottom edge the SD below the mean. The top-most horizontal line represents the highest measured value(s) for the category, and the bottom-most the lowest value.

It can be observed from figure 5.4 that there is no overlap between the three voicing contrasts. The aspirated stops show high VOT values, the unaspirated short values and the voiced show values that are below 0 ms.

An interaction between vowel and voicing was also explored. The results are shown in table 5.4 and the associated graph in figure 5.5.
### Case Processing Summary

<table>
<thead>
<tr>
<th>VOICING</th>
<th>VOWEL</th>
<th>Cases</th>
<th>Valid</th>
<th>Missing</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>N</td>
<td>Percent</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VOT</td>
<td>asp</td>
<td>a</td>
<td>16</td>
<td>100.0%</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>i</td>
<td>16</td>
<td>100.0%</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>o</td>
<td>16</td>
<td>100.0%</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>u</td>
<td>16</td>
<td>100.0%</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>unasp</td>
<td>a</td>
<td>16</td>
<td>100.0%</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>i</td>
<td>16</td>
<td>100.0%</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>o</td>
<td>16</td>
<td>100.0%</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>u</td>
<td>15</td>
<td>93.8%</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>voiced</td>
<td>a</td>
<td>14</td>
<td>87.5%</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>i</td>
<td>15</td>
<td>93.8%</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>o</td>
<td>13</td>
<td>81.3%</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>u</td>
<td>14</td>
<td>87.5%</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 5.4: Exploration of interaction between vowel quality and voicing contrast: case-processing summary. N represents the number of cases.

Figure 5.5: Exploration of the interaction between vowel quality and voicing contrast (with VOT as a dependent feature). N represents the number of cases for each category.
From figure 5.5 it can be seen that the vowel /i/ appears to be exerting more influence on VOT values than the other three vowels, leading to higher VOT and prevoicing values. The ranking of vowel influence on VOT and prevoicing appear to be in the following order: /i > u > o > a/ where / > / means ‘exerts more influence than’.

Interaction between place of articulation and voicing was also explored. The results are shown in table 5.5 and figure 5.6.

Table 5.5: Exploration of the interaction between place of articulation and voicing contrast: case-processing summary.
Although the difference seems to be rather small, it can be observed from figure 5.6 that the palatal stops appear to be showing higher VOT and prevoicing values than stops produced at other places of articulation. They are followed by the velar stops, then the bilabial and lastly the dental stops.

A univariate Analysis of Variance (ANOVA) was performed on the data in order to test the level of significance of the factors as well as that of the interaction between them. The results are shown in table 5.6. Full details of this analysis may be found in Appendix 5E.
<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>F-value</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Voicing</td>
<td>2</td>
<td>1623.182</td>
<td>.001</td>
</tr>
<tr>
<td>Place</td>
<td>3</td>
<td>81.965</td>
<td>.002</td>
</tr>
<tr>
<td>Vowel</td>
<td>3</td>
<td>3.526</td>
<td>.164</td>
</tr>
<tr>
<td>Sex</td>
<td>1</td>
<td>.559</td>
<td>.521</td>
</tr>
<tr>
<td>Voicing * Place</td>
<td>6</td>
<td>27.918</td>
<td>.000</td>
</tr>
<tr>
<td>Place * Vowel</td>
<td>9</td>
<td>2.891</td>
<td>.065</td>
</tr>
<tr>
<td>Place * Sex</td>
<td>3</td>
<td>.552</td>
<td>.742</td>
</tr>
<tr>
<td>Vowel * Sex</td>
<td>3</td>
<td>4.430</td>
<td>.606</td>
</tr>
<tr>
<td>Voic.<em>Place</em>Vowel</td>
<td>18</td>
<td>.462</td>
<td>.945</td>
</tr>
<tr>
<td>Voic.<em>Place</em>Sex</td>
<td>6</td>
<td>.467</td>
<td>.824</td>
</tr>
<tr>
<td>Voic.<em>Vowel</em>Sex</td>
<td>6</td>
<td>.307</td>
<td>.926</td>
</tr>
<tr>
<td>Place<em>Vowel</em>Sex</td>
<td>9</td>
<td>.908</td>
<td>.538</td>
</tr>
<tr>
<td>Voic.*Pla.<em>Vow</em>Sex</td>
<td>18</td>
<td>.545</td>
<td>.928</td>
</tr>
</tbody>
</table>

Table 5.6: ANOVA test of between-subjects effects.

It can be observed from table 5.6 that, for independent factors, the effect of voicing and place of articulation were highly significant statistically, whilst vowel quality was not so significant. For the two-way interactions, voicing-by-place and voicing-by-vowel quality were highly significant than place-by-vowel context. The rest of the interactions were not statistically significant.

5.7 Discussion

5.7.1 Effectiveness of VOT in distinguishing Shekgalagari voice types

VOT values for the three stop types presented in the results subsection above show long negative values for the voiced stops (which included voice decay during the stop occlusion, but not the short positive VOT after the burst), short positive values for the voiceless unaspirated stops and long positive values for the voiceless aspirated stops. There is no overlapping in time values between the three
stop types, and the ranges of the means between the three stop types are considerable. We may, therefore, consider VOT to be an effective distinguishing cue for the three stop types in Shekgalagari.

5.7.2 The effect of place of articulation and vowel quality on VOT

As has often been observed for other languages and reported in the literature, this investigation has shown that there is some interaction between VOT and place of articulation, and the quality of the following vowel. With respect to the relationship between VOT and place of articulation, my results have consistently showed that the palatals manifest larger VOT values than stops produced at other places of articulation for the voiceless stops. The velars (in most cases) come second to the palatals, which are then followed by the uvulars and the dentals and the labials. This observation does not conform in any simple way to the two hypotheses that have been pointed out to account for differences in VOT as a function of place of articulation. The first hypothesis is related to the aerodynamic factors involved due to the volume of the cavities above the larynx, and the second to an articulatory factor based on the size of the tongue body involved and its relative speed in the production of the stop.

According to the first hypothesis, as the tongue moves further back in the mouth, it gradually reduces the size of the supraglottal cavity behind. As a result of this, there is a high level of air pressure behind the obstruction for articulations that are made further back in the mouth. When the obstruction is released, there is some delay in balancing air pressure below and above the glottis, causing delay in the commencement of voicing for the subsequent vocalic segment. The further back in the mouth the obstruction is, the longer it will take for the development of air pressure required for voicing (Klatt 1975, Weismer 1980). In most of the languages reported in the literature, there has been a tendency for velar stops to have longer VOT values, followed by alveolar stops and lastly the bilabial stops. Lisker and Abramson (1964) made the earliest observation of this. Many subsequent studies have since made similar observations, e.g. Klatt (1975, Shimizu 1990).

According to the second hypothesis, VOT is affected by the speed of the active articulator involved. It has long been observed that larger parts of the tongue body, e.g. the back part, move relatively slower than the smaller parts, e.g. tip and blade. This means that back articulations are released relatively more slowly than front articulations. A consequence of this in terms of timing is that it
takes longer for the aerodynamic conditions necessary for voicing to be met in back articulations than in front articulations (Hardcastle 1973, Klatt 1975).

If we follow the first hypothesis, we should expect VOT values for the stops to increase as the place of articulation moves further back in the mouth. However, this prediction does not hold in the results obtained in this study, both for individual results and for pooled data. The VOT values obtained for Shekgalagari stops in this study ranks the stops in the following order: /c/ > /k/ > /q/ > /p/ > /t/. Contrary to expectation, uvular stops do not manifest larger VOT values than the stops uttered at anterior places of articulation. Rather, more agreement with other results may perhaps be considered if we regard the bilabial and dental stops as front articulations and the palatal, velar and uvular stops as back articulations. Also, whereas the classification of the velar place has always been clear — dorsal, that of the palatal has remained largely uncertain. Ladefoged (1993:7) writes that ‘palatal sounds are sometimes classified as coronal articulations, and sometimes as dorsal articulation’. Taking the side that classifies them as dorsals, then VOT values for the palatal stops may well be in accordance with the expectation.

According to hypothesis two, palatals manifest longer VOT values than stops produced at other places of articulation because they are made with a relatively large size of the body compared to the other stop. This means that the speed of the tongue is greatly reduced and thus takes longer to move towards and away from the palate after making contact, making the oral cavity remain constricted for a relatively longer time. The direct consequence of this is that it therefore takes longer for the air pressure inside the mouth to balance with that which is outside the mouth. At the glottis, the development of airflow required for voicing thus also takes longer, resulting in longer VOT (and prevoicing) values for the palatal stops. High vowels exert a similar influence on VOT values because their manner of articulation (i.e. raised tongue position) means that the oral cavity is narrowed. This configuration delays the achievement of the transglottal air pressure drop necessary for voicing, and VOT is larger than for other vowels. Palatal stops display similar articulatory configurations to high vowels, and this, in addition to the slower speed of the tongue, means that palatals manifest higher VOT values than that for other places of articulation.

As mentioned earlier, most of the languages investigated in the literature do not have palatal and uvular articulations, and the contribution of these particular places of articulation to VOT has therefore virtually been non-existent. We may still speculate that the velar position in these languages may possibly be a little fronted than in languages like Shekgalagari, giving them longer VOT values.
For the bilabial stops, table 5.1 reflects some relationship between tongue position for the vowel and VOT. For these stops the tongue is only involved in the production of the vowel, since it is the lips which make a constriction against each other for the stop. Consequently VOT values reflect interaction with tongue height for the vowel: bilabials which are followed by high values manifest longer VOT values than those which are followed by non-high vowels.

In the results obtained here, bilabial stops show VOT and prevoicing values which are higher than those of the dental stops, when, in fact the tendency in other languages has been the reverse: dental > bilabial. This result was not expected for physiological reasons. Since there is no cavity in front of the mouth, it is expected that once the constriction made at the lips is released, the oral air pressure will quickly balance with the atmospheric pressure and transglottal air flow quickly restored to the level required for voicing. This should produce VOT and prevoicing values which are lower than those for the dental place, which is posterior to the lips. That results here do not accord with this possibly means that these stops need to be investigated further.

Nevertheless, hypothesis two discussed above seems to account for the results obtained here better than hypothesis one.

The other factor that may affect VOT relates to the effect of the following vowel. It has been observed that, generally, stops (the velar stop /k/ being the most commonly reported) in syllables with high vowels tend to manifest greater VOT values than those in syllables with non-high vowels, e.g. Klatt (1975) and Port (1979). A possible explanation that has commonly been given for this is that the raising of the tongue for the high vowel narrows the oral cavity and creates some resistance to the flow of air from the lungs. This in turn delays the development to transglottal air pressure necessary for vocal fold vibration.

The data also shows that there is some interaction between the height of the vowel and VOT value for the stop in the syllable. The influence of the vowels has shown that [u] and [i] have always exerted more influence on VOT than [o/ɔ] and [e/ɛ] with [a] showing the least influence on VOT, and, for the voiceless stops, this trend appears to be fairly consistent in these data.

The VOT values obtained here for the voiceless unaspirated stops in Shekgalagari have generally been low (see table 5.1 (a)). Although higher values were obtained in some cases, this seem to accord more with the effect of place of articulation than with the voice type for the stops, as we explained above. Had high VOT values been a feature of these stops in this language, it would have occurred across all the five places of articulation. Even the high values obtained for some of these stops were no-where near the values obtained for the voiceless
aspirated stops, as has been reported for ejectives in some North American Indian and Caucasian languages. This is strongly suggestive of the possibility that the voiceless unaspirated stops in Shekgalagari may not be ejectives, at least as far as VOT is concerned.

With regard to the voiced stops voicing occurred for a considerable duration during the stop occlusion. Where there was voicing decay (for some of the speakers), this was relatively small. We may therefore regard these stops as truly voiced. The values obtained for the voiceless unaspirated stops seem to be consistent with those that are reported for other languages (see Shimizu 1990).

5.8 Summary

On the whole, the findings of this study with respect to VOT have demonstrated that this timing dimension can effectively distinguish Shekgalagari homorganic stops in word initial position. The three types of stop manifest considerable significant variation in VOT values: long prevoicing values for the voiced stops, short VOT lag for the voiceless unaspirated stops and long VOT lag for the aspirated stops; and this has been remarkably consistent for all the data.

VOT may be affected by place of articulation and vocalic context. In this study, palatal stops have manifested higher VOT and prevoicing values, followed by the velar stops and the by the uvulars. This has not been reported in the literature because most languages which have been analysed to not have palatal (and uvular) stops. Stops followed by high vowels also show duration values which are higher than those shown by stops followed by non-high vowels.

Other factors which may affect the value of VOT include the effect of the manner of production: citation form and within a frame sentence, the rate of speaking, and the influence of the second language; (English). It has long been observed that acquiring sounds from an L2 language may affect the production of sounds in the mother language (Shimizu 1990:42) The other factor which may affect VOT in this study is tone, but this was not investigated in this study, and recommend it for future research.

From the findings of this and the previous chapter, the subsequent discussions, it seems reasonable to conclude that Shekgalagari is a three-way category language, with voiced, voiceless unaspirated and voiceless aspirated stops, and belongs to the class of languages of type III in table 5.7 (adapted from Monaka, Abberton & Harris 1997) (see § 3.3.6).
It may also be concluded, at least on the basis of the findings of the previous and the present chapters, that the voiceless unaspirated stops in Shekgalagari may be simple plain voiceless unaspirated stops, and not ejectives. This conclusion is, however, tentative, since more investigation needs to be done on these stops, as well as more cross-linguistic analysis made, in order to determine whether these stops may possibly be weak ejectives. For example measurement of the burst amplitude, of closure duration, of intra-oral air pressure during the hold phase amongst other things.

\footnote{That is, Southern German.}
6 Articulatory Analysis: Electropalatography

6.1 Introduction

The aim of this chapter is to examine lingual articulatory characteristics of Shekgalagari stops by means of electropalatography (EPG). Qualitative and quantitative description will be provided on the type and extent of articulatory movements for Shekgalagari stops and for the variation observable between the voice types. Coarticulation will also be investigated by assessing lingual-palatal contact patterns for V-to-C co-articulation in symmetrical and asymmetrical vowel contexts, i.e. V₁C-V₂, focusing on variation between the voice types and place of articulation. Where relevant, the relationship between EPG data and stop duration measured acoustically will also be discussed. The findings of the experiment in this study will be discussed in the light of constraints on the tongue during the production of sounds, theories proposed to account for co-articulation in speech and in the light of the findings reported in the literature for similar studies in other languages.

The rest of the chapter proceeds as follows. Section 6.2 discusses some theoretical background information relating to the nature of tongue-to-palate contact patterns for the different voicing types for the stops and their co-articulatory behaviour. Section 6.3 deals with the instrumentation, conditions of recording, data and the informant of the study. Section 6.4 describes the method of analysis adopted in this investigation. The results are presented in section 6.5, followed by the discussion of the results in section 6.6. The results are discussed in the light of theories proposed to account for co-articulation in speech presented in Chapter Three. Section 6.7 summarises the whole chapter.

6.2 Theoretical background

6.2.1 Lingual-palatal contact patterns for the articulator and co-articulatory behaviour of segments

Different voicing types of stops as well as stops produced at different places of articulation seem to have differing tongue-to-palate contact characteristics (see, for example, Shimizu [1990:48-52], Fujimura, Tatsumi and Kagaya [1973] on Japanese, Kang [1998:19 ff] on Korean affricates and fricatives and Shin (1997) on Korean stops and affricates. For example, Shimizu (1990:52) observed that for
the Japanese [ta] and [da], ‘[ta] shows a greater contact area than [da], and the duration of maximum contact is longer in [ta] than in [da].’ Similar observations were made by Fujimura, Tatsumi and Kagaya (1973:53) whose comparison of palato-lingual contact patterns for [t] and [d] in Japanese showed some correlation between the degree of contact and the duration of closure, and that the distinction between contact patterns for these stops may be ‘characterised by different program values in one control dimension, which may be labelled as tense-lax.’ For the Korean affricates, Kang (1998:19-20) observed that, compared with the aspirated and the so-called ‘tense’ affricates, the lax type showed lesser tongue-to-palate contact area and that this difference appears ‘to be a common feature of the stops and affricates.’ Similar observations were made by Shin (1997) for Korean stops and affricates. Characteristics of tongue to palate contact for Shekgalagari stop production, both by voicing type and by place of articulation are investigated in the sections below.

6.2.2 Co-articulatory behaviour of segments

Coarticulation refers to a point in time when the configuration of the vocal tract is influenced by more than one sound (Farnetani 1997:371). This happens when the organs of speech interact with each other and overlap in time during the articulation of juxtaposed sounds. It is similar to a related process of assimilation which may refer to variation in speech sounds where one or more of a particular sound’s phonetic properties are altered and become similar to those of the neighbouring segments. But this definition could also be relevant to co-articulation. Both processes are context dependent. In reality, they overlap with each are very hard to distinguish from each other. Co-articulation and assimilation have already been discussed in detail in section 3.6. Nevertheless, whether and/or how co-articulation is distinctive from assimilation is still a matter of controversy, and will not be dealt with here. Coarticulation is not necessarily always audible. It may be described in a number of ways, e.g. in terms of the main articulators involved i.e. mechanically, or in linguistic/phonological terms (see Ibid. for more details). Coarticulation has been observed in most, i.e. not all of the languages so far investigated, and it seems to differ from language to language (Farnetani 1997:376).

The different parts of the tongue, particularly the tongue tip/blade and the tongue body can behave in a quasi-independent manner from each other in the production of sounds, resulting in overlap in time in the production of juxtaposed segments and different tongue configurations for different sound types. The
different stop types can manifest different co-articulatory behaviour in relation to different vowel contexts (Farnetani 1997:388 and references therein). Stops may also manifest different co-articulatory behaviour depending on their voice types and their places of production. This chapter will investigate the co-articulatory behaviour of Shekgalagari stops by voice types and by place of articulation.

6.3 Instrumentation, recording conditions, data and subjects.

The basic instrument used in this study is an artificial palate with embedded electrodes that respond to tongue contacts. This has been described in detail in Appendix 1A. Figure 6.1 shows the Reading artificial palate similar to the one used in this study, and an accompanying model showing the positions of the electrodes when an informant wears the artificial palate during the recording.

![Image of artificial palate](image)

Figure 6.1: The Reading EPG artificial palate similar to the one used in this study, and the plaster impression showing the position of the electrodes in the mouth (Hardcastle et al. 1989:3).

The electrodes mounted on the artificial palate are connected to lead-out wires, which come out of the corners of the informant’s mouth to connect directly to a computer which in turn stores data continuously during the recording. This data can then be analysed in real time or stored for future processing. In this study, the data was sampled at a rate of 300 frames per second, and the interval between the
frames was approximately 3.3 ms. A simultaneous audio recording was also made to aid later identification of the target sequences on the EPG frames. A digital audio recording was made into one channel of a Sony Digital Audio Tape Deck using a SHURE microphone 515 SD UNIDYNE B through a MaPLIN MPX-55 mixer/preamp. The waveform displays and spectrogram were computer displayed using Loughborough Sound Images Speech Workstation (LSWISW) software on a PC 686 MX 233 MHz Pentium 24 Megabytes of RAM. A sampling rate was 20 kHz. To synchronize the simultaneous recordings, a reference signal was made before each group of utterances by tapping a ruler on a desk, thereby enabling the two time scales to be matched using the tap as the starting point.

Data recorded for analysis consisted of a list of 81 words of the form CV\textsubscript{CVCVCV:CV}\textsuperscript{1} produced in isolation, where the target C could be [t\textsuperscript{b}, t, d/ c\textsuperscript{b}, c, \textsuperscript{j}/ k\textsuperscript{h}, k, g], and the flanking vowels could be [a, i, u]. For the relevant VCV sequence, all possible combinations of the stop and vowels were employed. Nonsense words were used where meaningful could not be found or generated. The target stop was in the antepenultimate syllable, which was suitable for controlling the effect of penultimate lengthening, if any. There was no effort made to control tone since this would have made the generation of meaningful words for all data almost impossible. The data can be found in Appendix 6F. The list was read five times, giving a total of 405 words, at a normal speaking rate, on the same day by one female speaker of the Shengologa dialect of Shekgalagari, who served as the informant.

6.4 Processing.

6.4.1 Divisions of the palate

Figure 6.2 shows a representation of a stylised EPG palate frame print-out.

\textsuperscript{1} As pointed out in Chapter Two (§ 2.4.3), Shekgalagari, like all South-eastern Bantu languages, has typical penultimate length (\textsuperscript{.})
The electrodes are arranged in eight horizontal rows. The front row carries six electrodes, and the rest eight. The Os indicate contact between the palate and the tongue, and reflect electric conductance as the circuit is completed by this contact. In order to see the overall shape of contact — which region has more contact, and how long the main constriction lasts, over time — the palate can be divided into regions, and this differs from author to author. A detailed discussion of the different divisions of the palate by different authors may be found in Shin (1997:70-74). It is also worth pointing out that different authors have adopted different methods of representing the electrodes. Some authors use circles or dots (e.g. Gibbon (1990:9), and others used rectangles or boxes (e.g. Gibbon et al. (1993:272) and Recasens (1993:219)).

In this study, a four-way division of the palate with rows 1-3 corresponding to the alveolar/dental place of articulation, row 4 to post-alveolar, 5-7 to the palatal and row 8 to the velar region is adopted. This is shown in figure 6.3.
This division was chosen because, from initial observation, the dental stops showed constriction at rows 1 to about the 3rd or 4th, palatal constriction occurred between rows 4 and 7 and velar from around row 7 to 8. Since the artificial palate can only go as far as the divisions between the soft and hard palate, uvular stops, which are produced beyond the velar place, cannot be used in this study (or for that matter in any study using the Reading artificial palate).

6.4.2 Criteria for analysing contact pattern and area

The contact pattern for stops is often described by observing lingual-palatal characteristics at maximum point of contact (MC). The number of on-electrodes is counted at MC as well as the differences in frequency of contact for segments in order to determine the area of contact for different segments. As would be expected, this varies with repetitions of the target token, both when the context is the same and when it is different. A number of proposals have therefore been made regarding how to analyse and represent tongue-palate contact. Shin (1997:59) groups these methods into three broad classes: the cut off method, the frequency representing method and the average representing method.

In the cut-off method, it is thought that tongue contact area for a stop may be described in terms of a common area of contact across the repetitions for the stop. This common area represents electrodes which are contacted 80% of the
time or more every time a token is repeated. The 80% is an arbitrary 'cut-off' point and used as a reference point for analysing and describing contact pattern and area for stops. Thus contact pattern for a stop is based on the lingual-palatal characteristics at a mutual point, i.e. where 80% of contact happens for all the repetitions of the token. Contact elsewhere is excluded from analysis. The 'cut-off' terminology carries the implication that the method virtually eliminates variation between repetitions of tokens, and this is the disadvantage of this method of analysis. Otherwise its advantage is that lingual-palatal characteristics of stops can be clearly displayed.

The cut-off method may be divided into the simple and the overlap type. In the simple type, lingual-palatal contact pattern for each segment is represented separately and analysed in a straightforward way as described above. The main problem with this type is that the difference in contact pattern for the different voice types of stops (which the overlap method attempts to show) may not be easy to spot quickly. This type is shown in figure 6.5. In this figure the authors represent the palate and the electrodes differently, based on the division of the hard palate shown in figure 6.4.

Figure 6.4: Division of the hard palate according to Dixit and Flege (1990:217).
The overlap method attempts to represent tongue to palate contact pattern for more than one segment, at a particular point in time, in a single stylised EPG.
frame. But showing more than three segments at a time is still neither easy to perform nor clear in displaying the necessary information.

The frequency representing method displays on a single stylised frame variation in the articulation of a token repeated a number of times, and is also divided into the simple and the overlap method. In the simple method, variation between repetitions of a token is shown by colouring or shading activated dots or boxes to varying degrees depending on the frequency of contact. Consider figure 6.6.

![Figure 6.6: Lingual-palatal contact for the English [kl] sequence and the [k] stop for one subject. The legend shows the frequency of contact at particular points on the palate (Gibbon et al. 1993:272).](image)

One of the disadvantages has often been that variation between articulatory patterns for repetitions of tokens has not been easy to distinguish clearly since the extents of the colouring of the dots or boxes has in some cases been hard to distinguish from each other.

The overlap version of the frequency representing method attempts to display variation between different voice types for the stops (e.g. t/d vs. k/g), by indicating the frequency of contact in each electrode. But this also can only use at most three schematised EPG frames. See figure 6.7.
The third method, the average representing method, is based on taking the mean value for contact between repetitions for a target sound. This mean value may then be represented as contour lines on a schematised artificial palate. An example of this representation is shown on figure 6.8.

Figure 6.7: The overlap type of the frequency representing method. Each palatogram shows contact pattern for the voiced and voiceless stops [t/d, and k/g] (Farnetani (1989), in Hardcastle et al. (1991:254)).

Figure 6.8: A contour example of the average representation method, showing consonants produced in the context of different vowels (Recasens 1984:130).
6.4.3 Criteria for analysing contact pattern and area, and the definition of MC adopted in this study

6.4.3.1 contact pattern and area. In this study, in addition to the conventional method of counting the number of on-electrodes in the production of a stop at maximum contact, we follow Shin (1997) in extending Farnetani (1990)'s method of 'tongue profile figures.' This method was originally used to study and quantify co-articulation and to present it graphically. Here we extend it to analyse tongue-to-palate contact pattern and area for the stops as well as study the influence of the flanking vowels on the production of the stops at MC. The results are presented graphically, and in this way variation between the different voice types for the stops — in terms of lingual-palatal contact patterns, tongue configuration and susceptibility to co-articulation — may be clearly seen.

According to Farnetani's method, a percentage for the electrode activation per row is calculated in the following way:

\[
\frac{X}{Y} \times 100 = Z\%
\]

In formula (1), \(X\) is the number of on-electrodes in each row; \(Y\) is the total number of electrodes in that row and \(Z\) is the activation percentage for each row. A tongue profile graph visualising tongue contact pattern and area for each voice type may then be plotted using the percentage activation value calculated using formula (1). The procedure for calculating activation percentage and the associated graph are illustrated in Figure 6.9 (a) and (b) respectively.
Figure 6.9 (a): Procedure for calculating activation percentage

<table>
<thead>
<tr>
<th>Row</th>
<th>Percentage activation per row</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0/6 * 100 = 100</td>
</tr>
<tr>
<td>2</td>
<td>7/8 * 100 = 87.5</td>
</tr>
<tr>
<td>3</td>
<td>4/6 * 100 = 50</td>
</tr>
<tr>
<td>4</td>
<td>2/8 * 100 = 25</td>
</tr>
<tr>
<td>5</td>
<td>2/8 * 100 = 25</td>
</tr>
<tr>
<td>6</td>
<td>2/8 * 100 = 25</td>
</tr>
<tr>
<td>7</td>
<td>3/8 * 100 = 37.5</td>
</tr>
<tr>
<td>8</td>
<td>4/8 * 100 = 50</td>
</tr>
</tbody>
</table>
Figure 6.9 (b) graphs a tongue profile visualising tongue contact pattern and area for the hypothetical segment shown by the EPG frame in figure 6.9 (a), using the percentage activation values calculated in figure 6.9 (a). The place (row(s)) where the main stricture occurs is indicated by a high percentage of on-electrode activation. This high percentage of electrode activation will extend over a number of rows where the segment being produced has a longer constriction length. In figure 6.9 (b), the main constriction happens on row 1, and does not extend further to any other row. This could indicate that the segment being produced is considerably short in duration, both spatially and temporally. Contact area is the

\[\text{Contact area} = \text{Percentage activation} \times \text{Number of rows} \]

In this particular case, brief duration is shown by complete closure happening in one row, but complete closure could happen over several rows and still be of short duration. It is also possible to suggest that brief duration could be related to relative speeds by different parts of the tongue. The tongue tip in this case makes complete closure, and this part of the tongue moves relatively faster than the body of the tongue,
portion below the line and this may be calculated by adding up the Y-values below the line for each row.

In addition to visualising the pattern and area of contact for the target stops as well as the differences between them for the different voice types in a figure like 6.9 (b), we may also infer information about the shape of the tongue during the production of the stop. Assuming that the part of the tongue that lies directly below any part of the hard palate is the one that makes contact with the part of the palate directly above it, several inferences could be made. If a constriction extends over several rows for a stop, it could indicate either longer constriction length (i.e. spatial aspects of tongue contact), or, where the same pattern of on electrodes is observed over several successive frames, longer temporal duration. Different voice types have been observed to have different constriction lengths and temporal durations: with the voiceless stops tending to manifest relatively longer constriction and duration than the voiced stops (Shin (1997) for Korean stops and affricates). It can therefore be inferred that for the voiceless stops, the tongue remains raised towards the roof of the mouth for a relatively longer period.

Place of articulation may also contribute to the configuration of the tongue. In stops produced at the palatal place where the centre of the tongue articulates with the hard palate, the tongue configuration may be inferred to be more convex than stops produced at other places of articulation.

6.4.3.2 Maximum contact (MC).

The MC pattern of the stops to be examined here was done with the articulatory closure and release defined as follows. The definition of onset of closure (C) was taken to be the first frame showing full contact in any row, and the release of closure (R) as the first frame not to have full contact in any row. Figure 6.10 illustrates MC for the segment [t] in the context /ata/ in the word [matapo:le] 'potatoes.' The stop closure occurs between frames 45 and 82, and MC at the frame or frames with the most number of on-electrodes. As mentioned above, in this study, the data was sampled at a rate of 300 frames per second, and the interval between the frames was approximately 3.3 ms.
6.5 Results

6.5.1 Spatial characteristic patterns of stop articulation: voice types and place of articulation

6.5.1.1 Voice types: (a) Number of on-electrodes at MC. Following the conventional method, the number of on-electrodes at MC was counted and averaged across repetitions and vowel contexts for the different voice types of the stops and by place of articulation. The results are plotted on Figure 6.11.
Mean number of on-electrodes for Shekgalagari voice types for stops

It can be seen from figure 6.11 that the contact area is different for the different voice types, with the voiceless aspirated stops showing the greatest area of contact for all the three places of articulation; the unaspirated stops also showing area of contact which is slightly greater than that of the voiced stops. Also, on average, the velar stops shows the least area of contact followed by dental stops and lastly the palatal stops, which have a greater area of contact than the stops produced...
at other places of articulation. The average numbers of on-electrodes at MC across the voice types by place of articulation are: palatal = 52.3, dental = 37.3 and velar 23.2.

6.5.1.1 (b) Voice types: lingual-palatal contact profile — percentage of on-electrodes. Percentages of on-electrodes for the voice types were calculated across vowel contexts in a straight-forward way as explained earlier (see section 6.4.3.1, formula (1)), and a tongue contact profile graph was plotted on Figure 6.12.
Contact area for Shekgalagari voice types for stops: percentage of on-electrodes at MC

Figure 6.12: Tongue contact profile: Average percentages of on-electrodes at MC for the stops across vowel contexts, by voice types and places of articulation.
As was pointed out earlier, this graph helps us to see how the voice types vary in contact pattern, and thus add a dimension which cannot be obtained from counting on-electrodes at MC only (cf. Figure 6.11). The three types of stops are produced with different overall contact pattern as can be seen in the figure. According to figure 6.12, and at least for the subject of this study, [t] and [d] show most contact in rows 1 and 2, suggesting that lingual palatal contact for these stops is made with the blade/tip of the tongue articulating with the region of the palate around the alveolar ridge. The results in this study help only this far. But as was noted in Chapter Two, there has been a discrepancy in the past with respect to place of articulation for these stops in Shekgalagari, with some authors describing them as alveolar stops whilst others thought they were dental stops. Therefore, in addition to the information provided by EPG, and since the author of this thesis was the subject of the study, she also noted that for most of the data containing these stops, the anterior part of the tongue was felt to be constricted against the buccal surface of the front teeth. This appears to suggest that these stops are more likely to be dental stops rather than alveolar, and thus agree with Dickens (1986, 1987) and Andersson and Janson (1997).

As will be noted for the other articulatory types, the voiceless aspirated stop [tʰ] shows higher percentage of on-electrodes than the other two, and there is a weak trend for the voiceless counterpart [t] to show slightly higher percentage of on-electrodes than its voiced cognate [d]. A similar situation was observed for Japanese (Shimizu 1990:52, see § 6.2.1). Both the aspirated and unaspirated voiceless stops recorded the highest percentage of on-electrodes (i.e. 100%) on rows 1 and 2. However, the aspirated cognates recorded a percentage of over 90 on the third row, whilst their unaspirated counterparts recorded a percentage which was below 80. On average, the highest number of on-electrodes (i.e. maximum contact) extended over 11 to 12 frames for the aspirated stops, 8 to 9 for the unaspirated stops and 6 to 7 for the voiced stops.

Compared to the other stops in this study which had slightly more number of frames for MC, this appears to suggest a relatively short constriction duration for dental stops. This may be expected since the front part of the tongue is relatively faster than other parts of the tongue. Also, for these stops, and at least in this study, the percentage of on-electrodes starts to rise again from the fifth row. This may be attributed to the influence of vowels on the back part of the tongue, which appears to be more sensitive to vowel context, particularly back vowels.

The palatal stops [cʰ], [c] and [j] show the highest percentage of on-electrodes with full contact being made on row 7, suggesting more tongue dorsum contact with the palate. Although the highest percentage of on-electrodes is
especially on row 7 for the palatal stops, it can be seen from figure 6.12 that the rise in the percentage of on-electrodes begins at around row 3. This appears to suggest that, in terms of place of articulation, the palatal stops have a bigger constriction, being produced with contact made from around the post-alveolar region to the most back part of the hard palate. Also, taking into consideration the fact that they are made with mostly the middle part of the tongue/tongue dorsum, it possibly takes a relatively longer time to build the constriction. Thus the palatal stops appear to have long constriction, both in terms of front-to-back place contact and of time. In is interesting to note that MC for these stops on average tended to extend over a number of frames in this way: the aspirated stops: around 15 to 16 frames; unaspirated: around 13 to 14 and the voiced stops around 11 to 12 frames; thereby appearing to confirm the long stricture and time factors just mentioned. As will be observed below, this would suggest that the tongue remains raised in the oral cavity for a relatively longer period, leading to a relatively constricted oral cavity. In terms of airflow, a constricted oral cavity means that transglottal air pressure takes a little while to develop and sustain vocal fold vibration. This in turn leads to delayed onset of voicing for the next vowel. This seems to support the observation in the previous chapter (see § 5.7.2) where longer VOT values were obtained for palatal stops than for stops produced at other places of articulation.

The highest percentage of on-electrodes for the velar stops is in row 8, suggesting that it is the tongue dorsum that articulates with the palate, and the length of constriction does not appear to be as long as that of the palatal stops, both in terms of place and time. The voiceless aspirated stop [kʰ] marginally shows higher percentage of on-electrodes than the unaspirated cognate, and the voiced stop [ɡ] shows the lowest percentage of on-electrodes compared with the other two. The average number of frames over which MC extended for the velar stops was as follows: frames for the aspirated stops were on the order of 12 – 13, the unaspirated; 9 – 10 and the voiced stops; 6 – 7. Interestingly, the average number of frames over which MC extends for the velar stops in this study is very much similar to those for the dental stops, inspite of the dental stops having larger contact area than the velar stops. In fact, for the voiceless stops, the velar stops appear to marginally have more frames than the dental stops. It will be proposed below that this could be related to the relative speeds of the different parts of the tongue, where velar articulations have relatively slower articulatory speed and

---

3 It should be remembered that the artificial palate can only go as far back in the mouth as the junction between the soft and the hard palate, and it is possible that part of the articulation for the velar stops may be missed out.
therefore maintain tongue-velar contact for a relatively longer time than the dental stops.

We also may infer information regarding the shape of the tongue during the production of the stops by observing figure 6.12. The difference in contact pattern between the stop types implies that the tongue has a slightly higher position for the voiceless stops than for the voiced ones across all the three places of articulation, since the voiceless counterparts consistently showed higher percentage of on-electrodes, albeit marginally.

With respect to place of articulation, it can be seen from Figure 6.12 that the pattern of contact for the stops appears to be different for different places of articulation. The palatal place shows the greatest area of contact (with full contact being formed in row 7) which appears to be maintained for a relatively longer time. The velar place shows the least lingual velar contact area, shown only on one row, row 8. We may deduce that for the palatal and velar stops, the tongue seems to be in a convex shape and remains so for a relatively longer period for palatal stops than for velar stops. According to figure 6.12, it appears that the tongue shape is concave for the dental stops in this study. For these, the main constriction extends over only one front row (row 1) and remains for the most part on row two and is dramatically lost afterwards. It was mentioned earlier that the configuration of the back part of the tongue appeared to be more sensitive to the influence of back vowels, and not necessarily as part of the production of the relevant stops.

Articulatory dynamics of stops. As mentioned in the introductory chapter (see also Appendix 1A), EPG allows for the examination of contact patterns between the tongue and the hard palate as well as recording the timing of these contacts in speech. Both contact pattern for the stops and changes in these patterns over time may therefore be studied. In this subsection, we analyse changes in contact pattern at MC over time for Shekgalagari stops, focusing on differences due to voice types and place of articulation. This may clearly be portrayed by means of a graph of percentage of on-electrodes vs. time. This graph was plotted in the following way.

Following the conventional method, the number of on-electrodes from C (contact) to R (release) was counted per frame per token and across all the

---

4 But, as we noted earlier, although velar stops can be investigated, this investigation may be partial, and this must be taken into consideration in the analysis of results of stops produced at the velar place. For this reason, our results for the velar place here may therefore be regarded as incomplete.
repetitions of the tokens. MC points for the tokens were identified and used as the point of reference to time align individual frames for the tokens, following Shin (1997:74,90). Where MC extended over more than a single frame, the middle frame was used if the number of MC frames resulted in an even number. If, however, the number of MC frames observed resulted in an odd number, the reference point was determined in the following way:

\[
\left\{ \left( \frac{\text{the total number of MC frames}}{2} \right) + 0.5^{\text{th}} \text{ frame} \right\}
\]

Every repetition of the tokens was then time aligned with the point of reference. In order to plot the graph, frames to the left of reference MC frame, which represent build up toward the major constriction of the stop, MC, were assigned negative values. Frames to the right of MC, representing the release of the major constriction, were assigned positive values. The point of reference itself was assigned the value 0 (cf. Shin 1997).

After counting the number of on-electrodes from C to R, the percentage of on electrodes was then determined using formula (1), and then averaged across vowel contexts, so that each point on the graph represented an average of the percentage of on-electrodes for 45 tokens (being 9 different vowel contexts \(X^5\) 5 repetitions of each token). This was done for all the three places of articulation and the voice types. The graphs are presented on figures 6.13 through 6.15.

\[X\] means multiply by.
Fig. 6.13: The percentage of on-electrodes vs. time for the dental stops
Fig. 6.14: The percentage of on-electrodes vs. time for the palatal stops
Fig. 6.15 The percentage of on-electrodes vs. time for the velar stops
The main concern here is to observe the changes in contact pattern over time for Shekgalagari stops, focusing on differences due to voice types and place of articulation. Regarding voice type, the graphs appear to suggest that, overall, at all levels; i.e. build-up to MC, at MC and during the release phase, the percentage of on-electrodes is largest for the voiceless aspirated stops, followed by the voiceless unaspirated stops and lastly the voiced stops within an articulatory place and for all the three places of articulation.

The stops are also articulated with an increasing number of on-electrodes, with more time being required during the build-up to maximum contact (MC) than during the release phase. However, of the three places of articulation, the dental stops appear to reach maximum contact and release the closure faster than the other stops. This could be attributed to the fact that the tongue tip/blade have been observed to move at a faster speed than the other parts of the tongue.

For the voice types within articulatory places and for all the places of articulation, the percentage of on-electrodes increases considerably at maximum contact, and decreases dramatically after MC. Generally, the ranking order of the factors at MC is as follows: aspirated > unaspirated > voiced - (this seems to correlate with observations made earlier (cf. figure 6.11), where it was observed that the voiceless aspirated stops showed the largest area of contact and the unaspirated stops showing area of contact which is slightly greater than that of the voiced stops). With respect to place of articulation: palatals > velars > dentals. In figure 6.11, on average, the velar stops showed the least area of contact followed by dental stops and lastly the palatal stops.

In figure 6.11, we observed the least area of contact for the velar stops, but here we observe that velar stops nevertheless appear to require more time for build-up to MC and for the release phase than the dental stops. This could be an indication that large contact area does not necessarily mean more time required for the articulation. As will be pointed out for the palatal stops in the next paragraph, this may be accounted for in terms of the relative speeds of the different parts of the tongue. Also, for the velar articulation, as already pointed out, (cf. App 1A) the limitations of the technique mean that the sounds can only be partially investigated.

The release phase for the palatal also seems to be relatively slower than that of the velar stops, albeit only marginally. We may attribute this to the fact that the part of the tongue that articulates with the palatal and post palatal region of the hard palate moves at a relatively slower speed than the front-most part of the tongue. This could mean that tongue-palate contact in posterior articulation is maintained for relatively longer period than in anterior articulations.
It is not possible to determine the influence of different vowel contexts here because data was pooled across vowel contexts. This could be done in future research.

6.5.2 Characteristics of co-articulation: spatial changes

This section investigates the effect of flanking vowels on the production of stops, focusing on the spatial aspect, and Shin’s (1997) extension of Farnetani’s percentage method of plotting a graph visualising this effect is again used. The percentages of on-electrodes were calculated for each context per voice type and for each place of articulation for the stops. The results are shown on Figures 6.16 through 6.18.
Fig. 6.16(a): The influence of flanking vowels on the production of the voiceless aspirated stop [th]
Fig. 6.16(b): The influence of flanking vowels on the production of the voiceless unaspirated dental stop [t]
Figure 6.16(c). The influence of flanking vowels on the production of the voiced dental stop [d].
Fig. 6.17(a): The influence of flanking vowels on the production of the voiceless aspirated palatal stop [ch]
Fig. 6.17(b): The influence of flanking vowels on the production of the voiceless unaspirated palatal stop [c]
Fig. p. 17(c): The influence of flanking vowels on the production of the voiced palatal stop [f].
Fig. 6.18(a): The influence of flanking vowels on the production of the voiceless aspirated velar stop [kh]
Fig. 6.18(b): The influence of flanking vowels on the production of the voiceless unaspirated velar stop [k]
Fig. 6.18(c): The influence of flanking vowels on the production of the voiced velar stop [g]
The focus of this section is to investigate the effect of flanking vowels on the production of stops and to compare this effect between the different voice types as well as places of articulation. Susceptibility to co-articulation may be determined from the graphs by observing the regularity of the traces and how separated from each other they are. The results obtained show that the stops show some susceptibility to co-articulation, both by voice type and by place of articulation. The degree of this susceptibility, however, appears to be different for the voice types and according to place of articulation.

With regard to place of articulation, the palatal stops appear to show less accommodation to the influence of flanking vowels, with the voiceless aspirated stops showing less influence, followed by the voiceless unaspirated stops. The voiced cognates appear to be relatively more susceptible to co-articulation (see figure 6.17 (a ~ (c)). It may be noted that the traces on figure 6.17 (a) ~ (c) appear to be closer together than the lines on figures 6.16 (a) ~ (c) and 6.18 (a) ~ (c), and that, within this articulatory type, the lines on figure 6.17 (a) appear to be closer together than the lines on figure 6.17 (b), which also in turn seem to be closer than the ones on figure 6.17 (c). It may be remembered that it was observed earlier, (see figure 6.11), that palatal stops seem to have a contact area which is larger than that observed for stops produced at other places of articulation. It will be argued in section 6.6 that observations made for the palatal stops (as well as for the other articulatory places) seem to support the co-articulation resistance theory, which proposes that the contact area of a segment and its susceptibility to co-articulation have an inverse relationship with each other.

It may also be worth pointing out that, as will be observed for the other articulatory types, the graphs seem to suggest that both the front and back parts of the tongue are susceptible to co-articulation, with the back part accommodating more to contextual influence than the front part (compare the behaviour of the tongue on row 1 with that on row 8 on figure 6.17 (a) ~ (c)). As mentioned earlier, the lines on the graphs are rather revealing with regard to the shape of the tongue. It may be observed from figure 6.17 (a) ~ (c) that, for all the different vowel contexts, the percentage of on-electrodes begins to rise from row 3 of the artificial palate and onwards, giving 100% contact on row 7 (and, for the voiceless counterparts, and for some contexts, on row 8 too). This gives the general impression of a very tense, rather convex tongue configuration. It will be argued in section 6.6 that, for the palatal stops in particular, this observation appears to suggest that, in addition to the inverse relationship between the contact area of a segment and its susceptibility to co-articulation as the co-articulation resistance theory proposes, tongue configuration may be another dimension worth
considering when studying the co-articulatory behaviour of segments. Russian palatal consonants will also be cited in support of this.

The vowel /i/ also seems to be exerting more influence than the other vowels, irrespective of whether it occurs in a pre-consonantal or post-consonantal position. Its greatest influence is, however, observed when it preceded and follows the stop, i.e. in the /i _ i/ context for both the voiceless and the voiced stops.

The dental stops come second to the palatals in the magnitude of contextual influence that they exhibit. For this place of articulation also, the voiceless cognates show less influence than their voiced counterparts (see figure 6.16 (a), (b) vs. (c)). It may be observed that the lines on figure 6.16 (a) ~ (c) appear to be closer together than the lines on figure 6.18 (a) ~ (c), but relatively more spread than those on figures 6.17 (a) ~ (c). Also, within this articulatory type, the lines on figure 6.16 (a) appear to be closer together than the lines on figure 6.16 (b), and those on figure 6.16 (b) closer than those on figure 6.16(c). It may be remembered that it was observed earlier, (see figure 6.11), that the dental stops showed contact area which is smaller than that observed for the palatal stop but larger than that observed for stops produced at the velar place of articulation. It will be argued in section 6.6 this observation also seems to support the co-articulation resistance theory.

The graphs also appear to suggest that the front and back parts of the tongue have different degrees of susceptibility to co-articulation. The back part seems to accommodate more to contextual influence than the front part (compare the behaviour of the tongue on row 1 with that on row 8 on figure 6.16 (a) ~ (c)). As mentioned earlier, the shape of the tongue may be appreciated by observing the lines on the graphs. It may be observed from figure 6.16 (a) ~ (c)) that, for all the different vowel contexts, the percentage of on-electrodes is highest on row 1 (rows 2 and 3), giving 100% contact for the voiceless counterparts for all vowel contexts, and then it dramatically falls to between 35% and 52% mostly from row 5 to 8. This seems to suggest a generally concave tongue configuration. It will be mentioned in section 6.6 that observations made for the dental stops (as well as the velar stops) appear to support the co-articulation resistance theory more on tongue contact area than on tongue configuration.

Again the vowel /i/ also seems to be exerting more influence than the other vowels, both in pre-consonantal or post-consonantal positions, although its greatest influence appears to be observed in the symmetrical /i _ i/ context for both the voiceless and the voiced stops within the dental articulatory place.
Out of the three places of articulation, the velar stops seem to be more accommodating to contextual influence than the other two articulatory types. It may be observed that the lines on figure 6.18 (a) ~ (c) appear to be relatively more separated than the lines on figures 6.16 (a) ~ (c) and those on figures 6.17 (a) ~ (c). Also, within this articulatory type, the lines on figure 6.18 (a) appear to be closer together than the lines on figure 6.18 (b), and the lines on figure 6.18 (b) closer than those on figure 6.18 (c). This indicates that, for this place of articulation, the voiceless cognates show less influence than their voiced counterparts. It was observed earlier (see figure 6.11), that the velar stops showed a contact area which is smaller that that observed for the palatal stops and the dental stops. This observation also seems to support the co-articulation resistance theory.

The front and back parts of the tongue have different magnitudes of susceptibility to co-articulation. Contact for the back part of the tongue seems to change more as a function of flanking vowels compared to the front part of the tongue (compare row 1 with that on row 8 on figures 6.18 (a) ~ (c)). With regard to the shape of the tongue, for all the different vowel contexts, the percentage of on-electrodes is highest on row 8, giving percentages that are above 90, but it is considerably lower for rows 1 through 6. This seems to suggest a generally convex tongue shape. As with the dental stops, it will be mentioned in section 6.6 that observations made for the velar stops appear to support the co-articulation resistance theory more on tongue contact area than on tongue configuration. However, it should be kept in mind that the study of contact pattern for the velar stops by means of EPG may be hampered by the limitation of the technique. The artificial palate does not go beyond the area where the soft palate and hard palate join, otherwise it introduces discomfort and interferes with natural speech. This means that some of the velar articulations may therefore not be detected.

Again the vowel /i/ also seems to be exerting more influence than the other vowels, both before and after the stop, although its greatest influence appears to be observed in the symmetrical /i _ i/ context for both the voiceless and the voiced stops at the velar articulatory place.

### 6.6 Discussion of results

This subsection discusses the implications of the results obtained in this study for the theories of co-articulation presented in Part One of this thesis, with the
intention of finding which model can best account for the results obtained here.\footnote{Only the feature spreading model, the window model of Keating and the co-articulatory resistance theory will be evaluated here. Other aspects needed to fairly evaluate the other models have not been investigated in this study. For example, the assessment of the co-production model requires the measurement of temporal aspects of the intervening stop: closure duration, time distance between $V_1$ and the stop, the stop and $V_2$ and $V_1$ and $V_2$ in order to determine whether co-articulation is more sensitive to the time distance between segments irrespective of context. These have not been measured.} To assess the effectiveness of the models, their main proposals are first summarised.

6.6.1 The feature-spreading model

As pointed out in Part One, according to the feature-spreading model, co-articulation is heavily context dependent, and occurs because a feature spreads from one segment to the neighbouring one. For example Henke (1966), after studying anticipatory labial co-articulation in Russian, proposed that speech segments have articulatory targets and may be assigned binary phonological features $[\pm]$. Segments without articulatory goals may be considered to be neutral and hence assigned the value 0. In any given domain of co-articulation, all preceding unspecified segments are then given a feature of a specified (following) segment in a "look ahead" fashion. In the event that unspecified segment(s) also follow the specified one, the feature of a specified one spreads onto those segments too, although this carry-over effect is accounted for in physical terms, as resulting from "muscle delay" (Henke 1966:47). The spread of a feature may only be blocked by a segment with its own inherent articulatory specification.

Daniloff and Moll (1968) and Moll and Daniloff (1971) studied co-articulation movements for anticipatory lip protrusion and velum lowering in American English. They observed that co-articulation movements for anticipatory lip protrusion could actually start approximately two segments prior to the segment specified $[+\text{round}]$, (thus the feature $[+\text{round}]$ actually spreads onto the unspecified two segments), and that velum lowering may be triggered at the start of the first vowel in CVVN sequences (thus the feature $[+\text{nasal}]$ spreads from the nasal segment ahead onto the two preceding vowels). According to this model, co-articulation happens simply when a feature spreads from one segment onto another as these experiments illustrate.

In the present study, segments have rather exhibited differences in the degree of co-articulation rather than the presence or absence of a feature, both within place of articulation and by voice type within articulatory type. Take, for
instance, place of articulation. Stops of the same articulatory type have exactly the
same tongue body features for place, and differ with regard to laryngeal features
only. Dental stops have the feature [+ dental], palatal stops have the feature [+ palatal] and velar stops the feature [+ velar]. Yet stops produced at the same place
of articulation manifest different degrees of co-articulation. Since these stops have
exactly the same feature specification for place, nothing is left for the feature-
spreading model to claim has spread onto them to account for the differing
degrees of co-articulation that the stops exhibit.

It was also observed that different places of articulation manifest co-
articulation to a varying extent in this order: palatal < dental < velar. It is difficult
to see how the feature-spreading model may explain this variation in terms of the
spread of a feature, since this will have to occur across places of articulation. As
mentioned above, this variation rather appears to be dependent on tongue contact
area (see figure 6.11), and explainable in scalar terms rather than in terms of the
presence or absence of binary phonological features.

With regard to voice type within an articulatory type, there has been a
weak trend of the voiceless stops to show consistently less accommodation to the
influence of flanking vowels than their voiced counterparts. This was shown by
the fact the lines of the graphs for the voiceless stops appeared to be closer
together and smoother than those for the voiced stops. Again co-articulation
between the voice types seems to happen in degree rather than in terms of the
presence or absence of a feature. There is, therefore, an incompatibility between
the co-articulatory behaviour of the stops in this study and the feature-spreading
account of co-articulation. The model does not fit the results and is not suitable
for explaining them.

It was also observed that the high vowel /i/ consistently exerted more
influence than the other vowels, both before and after the stop, with its greatest
influence observed in the symmetrical /i _ i/ context for both the voiceless and the
voiced stops and for all the three places of articulation. Although this vowel may
be specified with the feature [+ high], it is still difficult to see how this feature can
usefully explain the co-articulatory behaviour of the stops.

6.6.2 The window width theory

As presented in Part One, the window model, proposed by Keating in a series of
papers (e.g. 1990a) accounts for co-articulation in terms of windows. Windows
have a temporal length and a spatial width, where the widths of windows
represent the maximum and minimum range of contextual variation of a segment.
A segment which shows more sensitivity to context may be assigned a wider window width than a segment which does not accommodate much to influence from neighbouring sounds.

This model appears to be compatible with the results obtained in this study, both for articulatory types and voice types, in that segments may be assigned different window widths depending on their co-articulatory behaviour. Palatal stops were less susceptible to co-articulation than stops produced at other places of articulation. Therefore, out of the three articulatory types, they may be assigned the narrowest window width. The velar stops showed more sensitivity to context, and may therefore be assigned the widest window width. The dental stops were somewhat in between, and may be assigned a medium sized window width. Similarly, with regard to voice type within place of articulation, the voiceless stops may be assigned a narrower window width than their voiced cognates since they manifested less susceptibility to co-articulation than the voiced stops, albeit marginally. Consider figure 6.19.
Figure 6.19: Window widths for Shekgalagari voiced and voiceless unaspirated and voiceless aspirated stops. \( V_1 \) and \( V_2 \) represent the same vowel, e.g. [a], hence the same window widths (on the sides of the pictures above the \( V_1, V_2 \) symbols) for these across places of articulation. The brown line represents contour lines connecting windows. The black dotted lines represent window widths for the aspirated stops, the blue unbroken lines represents width for the unaspirated stops and the green broken lines represent the voiced stops.
As pointed out in Part One, this model accommodates phonetic variation in individual languages. Window widths may therefore differ from one language to another, and/or from dialect to dialect. In addition, phonetic rules of individual languages and their phonemic inventories may also affect widths of windows for segments. But, no criterion is provided by which window widths may be determined. Thus with no constraint given by the model, exhaustive investigation of the co-articulatory behaviour of segments is left as the only deciding factor. And, given that the widths of windows represent the maximum and minimum range of contextual variation of a segment, it is rather difficult to determine the exact size of a window width for a segment in a language, and comparison with other languages becomes rather unrevealing. The best that could be done in studying the co-articulatory behaviour of segments in a language is to say, for instance, that in Shekgalagari, palatal stops have the narrowest window width, followed by dental stops and lastly velar stops, and that the voiceless stops have a slightly narrower width than their voiced cognates within places of articulation in the language.

Also, in this model, adjacent windows are linked to each other by means of contour lines. It was discussed in Chapter Three that contour lines give information about the articulatory (and/or acoustic) variation of a segment over time within a particular context. These contour lines are governed by 'smoothness and minimal articulatory effort.' To give examples, in a sequence like VCV, an intervening window for the C (stop) with a wide width (indicating that it is unspecified for tongue body position) allows for direct interaction between the preceding and the following (vowel) segments. Thus vowel-to-vowel co-articulation occurs with very minimal intervention from the intervening stop. If, however, the intervening consonant is specified for tongue body position, then its associated window width will be narrow, and will discourage quick interaction between flanking vowels.

Also, 'windows have no internal temporal structure allowing them to stretch or compress in time,' (Farnetani 1997:393). The widths are rather limited to representing all possible contextual variations only. At any rate, the model would become rather weak if window widths associated with phonological features could also stretch or compress due to factors other than the phonological features with which they are associated.
6.6.3 Co-articulation resistance theory

To recap, the co-articulation resistance theory relates to the co-articulatory resistance behaviour of segments and is based on two different assumptions. The first one states that segments may be assigned different co-articulatory resistance values, depending on how they resist contextual influence (Bladon & Al-Bamerni (1976). This is based on the observation that the co-articulatory behaviour of any given segment may differ from language to language and/or dialect to dialect. For example, Bladon & Al-Bamerni (1976) studied the co-articulatory behaviour of the allophones of the phoneme /l/ in English and noted that these allophones co-articulated to different degrees with the surroundings in a VCV environment, with clear [l] showing the least co-articulation resistance, syllabic dark [ʎ] showing the most resistance to co-articulation and dark [t] somewhat in the middle. This graded nature of co-articulation resistance for these allophones was also observed in relation to co-articulated voicelessness. They then proposed that these allophones therefore be given different co-articulatory resistance values. Thus syllabic [ʎ] could be assigned a high co-articulatory resistance value, followed by [t] and lastly [l].

This theory appears to be compatible with the results obtained in this study, in that palatal stops may be assigned high co-articulatory resistance value, followed by dental stops and lastly velar stops with the least resistance value. And, similarly, slightly higher values for the voiceless cognates than the voiced cognates within an articulatory type.

If high resistance value is equated to narrow window width, low resistance value to wide window width and medium sized width to average resistance value, the co-articulatory resistance approach becomes very much similar to the window theory of Keating. Also, resistance values may change from language to language and/or dialect to dialect; and again, there is no constraint provided against which the assignment of values may be done.

The second hypothesis states that the lingual palatal contact area of a segment may be inversely related to the degree of co-articulation it exhibits (Recasens, e.g. 1987, 1990, 1991). Segments with large contact area are less susceptible to co-articulation, and segments with small contact area are more susceptible to co-articulation. In other words, the co-articulatory resistance behaviour of a segment in any language of the world may be predicted from its articulatory characteristics. This hypothesis thus introduces a constraint in the form of phonetic properties of a segment. Also, it appears to suggest universality
of this aspect, and does not need to be specified in the grammar of any individual language.

The results obtained in this study seem to support this hypothesis. With regard to articulatory type, the palatal stops, which showed the largest contact area compared with stops produced at other places of articulation, also showed the most resistance to contextual influence. The velar stops showed the smallest contact area and appeared to manifest less resistance to the influence of context. The dental stops came somewhere in the middle.

It is worth pointing out that in some investigations, the overall tongue configuration has also been mentioned as a possible contributing factor in studying the co-articulatory behaviour of segments. For example, in Russian, both palatal and non-palatalised consonants are not produced with a large contact area (Bolla 1982), yet they both appear to manifest high resistance to co-articulation. Both, however, are produced by raising the back of the tongue towards the roof of the mouth since, “Russian non-palatalised consonants are often, in fact, strongly velarized and they seem to influence the quality of neighbouring segments” (Shin 1997:339). This gives the general impression of a very tense tongue shape. Such a configuration would not readily accommodate influence from flanking sounds.

In this study, it was observed that the percentage of on-electrodes for the palatal stops began to rise from row 3 of the EPG palate, with full contact being made on row 7. This gives the general impression of a very constrained, rather convex tongue configuration. For the palatal stops in particular, this observation appears to suggest that in addition to the contact area of a segment and its susceptibility to co-articulation tongue configuration may be another dimension worth of consideration when studying the co-articulatory behaviour of Shekgalagari palatal stops, needless to say that, of course, physical studies still have to be done to examine tongue configuration during their production.

6.7 Summary

The aim of this chapter was to examine articulatory characteristics of Shekgalagari stops by means of electropalatography (EPG), providing qualitative and quantitative description of the type and extent of articulatory movements for Shekgalagari stops and for the variation observable between the voice types. It was observed that contact area is different for the different voice types, with the voiceless unaspirated stops showing area of contact which is slightly greater than that of the voiced stops. With regard to place of articulation, the velar stops
Chapter 6. Articulatory analysis: Electropalatography

showed the least area of contact followed by dental stops and lastly the palatal. The difference in contact pattern between the stop types implies that the tongue has a slightly higher position for the voiceless stops than for the voiced ones, since the voiceless counterparts consistently showed higher percentage of on-electrodes, albeit marginally.

With respect to place of articulation, it can be seen from Figure 6.12 that the pattern of contact for the stops appears to be different for different places of articulation, with the palatal place showing the greatest area of contact, and is followed by the dental place. The velar place shows the least lingual palatal contact. It may be deduced that for the palatal and velar stops, the tongue seems to be in a convex shape and remains so for a relatively longer period for palatal stops than for velar stops. It appears that the tongue shape is concave for the dental stops in this study, although the back part of the tongue may be raised due to the influence of back vowels, and not necessarily as part of the production of the relevant stops.

Coarticulation was also investigated by analysing lingual-palatal contact patterns for V-to-C co-articulation in symmetrical and asymmetrical vowel contexts. The stops showed some susceptibility to co-articulation, both by voice type and by place of articulation. The degree of this susceptibility was different for the voice types and according to place of articulation. The palatal stops appear to show most resistance to co-articulation, followed by dental stops and lastly by velar stops. The back part of the tongue also seemed to accommodate to more contextual influence than the front part. The vowel /i/ also seemed to be exerting more influence than the other vowels, particularly in the /i _ i/ context for both the voiceless and the voiced stops within an articulatory place.

The findings of the experiment in this study were discussed in the light of constraints on the tongue during the production of sounds, theories proposed to account for co-articulation in speech and in the light of the findings reported in the literature for similar studies in other languages.
Part Three: Implications for phonetics and phonology

(Representation of Shekgalagari Stops)
7 Some implications of the acoustic investigation for general phonetics and phonology

7.1 Introduction

This chapter attempts, on a more general level, to addresses the contribution of the findings of the experiments performed in Part Two of this thesis to linguistic theory, and, on a specific level, to the theories presented in Part One, section 3.3. It starts by addressing the rather controversial relationship between phonetics and phonology (§ 7.2), a discussion of the phonological features is given in section 7.3, and issues of stop representation are also discussed (§ 7.4), in order to address the issue of the phonetics and phonology of Shekgalagari stops, (§ 7.5), based on the theories presented in Chapter Three, Part One, and the experiments performed in Part Two of this thesis. Cross language analysis is also made in section 7.6. Section 7.7 gives a summary of the chapter.

7.2 The phonetics-phonology correlation

Phonetics, on the one hand, may be described as a discipline concerned with the study of the concrete physical properties of speech sounds. In other words, phonetics deals with finding out and describing the vocal sounds utilizable in human speech, and studying their articulatory, acoustic and perceptual properties. Phonology, on the other hand, may be described as a discipline concerned with the more abstract aspect of speech sounds: that is, with how the sounds are used in a language, how and why they pattern together or function, and how they are represented and used in the mental grammar of the speaker. With these descriptions of phonetics and phonology, it may appear reasonable to conclude that these two disciplines are indisputably related to each other. But the fact that the relationship between phonetics and phonology has been one of the areas of controversy amongst linguists for a long time and still largely remains so would seem to indicate that such conclusion may not be so easily reached.

The crux of the arguments regarding the connection between phonetics and phonology has been their relevance to linguistics. Opinions have been varied. Some researchers have argued that phonetics, unlike phonology, does not belong to linguistics and is not governed by linguistic principles. Proponents of this view argue that properties of (speech) sounds derived from the main branches of
phonetics — articulation, acoustics and perception — are a result of human physiology (and anatomy) and the aerodynamics involved, and are not necessarily relevant to linguistics. More recently, for instance, Harrison (1999:198-199) argues that articulatory phonetics is 'in part straightforwardly a branch of physiology (and anatomy) (1999:199). As far as he is concerned, parameters described in articulatory phonetics do not deliver any linguistic information. He goes on to argue that acoustic phonetics is a part of a branch of physics which investigates sound, and adds only 'when that sound is made by the human vocal tract and used in the service of language' (1999:198) (italics original). He insists that acoustic data cannot be restricted to linguistically significant speech only. Any sound type: e.g. speech, music, noise, may also be investigated. Also, according to Harrison, in auditory phonetics, 'the field of enquiry is once again squarely set within either a physiology, an anatomy or a physics framework' (1999:199). He concludes by saying that 'The three branches of phonetic science, then, do not involve themselves with the linguistic' (Ibid.).

Other views which have been expressed will not be discussed in detail here but will only be mentioned briefly. Some of these opinions include those of researchers who see phonetics and phonology as inter-connected disciplines, neither of which belong to linguistics. According to this view, phonology is only regarded as a link between phonetics and linguistics, and is not genuinely linguistic. Also, according to some of the definitions given by some researchers, phonetics and phonology may well be one and the same thing. Consider, for example, this one given by Crystal (1991:259); 'PHONETIC FORM: the output of the phonological component of the grammar, or the phonological component itself.' Some other opinions commit to the view that phonetics and phonology are inter-connected disciplines within linguistic theory, although the precise nature of interface between the two, and where it occurs remains a matter of controversy.

In fact, nowadays there appears to be some general agreement that phonetics and phonology are related to (but also differ from) each other, may and should be studied in conjunction with each other, and that both belong within linguistic theory. The following section focuses on the theory of features, and briefly explores just how the study of features has helped to understand and

---

1 In the past decades, there have been rather extreme versions of this view, where both phonetics and phonology were regarded as completely independent of each other, each capable of being pursued without reference to the other (some of the proponents of such views are mentioned in Jessen 1997: 21-22).

2 For a detailed discussion of these, the reader is referred to the references in Harrison (1999:198).
possibly acknowledge the connection between phonetics and phonology, and perhaps therefore the relevance of both disciplines to linguistics.

7.3 Features

7.3.1 Features and the phonetics and phonology connection

One of the interesting aspects of the study of speech sounds has been the observation that they tend to pattern together in languages. Researchers have been curious to find out how this patterning happens and why. One of the ways of finding just this has been by way of features, where features were thought to be linguistically significant properties of speech sounds. If we abide by the descriptions of phonetics, and phonology given earlier on (§ 7.2), that phonetics deals with physical properties of speech sounds, and phonology with how the sounds are used in a language: how and why they pattern together or function etc, then it becomes clear that phonetic details of speech sounds have been used in the phonological study of languages. Phonologists viewed segments as being made up of hierarchical features. Thus features acknowledged a link between phonetics and phonology, and hence implied at least some connection between phonetics and general linguistic theory, since phonology has generally been acknowledged to belong to linguistics.

The acknowledgement of the relevance of phonetic realization to phonology (and hence to linguistics) has also brought with it the issue of just to what extent phonetics was relevant to linguistics. Some researchers, e.g. Chomsky and Halle (1968), assumed that all phonetic detail was automatic and language universal. Others, however, thought that there were phonetic details of speech sounds which were subject to variation in individual languages. Keating is one of the proponents of this view. In the (1990b) article, she argues that phonetic details of segments are not always universal. She insists that phonetic realization may differ from one language to the other, being subject to language specific rules. The difference in vowel lengths before stops with different voicing configurations in English is one of the examples she gives. Vowels tend to be longer before voiced stops than before voiceless ones. This observation is not relevant for Czech, a West Slavic language, where the difference in the vowel duration is not consistent and may not even be existent at all. Keating therefore argues that phonetic realization is not always universal, and proposes that particular aspects of phonetic implementation be incorporated in the grammars of individual languages.
Chapter 7: Some implications of the acoustic investigation for general phonetics and phonology

In a way, all this implies some appreciation of the relevance of phonetics to linguistics as well as the fact that phonetics and phonology are correlated: the objects of study in both are speech sounds.

7.3.2 Distinctive features

All the three branches of phonetic science mentioned above: articulatory phonetics, acoustic phonetics and auditory phonetics have a potential of each generating an infinite number of features (articulatory features, acoustic features and perceptual features, i.e. with each branch providing a different view on speech) which may be used as descriptive parameters, that is to describe speech sounds as accurately as possible. Thus there exists the potential for uneconomic over-generation of features which may be redundant in describing sounds. Considering the fact that speech sounds make up linguistic systems, it appears obvious that they also exist in order to contrast with one another within these systems, as well as pattern together and form groups or classes. This is where distinctive features become important. Distinctive features only describe contrast between segments and classes of sounds within (and across) a language(s). The description between segments in a language is mostly based on listeners' categorical discrimination of sounds. Distinctive features could also be used to classify classes of sounds within a linguistic system. But the relationship between these (phonetic) features, and the phonological contrast, however, is not a straightforward, one-to-one relationship, as it constitutes one of the problems that lie in the phonetics-phonology connection.

7.3.2.1 Distinctive feature geometry: the issue of universality

There has been emphasis, particularly in generative phonology, placed on the need for distinctive features to be universally applicable. Feature geometry ought to be able to function both as a universal inventory and as parameters for language specific description. Consider, for instance, the feature [± tense]. At a universal level, it is described by Jacobson and Halle (1956:29 ff) in the following phonetic terms.
Chapter 7: Some implications of the acoustic investigation for general phonetics and phonology

Acoustic: high energy with greater spread across spectrum and longer duration.
Articulatory: greater deformation of vocal tract from its rest position.

Cross linguistic sounds with these particular phonetic characteristics would be described as [+ tense], whilst those without would be [- tense]. At a language specific level, the feature [± tense] would be used to distinguish between two different classes of sounds within that language, e.g. [p, t, k] would be classified as [+ tense] and [b, d, g] as [- tense] in German (cf. Jessen 1997).

Supporting the universal validity of distinctive features, Chomsky and Halle (1968:295-5) argue that 'The total set of features is identical with the set of phonetic properties that can in principle be controlled in speech; they represent the phonetic capabilities of man and, we would assume, are therefore the same for all languages'. This appears to imply that articulatory, acoustic and perceptual (auditory) features could be correlated. In fact, however, many of the terms given to features being used in linguistics could in fact be biased towards one branch of phonetics more than others. For example features like [± coronal], [± anterior], [± continuant] seem to be biased towards articulatory phonetics, whilst features like [± strident] and [± sonorant] appear to be biased towards acoustic or auditory phonetics (features based on Chomsky & Halle 1968). This, nevertheless, seems to support the fact that speech is indeed an integration of the three levels of phonetics. Regarding this Clark & Yallop (1990:317) write:

Thus whilst an acoustic signal must be analysed in its own terms (intensity, frequency, etc.), the criteria by which features and parameters are selected and assigned values must refer to linguistic activity. In short, acoustic features are treated as correlates or realization of other features. The values of formants within an acoustic spectrum are measured not because they are objective properties of acoustic reality but because they are believed to reflect articulatory settings and serve as cues in human perception. A similar point can be made about articulatory features.

7.3.2.2 Distinctive features and their correlates

Because distinctive features are understood to have a phonetic base, they are often defined in phonetic terms, i.e. using phonetic features. Some features have been described in terms of acoustic features (e.g. Jacobson et al., 1952), others mainly in articulatory terms (e.g. Chomsky & Halle 1968), whilst others
have been defined in terms of both acoustic and articulatory features (e.g. Jacobson & Halle (1956), Ladefoged (1982:254)). The features are binary, with plus and minus values. But, as mentioned earlier (§ 7.3.2), mapping phonetic information onto phonological features is not a straightforward, one-to-one relationship. Often, these phonological features are described using several phonetic features or correlates (articulatory and acoustic). Consider the following illustration for the phonological feature [± voice], for a language, say French.

These several phonetic correlates of phonological features have a hierarchy, depending on whether they can express the feature in all or most contexts and conditions in the language. A correlate which expresses the feature in all or most contexts and conditions in the language is a basic correlate of the feature, and others, which are contextually less frequent, are non-basic correlates. Other conditions that might be considered in establishing the basic and non-basic status of the correlate include statistical reliability and the contribution of the feature towards perceptual discrimination of a segment. In addition to basic and non-basic correlates, some researchers have proposed the term concomitant correlates, which they define as those correlates which occur as a result of the basic correlate (see, for instance, Jessen (1997:220)).
7.3.2.3 Distinctive features and their correlates for three-way category languages

In this subsection distinctive features of some of the three-way contrasting languages as well as their correlates is attempted basing on the literature. Only stops occurring in word initial context will be dealt with, although it is acknowledged that, as indicated above, basic and non-basic correlates are determined on the basis of reliable occurrence in all or almost all contexts; statistical reliability and significance in contribution to perceptual distinction may be considered where necessary. This investigation is limited to this position largely because the Shekgalagari stops analysed in this thesis were restricted to word initial position. The languages to be discussed in this subsection are Thai, Burmese and Korean, and phonetic features will include the activity of the vocal folds during stop closure, VOT, F0 onset and contour, and F1 onset.

Thai. Thai is a three-way contrasting Asian language, with stops produced at three places of articulation; viz.: [b, d], [p, t, k] and [pʰ, tʰ, kʰ]. Thai does not have voiced velar stops in word initial position. Also, stops are not released in word final context; the contrast between the stops is therefore neutralized, although there may be vowel-to-consonant transition cues to the place of articulation of the following consonant. Thai is also a tone language, where tone has lexical contrast significance.

Articulatory analysis. With regard to voicing, acoustic waveforms for Thai stops showed voicing lead before release for the [b, d] series, considerable turbulence before the commencement of voicing for the [pʰ, tʰ, kʰ] series, and simultaneous burst release and start of voicing for the [p, t, k] series (cf. Shimizu 1990).

Most of the discussion here will be based on Shimizu (1990), who investigated VOT, F0 onset and contour, F1 onset as well as spectral analysis of these languages. Other investigation will also be referred to.

As pointed out in Chapter Three, section 3.2.2, stops with different voice types are believed to affect the fundamental frequency (F0) of the (initial portion of the) following vowel differently. In many languages (e.g. English (Ohde 1984), Thai (Shimizu 1990)), the F0 contour has often been reported to begin at a high point and falling for the voiceless aspirated stops, whilst for the voiced, it has been thought to begin at a low point and rising (cf. Ibid., Hombert et al. 1979). This rise/fall dichotomy, however, remains controversial. For instance, in some studies (e.g. German (Jessen 1997:76-87)), although higher F0 values were obtained for tense stops than for lax stops, the F0 contour was not necessarily of the rise/fall pattern. Rather, the falling pattern was observed for both of the voicing types. The discussion to be presented here for three-way contrasting languages is based on the results of the studies performed on those languages (Shimizu 1990).
Chapter 7: Some implications of the acoustic investigation for general phonetics and phonology

1990:123-124). Thus the different voice types were manifested by different glottal states in the larynx.

*Acoustic analysis.* Voice onset time: VOT values for the stops were consistently; large and negative for the [b, d] series, large and positive for the [pʰ, tʰ, kʰ] series and small and positive for the [p, t, k] series (Lisker & Abramson, 1964; Shimizu 1990:126-128, 137-138). The difference between the VOT values obtained for the three groups were statistically significant. With respect to F0 perturbation and contour, the results obtained by researchers have rather been inconsistent. F0 values obtained by Shimizu (1990:128-129) were higher for the [pʰ, tʰ, kʰ] series than for the other two, which did not show any noticeable F0 difference between them. Also, the difference between values obtained for the [pʰ, tʰ, kʰ] series and the [p, t, k] (and possibly the series [b, d] series?) was statistically significant, but that between the [p, t, k] series and the [b, d] series was not statistically significant. This appears to signal that F0 onset values may not be a reliable distinguishing cue for stops in Thai. An earlier study by Gandour (1974), however, reported higher F0 values for the [p, t, k] series than for the [pʰ, tʰ, kʰ] series. F0 contours were also different for the different voice types. For the [pʰ, tʰ, kʰ] series, the F0 contour was high and falling. For the for the [b, d] series, the contour appeared to be rising and then becoming level, whilst the [p, t, k] series showed a somewhat level pattern. Gandour (1974) found that for both the [p, t, k] and the [pʰ, tʰ, kʰ] series, F0 appeared to be high and falling, whilst for the [b, d] series, it appeared to be low and rising. But further research needs to be done on F0 contour on Thai stops due to inconsistencies mentioned above. As far as F1 onset was concerned, the [pʰ, tʰ, kʰ] series again showed higher values than the other two.

**Phonetic features as the bases for a distinctive phonological feature in Thai.**

As mentioned above, distinctive features have a phonetic basis, and are described using phonetic features. Basing on the articulatory and acoustic characteristics examined above, the next step is to determine which phonological features can best distinguish Thai stops.

In most feature geometries, the feature [± voice] is used to express the basic distinction between voiced and voiceless between segments in the languages of the world. In articulatory terms, it is determined by adjustments in the glottis, and is realized by quasi-periodic low frequency vibration of the vocal folds under appropriate aerodynamic conditions. Acoustically, it is expressed by means of
timing: voice onset time (VOT). Three different categories may be adequately distinguished by VOT: voicing lead (prevoicing), short voicing lag and long voicing lag. All these three categories have been observed for Thai stops. The feature [± voice], therefore, appears to be one of the distinctive phonological features in Thai.

Articulatory glottal adjustments may also cue another phonological feature [± aspiration], a feature notably absent in some of the earlier feature geometries, e.g. Jacobson and Halle (1956:29 ff.) and Chomsky and Halle (1968:298 ff.), but present in Ladefoged's traditional features (Ladefoged 1982:25 ff.). Aspiration is also described in terms of timing relations between events in the glottis and in the supraglottal cavities. More specifically, it involves duration between the release burst in the mouth, signalled by abrupt increase in energy on the waveform, and the commencement of phonation in the larynx. Ladefoged (1982) defines aspiration simply as 'delay in onset of voicing'.

Opinions have been varied with regard to the determination of aspiration acoustically. The point of argument has been where aspiration ends. Some researchers have simply adopted the positive VOT technique, where the duration between the burst, marked by abrupt increase in energy on the (Sp) waveform, and the start of voicing, indicated by the first recognizable trace on the Lx waveform, was thought to be adequate for determining aspiration. Other researchers have argued that positive VOT simply indicates where vocal fold vibration for the following vowel begins, which is not necessarily where aspiration ends. Such researchers have proposed that the end of aspiration appears to be more detectable around the start of F2. Nevertheless, there is general agreement that aspiration entails a delay of voicing onset.

5 Voicing may also be determined on the basis of how many and/or how often tokens in an experiment (in a language) are pronounced with vocal cord vibration, particularly in post voiceless/utterance initial context(s). This, rather than measuring the duration of prevoicing, could actually be more revealing for languages like English, where the voice status of stops in this contexts is, to say the least, rather inconclusive. In English, the [b, d, g] series of stops may be realised without voicing in post voiceless/utterance initial context(s), and if there is any at all, it often occurs rather towards the release burst, being considerably short in duration.

6 As will be pointed out, some researchers determine this acoustically using (positive) VOT. VOT thus captures, in a single parameter, both voicing and aspiration. Where negative VOT is often called prevoicing.

7 This, though, may be regarded by other researchers as breathy onset of the following vowel.
In articulatory terms, aspiration is expressed by an out rush of turbulent, airflow from an open glottis after the release burst. Acoustically, it is realized by several phonetic features: long positive VOT values, higher range of F0 values (and contour), and F1 onset frequencies than for the unaspirated stops, and aperiodic noise on the Sp waveform after the burst. These were observed for Thai stops (although some of them could not reliably distinguish between the voicing categories, e.g. F0 and F1), making the feature [± aspiration] one of the possible distinctive phonological features in Thai.

Thus it appears that Thai uses both [± voice] and [± aspiration] distinctively in word initial position. Based on the studies\(^8\) and discussions presented above, the distinctive feature(s) for Thai and their basic and non-basic correlates may be reasonably summarised as follows.

---

\(^8\) It is, of course, acknowledged that in order to accurately and fairly present all the basic and non-basic correlates of the phonological features presented here for Thai, other phonetic feature which were not investigated in the mentioned studies, e.g. stop closure duration, duration of the preceding vowel, burst intensity, breathy phonation, etc. also have to be incorporated.
Chapter 7: Some implications of the acoustic investigation for general phonetics and phonology

Feature specification for Thai stops would then be as follows: [+ voice] [-aspiration] [b, d]; [- voice][+aspiration] [pʰ, th, kʰ], and stops with the [- voice] [-aspiration] specification would then be the [p, t, k] series. It is acknowledged that other phonetic features, e.g. stop closure duration, burst intensity, duration of the preceding vowel, breathy phonation, etc need to be included to complete the picture.

Korean. Korean is also a three-category language, with stops produced at three places of articulation: bilabial, alveolar/dental, and (palatal)-velar, and all being voiceless in word-initial environment. The terminology used to describe Korean stops has been a matter of great controversy amongst phoneticians and phonologists. The main problem hinges on the classification of the stops according to manner of articulation. The different stop series have acquired more than one description because of a great deal of variation and interchange of description for the stop types amongst phoneticians and linguists. For example, in the literature, the [p, t, k] series has been described as: 'forced', 'strong', 'long', 'fortis', 'checked', 'voiceless', 'tense', 'unaspirated', 'glottalized' 'laryngealized'; the [p*, t*, k*] series as 'voiceless', 'weakly/slightly aspirated', 'weak', 'lax'; and the [pʰ, th, kʰ] series as 'voiceless', 'tense', 'strongly/heavily aspirated' 'lax' (see, for instance, Han & Weitzman (1970), Abramson & Lisker (1972), Hardcastle (1973), Ladefoged (1973), Kagaya (1974), Hirose & al. (1974), Iverson (1983), Shimizu (1990), Jessen 1997:102-104) to name a few). Sometimes the stop categories are simply referred to as Type/Series/Category I, II and III, respectively. This ambiguity and confusion in describing and/or classifying Korean stops highlights a great deal of the difficulty faced by phoneticians and linguists in characterizing their phonetic properties, particularly on purely phonetic grounds (cf. Shimizu 1990:74).

There has also been a lot of confusion in the literature about the symbols used particularly for the Korean unaspirated and weakly aspirated stops. Sometimes the unaspirated stops are clearly distinguished from their weakly aspirated cognates by using capital letters: [P, T, K], e.g. Hardcastle (1973). However, symbols like [p, t, k] (e.g. Shimizu 1990) and [p', t', k'] (e.g. Kim 1970, Shin 1997) have also been used to represent the unaspirated stops, and [p*, t*, k*] (e.g. Shimizu 1990) and [p', t', k'] (e.g. Lisker & Abramson 1964) to represent the weakly aspirated stops. It appears that it is up to individual researchers to choose which symbols they use to represent the unaspirated and weakly aspirated stops in Korean. This may also contribute to the confusion faced by researchers in studying Korean stops in the literature. In the discussion presented below, the
Korean stops will be named as follows. \([p, t, k]\) will simply be referred as the unaspirated series; \([p^*, t^*, k^*]\) as the weakly aspirated series and \([p^b, t^b, k^b]\) as the aspirated series.

Studies on stops in Korean stops has been conducted from a wide variety of angles: physiological: e.g. Kim (1965), Hirose et al. (1973); acoustic: e.g. Lisker and Abramson (1964), Hardcastle (1973); perceptual: e.g. Han and Weitzman (1970), laryngographic: e.g. Abberton (1972), and aerodynamic: e.g. Dart (1987) and Kagaya (1974). A number of studies amongst these have investigated a combination of these areas.

**Physiological characteristic.** Kim (1965, 1970), conducted a cineradiographic study on Korean stops, focussing on laryngeal gestures during the articulation of these stops. He argued that the three Korean voice types differed principally because of the different glottal widths they exhibit at the plosive release. Glottal width and aspiration duration are correlated. When the glottal width is widest, aspiration is longest, and vice versa (see also Kagaya 1974:169-170, 173). Thus glottal width at the release is the most significant factor in differentiating between the stops, and aspiration duration is a mere concomitant of this. Regarding the development of glottal opening (and the size/shape of the glottal width) for the different stops in Korean, a fiberoptic investigation by Kagaya (1974) reported that in the aspirated stops, the glottis begins to increase at the start of the utterance (where the target consonant was in word initial environment (CV)) and reaches its local maximum at the articulatory release, where its width ‘is comparable to that of respiratory position’, and decreases dramatically for the voicing of the following vowel. In the unaspirated stops, ‘the glottal width seems always to decrease monotonically’ (pg. 166), and the vocal processes actually make contact before the stop is released, at which point a spindle-shaped gap could be observed in the membranous portion of the glottis. The stop burst happens just before or at voice onset. In the weakly aspirated stops, the development of glottal opening as well as its closure occurs more gradually compared to the other stop types. Thus, the aspirated stops have maximal glottal width of the three stop types, followed by the weakly aspirated and lastly by the unaspirated. Glottal width is also correlated to the timing of glottal adjustments after the release burst. Wide glottal width in aspirated stops results in delay of voice onset and thus large positive VOT, narrow glottal width in unaspirated stops result in almost coincident voice onset after release, with the slightly aspirated stop somewhere in the middle.

Another physiological study by Hirose et al. (1973), using electromyographic technique, reported that during the production of word initial
unaspirated stops, lateral cricoarytenoid (LCA) and especially the vocalis (VOC) muscle 'showed a marked increase in activity before the stop release ... which presumably resulted in an increase in inner tension of the vocal folds as well as in constriction of the glottis during or immediately after the articulatory closure (pg. 151). This increased activity of the LCA and VOC muscles was dramatically reduced from the peak, which, for the unaspirated series, occurred before the commencement of voicing, for the weakly aspirated stop after voice onset and for the aspirated stop at or immediately after voicing commencement.

In addition, Lee and Smith (1973)\(^9\) measured subglottal and intra-oral pressure for the three Korean stop types in word initial position. They found that, at the release burst, the mean values for subglottal pressure was 9.2 cm H\(_2\)O for the aspirated stops, 8.6 cm H\(_2\)O for the unaspirated stops and 8.2 cm H\(_2\)O for the weakly aspirated stops. Acoustic consequence for this in terms of F0, as will be pointed out below, is that both aspirated and unaspirated stops show higher F0 values due to rather stiff vocal folds and increased transglottal pressure difference than the weakly aspirated stops (cf. Kagaya 1974:175). Also, F0 contours for these stops begin at a higher value than for the weakly aspirated stops. It was pointed out above that the glottal width for the unaspirated stops appears to always be decreasing to a spindle-shaped gap before the burst. In this conditions, the transglottal air pressure difference necessary for voicing in the unaspirated stops may be created by expanding the supraglottal cavity, which could be achieved by positively lowering the larynx. This, in turn, means a drop in transglottal air pressure just before voicing. This was observed in the measurements performed by Lee and Smith (1973), and noticeable in the F0 contour investigated by Shimizu (1990) discussed below. Kagaya (1974:177) proposed 'stiffness' (of the vocal folds) as one of the features of Korean unaspirated stops,\(^{10}\) as well as subglottal pressure (see also Hirose et al. (1974:152), and Hardcastle 1974:(269-271)).

Further investigation on laryngeal activity during the production of Korean stops has also been carried out by Abberton (1972:75). For the weakly aspirated stops, she writes that they 'are associated with a characteristic laryngeal activity'\(^{9}\)

---

\(^{9}\) Due to the fact that I could not consult the article, I am indebted to Kagaya (1973:175) for this information.

\(^{10}\) On the basis of the different physiological and physical properties made for the unaspirated and aspirated stops, Kagaya (1973:178) argued that, pending further research, they could not therefore be classified under the same phonological feature [+ tense] (see also Hirose et al. 1974:151).
at least for some of her speakers. This was not observed for the other two stop types.

All this evidence has led some researchers to propose that another feature, the tensity feature, is needful to adequately distinguish Korean stops. Such researchers include Kim (1965) who also examined the muscular tension of the vocal folds and proposed that the tensity feature may also be employed usefully in differentiating Korean stops. Thus, an independent feature of tensity, with the opposition tense/lax, could be used together with voicing features of voice/voiceless and aspirated/unaspirated to describe and differentiate Korean stops. This would help to alleviate the frequently observed overlap in VOT values reported by a lot of analysts. Tensity as an independent differentiating feature in Korean stops has also been proposed by, for instance, Hardcastle (1973)\(^{11}\). But inspite of the suggestion of this tensity feature, as pointed above, the controversy has remained as to which stop category in the language may be described as ‘tense’. The two categories which have often been described with this feature are the aspirated and the weakly aspirated.

In acoustic terms, ‘tense’ stops generally have been reported to differ from the lax stops (or from the stops lacking this feature) in the following way. Tense stops exhibit higher F0 values at the start of the following vowel. Pulses in the initial portion of the vowel are noticeably large, sharp (i.e. high amplitude of the pulses at vowel onset) and undamped. Lax stops are expressed by low burst intensity, low amplitude of pulses at vowel onset, low energy distribution after the burst, i.e. during the aspiration phase (cf. Kim (1965), Hardcastle (1973), Han and Weitzman (1970)).

**Acoustic analysis.** This has focussed mainly on measuring VOT. Researchers include Lisker and Abramson (1964), Kim (1965), Han and Weitzman (1970), Shin (1997), etc. Most of the studies have reported long (positive) VOT values for the heavily aspirated stops, and rather overlapping values for the other stop types.

Shimizu (1990:79 ff) analysed F0 values and contour at vowel onset, F1 onset as well as other phonetic characteristics of Korean stops. With regard to F0 perturbation, the unaspirated series and the aspirated series consistently showed higher F0 values than the weakly aspirated series, and this was attributed to ‘increased tension of laryngeal muscles and increased air-flow rate in the transglottal area’ (pg. 79, see also Hardcastle (1973), Kagaya (1973) and Hirose et al. (1990)).

\(^{11}\) Hardcastle (1973:264) in fact uses the term in a specific sense, defining it as ‘referring to an increase in isometric muscular tension, primarily in the vocal cords and the pharynx. Muscular tension, it would appear, would be the basic (physiological) correlate of the feature ‘tensity’.
For the F0 contour, the \([p^h, t^h, k^h]\) series showed high values with a rather steady falling contour pattern. The weakly aspirated series showed high values but an abrupt falling pattern compared to the aspirated series. The unaspirated series began with lower values than the other series, and showed a rather steady, somewhat level pattern.

Physiological and aerodynamic factors have been used to explain these different F0 curves. As pointed out above, the unaspirated stops have been associated with some characteristic laryngeal activity or increased tension in the vocal folds. It is postulated that lowering of the larynx to expand the subglottal cavity is what possibly causes dramatic change in the F0 curve. Vocal fold tension is also present in the production of the \([p^h, t^h, k^h]\) series. But with this series, high airflow after the release means the folds remain stiff and hence keeps the F0 values relatively steady, which explain the rather steady falling nature of the curve. F1 onset values for the heavily aspirated stop series \([p^h, t^h, k^h]\), was also observed to be higher than for the other stop types: \([p, t, k]\) and \([p^*, t^*, k^*]\). The difference between the values obtained for the heavily aspirated and the weakly aspirated series was statistically significant, whilst that between the weakly aspirated and the unaspirated series was not.

Shin (1997:115-142) measured closure duration for the Korean stops, and observed that the unaspirated stops showed longer closure duration, followed by the aspirated stops and lastly the weakly aspirated stops. The differences between the values obtained for the stops were statistically significant, indicating that closure duration is a highly reliable distinguishing factor in Korean stops (pg. 118-119).

As may be apparent from the discussion above, more extensive research on Korean stops is clearly still necessary. However, if we accept the view that the production of the unaspirated series entails some tension in the laryngeal muscles, then we may consider that some of the acoustic correlates of this articulatory behaviour may well be as stipulated above: higher F0 values and F1 onset than the weakly aspirated cognates.\(^{12}\)

Korean distinctive features. Based on the studies and discussions presented above, the distinctive feature(s) for Korean and their basic and non-basic

\(^{12}\) As part of the investigation articulatory characteristics of Korean stops, Shin (1997:207 ff.) also examined variation in contact pattern for the stops as a result of flanking vowels, and noted that 'the lax stop (weakly aspirated stops) shows more variation in contact area as a function of vowel contexts than the aspirated or tense (unaspirated) stops (pg. 213).
correlates may be attempted in the following way. As has been pointed out, all word-initial Korean stops fall on the positive half of the VOT continuum, and are therefore [- voice]. The feature [± aspiration] is also applicable since aspiration functions distinctively in the language. It was also pointed out that the tensity feature, [± tense] is necessary to satisfactorily differentiate between the stops. Consider figure 7.3.

Basic and non-basic correlates for the distinctive phonological feature [voice] in Korean

![Diagram showing basic and non-basic correlates for [voice] in Korean](image-url)
Basic and non-basic correlates for the distinctive phonological feature [tense] in Korean

\[
- \text{tense}
\]

- Basic correlate
- Non-basic correlates

- Glottal tension
- High intra-oral pressure
- High F0 values
- High F1 onset
- (Long?) closure duration

Fig. 7.3 (b): The basic and non-basic correlates for the distinctive phonological feature [tense] in Korean. (The non-basic correlates are not necessarily arranged in order of priority.)

Basic and non-basic correlate for the distinctive phonological feature [aspiration]

\[
+ \text{aspiration}
\]

- Basic Correlate
- Non-basic Correlates

1. Delayed phonation after burst
2. Glottal tension
3. F0 onset and contour
4. F1 onset
5. Closure duration

Figure 7.3 (c): Basic and non-basic correlates for the distinctive phonological features [aspiration] in Korean. (The non-basic correlates are not necessarily arranged in order of priority.)
Feature specification for Korean stops would then be as follows: [- voice] [+ tense] [+ aspired] = [pʰ, tʰ, kʰ], the aspirated series; [- voice] [+ tense] [- aspiration] = [p, t, k], the unaspirated series; and [- voice] [- tense] [- aspiration], the weakly aspirated series, sometimes known as lax stops, which are voiced intervocally.

Jessen (1997:102-104), proposed an alternative distinctive feature representation for Korean stop and argued that the distinction between the weakly aspirated stops and the unaspirated stops in Korean could best be represented by using the feature [± checked] as well as [± tense]. He operates within the Jacobson et al. (1952) distinctive feature geometry where [+ checked] is articulatorily defined as sounds with a ‘compression or closure of the glottis’ (Ibid.: 23). His focus, however, is to argue for the validity of the feature [± tense] as a distinctive feature (at least in German). This seems to support the proposal of the tensity feature which could help to differentiate Korean stops. The introduction of the feature [± checked] highlights the terminology problem at the heart of the classification of Korean stops.

Burmese. Like Thai and Korean discussed above, Burmese has a three-way voicing contrast in its stops system, with stops being produced at three places of articulation. These are: [b, d, g], [p, t, k] and [pʰ, tʰ, kʰ]. Burmese also has tonal contrasts, like Thai.

**Acoustic analysis.** Voice onset time: VOT values obtained for the voiced series [b, d, g] were large and negative\(^\text{13}\), although this was not consistently so for all the informants in the study (Shimizu, 1990:99). The aspirated series; [pʰ, tʰ; kʰ] manifested large positive VOT values and the unaspirated series; [p, t, k], showed small positive VOT values. With respect to F0 perturbation and contour, F0 values obtained were highest for the [pʰ, tʰ; kʰ] series, followed by the voiceless unaspirated series: [p, t, k], and lastly the voiced stops [b, d, g] (Shimizu 1990:128-129). The difference between F0 values for these stops was statistically significant. F0 contours were also different for the different voice types. For both the [p, t, k] and the [pʰ, tʰ, kʰ] series, F0 appeared to be high and falling, whilst for the [b, d, g] series, it appeared to be low and rising. As far as F1 onset was concerned, the [pʰ, tʰ, kʰ] series again showed higher values than the other two,

---

\(^{13}\) This, however, was not consistently so for all the speakers used in some of the studies (e.g. Shimizu, 1990:99). It is however indicated that laryngeal timing for speakers who did not realise voicing lead for the voiced stop may possibly have been affected by the acquisition of the second language (English), since these speakers had been living in an English-speaking community for a long time.
although this was affected by vowel quality since the difference in F1 onset was more clearly observed when the subsequent vowel was not a non-high vowel (cf. Shimizu 198).

**Phonetic features as the bases for distinctive phonological feature in Burmese.** It appears that Burmese uses both \[± voice\] and \[± aspiration\] distinctively in word initial position. Basing on the studies and discussions presented above, the distinctive feature(s) for Thai and their basic and non-basic correlates may be reasonably summarised as follows.

Basic and non-basic correlates for the distinctive phonological feature \[voice\] in Burmese

![Diagram](https://via.placeholder.com/150)

Figure 7.4 (a): Basic and non-basic correlates for the distinctive phonological feature \[voice\] in Burmese

Basic and non-basic correlates for the distinctive phonological feature \[aspiration\] in Burmese

![Diagram](https://via.placeholder.com/150)

Figure 7.4 (b): Basic and non-basic correlates for the distinctive phonological feature \[aspiration\] in Burmese
Feature specification for Burmese stops would then be as follows: [+ voice][- aspiration] [b, d, g], [- voice][+aspiration] [pʰ, tʰ, kʰ], and [- voice][- aspiration] specification would then be the [p, t, k] stops. It is acknowledged that, as for Thai, other phonetic features, e.g. stop closure duration, burst intensity, duration of the preceding vowel, breathy phonation at vowel onset, etc. also need to be included.

To summarize, from the studies reviewed on the languages discussed above, it appears reasonable to conclude that languages appear to select different phonological features from the universal distinctive features to express contrast for stop voicing (in word initial context) in their own systems. This is acoustically manifested along several dimensions, e.g. F0 and F1 onset characteristics, VOT, glottal gestures.

7.4 Shekgalagari stops: distinctive features

This section focuses on Shekgalagari stops, assessing the contribution of the findings of the acoustic experiments performed in Part Two of this thesis to linguistic theory, and, on a specific level, to the feature theories presented in Chapter Three in Part One of this thesis. In order to explore the phonetics and phonology of Shekgalagari stops in greater detail, the results of the experiments will first be reviewed briefly.

7.4.1 Review of acoustic experimental results

7.4.1.1 The spectrogram, Speech and Lx waveforms

In the analysis of the waveforms presented in Chapter Four, three main categories of stops were identified from the data: voiced, voiceless unaspirated and voiceless aspirated.

The voiced stops were characterized by modal phonation during the occlusion of the stop, which continued after the burst into the vowel. Although irregular vibration was observed for some tokens and for some speakers, the total number of these was very small number and could be attributed to random variability in the production of tokens, considering the fact that the majority of voiced stops were produced with modal voice. It was pointed out that on the whole the waveform characteristics for these stops conformed to their traditional description as ‘voiced’ stops.
Chapter 7: Some implications of the acoustic investigation for general phonetics and phonology

For two of the speakers the voiced stops manifested decay of voicing towards the release explosion and a brief delay of voice onset after the burst for tokens produced in isolation and within a frame sentence (see figures 4.6 (b) in Chapter Four and (d), and § 5.6 in Chapter Five). It was interesting to observe that the cessation of voicing did not only remain for the rest of the stop up to the burst but continued for a brief period after the release burst until the start of the subsequent vowel. Voicing did not actually start immediately after the release burst. This decay of voicing was also observed in a pilot study. Thus from the burst to the vowel these voiced stops seemed to behave more or less like the voiceless unaspirated stops in this language.\textsuperscript{14} Cross-linguistic study has shown that sounds that are categorized similarly, e.g. 'voiced', may often be produced differently in different languages and may even be produced differently by different speakers in a given language (Shimizu, 1990). Shekgalagari voiced stops (at least for the two speakers of this study) may be considered to be an example of this. Nevertheless, we pointed out that Shekgalagari voiced stops may still be regarded as truly voiced stops since even when there is voicing decay, active voicing happens for most of the stop duration.

For the next category, voiceless unaspirated stops (traditionally described as ejectives), a large proportion of the data was produced with modal phonation on the Lx signal prior to the stop closure and at the following vowel onset (at least for the subjects used in this investigation). There was no cycle-by-cycle fluctuation in period and intensity, no gradual build-up of intensity was observed, and a small, rather insignificant burst was a feature in most of the waveforms for these stops, with the burst not even observable in some of the waveforms.

Ejection in stops has been reported to exhibit: irregular mode of vocal fold vibration prior to the target stop and/or at vowel onset, with period-by-period fluctuation in amplitude, e.g. Hausa (Lindsey, Hayward & Haruna 1992:520), Korean, (Abberton 1972:73); long VOT values, e.g. 80+ ms for some North

\textsuperscript{14} It was pointed out in Chapter Four that decay of voicing during the occlusion for a voiced plosive has also been observed for languages like Degema, (Lindau 1984:148-149) and Xhosa (cf. Jessen 1999:2, Jessen & Roux 1999:5), and English. In Xhosa, the interesting situation being lack of, or near lack of, prevoicing even in intervocalic environments where most languages show the presence of voicing of some form and to some extent during the stop closure. This situation makes the relationship between Xhosa 'voiced' stops (in word initial context) and their voiceless unejected unaspirated cognates very interesting.

The phonation status of English word initial 'voiced' stops is rather complicated. There may be phonation near the burst, immediately after the burst or it may be delayed after the release burst. In intervocalic position, voicing may die out and start again during the articulation of the stop.
Chapter 7: Some implications of the acoustic investigation for general phonetics and phonology

American languages and some Caucasian languages; long closure duration, e.g. Navaho (Lindau 1984); high burst amplitudes, e.g. Xhosa (Jessen 1999:28) and Tigrinya (Fre Wuldo 1985); higher oral air pressure than other voiceless stops in some languages, e.g. Tigrinya (Fre Wuldo 1985:126 ff); greater low frequency perturbation (slow up and down deflection indicating decreased open quotient) in the Lx waveform at vowel onset, e.g. Hausa (Lindsey, Hayward & Haruna 1992:520); there could even be evidence of whole larynx movement, e.g. Korean, (Fourcin & Abberton 1977:313), Xhosa (cf. Chapter Four of this thesis). Often languages select how they cue their own ejectives, and what would be a significant distinctive cue for ejectives in one language may not necessarily be a cue for the “same” types on stops in another language.

It was pointed out that different languages show different signal characteristics, and sometimes speakers also show considerable variation in their production of sounds. Considerable variation may also exist when the same ejective stop is produced several times, where a token could sometimes be produced in a plain manner and sometimes as an ejective stop.

Although a few tokens were produced with irregularity of vibration at the commencement of the vowel for one or two informants for the voiceless unaspirated stops in Shekgalagari, this could most probably be due to speaker variation and/or token-to-token variation, which is not an uncommon observation in studies like the present one. It was proposed that, on the whole, the observations made for Shekgalagari voiceless unaspirated stops seem to suggest that these were simple plain voiceless unaspirated stops, and not ejectives.

As regards the voiceless aspirated stops, the results showed turbulence after the burst and vowel onset displayed breathiness. This seems to conform to, as well as confirm, their traditional description as voiceless aspirated stops.

**Gx analysis**

The Gx signal appeared to show some different movement between the stops according to their voicing structure: distinctive downward movement was made for the voiced stops in contrast with the movement it made for the voiceless stops.

*The voiceless stops.* The shapes of the Gx trace observed for the voiceless unaspirated stops and for the voiceless aspirated stops were very similar (cf. figures 4.10 (b) and 4.11) — most tokens show a relatively level Gx waveform for these stops (see figure 4.10 (a) ~ (b) and 4.11). Gx behaviour after the release of the burst for the aspirated and unaspirated voiceless stops seemed to be affected
Chapter 7: Some implications of the acoustic investigation for general phonetics and phonology

by properties of the subsequent segments e.g. tone type, vowel height and F0, and possibly by a combination of all the factors. These shapes were not observed for the voiced stops. This appeared to suggest that Gx movement was possibly more in keeping with the voicing status of the stops than with the displacement of the larynx due to airstream mechanism.

It was pointed out that the behaviour of Gx in the production of voiced vs. voiceless stops in this study appeared to support Hombert et al. (1979) and Ewan and Krones (1974)'s larynx height hypothesis, where the voiced-voiceless dichotomy of stops is distinguished by the vertical height of the larynx. They, however, continue to say that 'additional evidence in favour 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). Other factors influencing the behaviour of the Gx trace at vowel onset which may not necessarily part of the properties of the stops include, e.g. tone, F0 perturbation, by high vowel, or, possibly, a combination of all these three factors.

With respect to Xhosa stops, there was upward deflection of Gx in the production of the Xhosa ejective [p'] in the word [p'enap'ema] 'throb with pain' produced in isolation and within a carrier sentence. It was also pointed out that this could plausibly be attributed to the abrupt displacement of the larynx as it initiates the airstream mechanism required for the production of these stops in Xhosa, at least for our speaker.

This distinctive, sharp upward deflection of Gx was however not observed for the voiceless unaspirated stops in Shekgalagari, which may possibly be suggestive of the probability that Shekgalagari voiceless unaspirated stops may not be produced on the glottalic airstream mechanism, and may therefore not be ejectives as they have traditionally been thought to be.

As was mentioned in Chapter Four, only one person served as an informant for the investigation of Xhosa stops. For this reason, Gx observations made here with regard to the glottalic stops in this language may not be conclusive. Other (acoustic based) studies (e.g. Jessen 1999, Jessen & Roux 1999) have since shown that ejection in Xhosa seems to be speaker dependent, and that there was also variation on this aspect from token to token. Other features which have been associated with ejectives, e.g. creaky voice at vowel onset were uncommon for most speakers in these studies, and occurred in other stop types which were not necessarily glottalic. Jessen (1999:29) concludes that due to this instability, it is difficult to consider ejection as a phonological aspect of Xhosa, although it does seem to be a characteristic of Xhosa stop system which has to be learned.
7.4.1.2 Voice onset time

VOT as an effective differentiating parameter in Shekgalagari stops

VOT values for the three stop types showed long negative values for the voiced stops (which included voice decay during the stop occlusion, but not the short positive VOT after the burst), short positive values for the voiceless unaspirated stops and long positive values for the voiceless aspirated stops. No overlapping values were obtained between any of the categories. The ranges of the means between the three stop types were considerable. The VOT values obtained here for the voiceless unaspirated stops in Shekgalagari have generally been low (see table 5.1 (a)). Any higher values obtained appeared to correlate more with the effect of place of articulation than with the voice type for the stops, and were no-where near the values obtained for the voiceless aspirated stops, as has been reported for ejectives in some North American Indian and Caucasian languages. Had high VOT values been a feature of these stops in this language, it would have occurred across all the five places of articulation. We pointed out that this was strongly suggestive of the possibility that the voiceless unaspirated stops in Shekgalagari may not be ejectives, at least as far as VOT is concerned, and for the subjects of the study.  

7.4.2 Evaluation of feature theories in the light of the results obtained for Shekgalagari

Seven different feature theories were presented in Chapter Three, section 3.3. The situation of the word initial voicing structure of Shekgalagari stops will now be assessed in detail in the light of each of these theories.

16 Also, there was some interaction between VOT and place of articulation, and the quality of the following vowel. The ranking for place of articulation was in the following order: palatal > velars > uvulars > dentals > labials, and that for the vowels was as follows: [u], [i], > [o/o], [e/e] > [a], and, for the voiceless stops, this trend appeared to be fairly consistent in these data.
Chapter 7: Some implications of the acoustic investigation for general phonetics and phonology

7.4.2.1 The Sound Pattern of English model (SPE)

In the SPE model of representation of speech sounds reviewed in section 3.3.1, segments are described in terms of inventories of features called feature matrices, with features being binary valued. There are a number of rows containing the phonetic properties (features) of the segment, and there is a column, which represents the segment being described. The theory comprises two types of rules: phonetic rules and phonological rules. Phonetic rules operate on the features in a language specific manner. The terminal output of the phonetic rules is the phonetic transcription which also encodes information relating to the pronunciation of the output as determined by the grammar of the language concerned. Phonological rules may alter the values for the features, they may insert or remove segments, but they may not change the content of the feature matrices — which specify segments. Distinctive description between the natural classes of segments is done by this component of the grammar. The specification for Shekgalagari stops would be as follows.

(1) The voiceless unaspirated stop

[p]
[+ consonantal]
[- sonorant]
[- vocalic]
[- tense]
[- voice]
[- stress]
Chapter 7: Some implications of the acoustic investigation for general phonetics and phonology

(2) The voiceless aspirated stop

\[ p^h \]
\[ + \text{consonantal} \]
\[ - \text{sonorant} \]
\[ - \text{vocalic} \]
\[ + \text{tense} \]
\[ - \text{voice} \]
\[ - \text{stress} \]

(3) The voiced stop

\[ b \]
\[ + \text{consonantal} \]
\[ - \text{sonorant} \]
\[ - \text{vocalic} \]
\[ - \text{tense} \]
\[ + \text{voice} \]
\[ - \text{stress} \]

(where the dots at the end indicate that the specification for the stop is not complete). Similar feature matrices could be drawn for stops produced at other places of articulation.

It is obvious that an endless list of features would be required to provide, if it were possible, a complete specification of the stops. Both unique phonetic description and phonological information is provided by one and the same matrix for any given stop. At a cross-linguistic level, the SPE feature geometry does not have a single phonological feature of 'voicing' which distinguishes, say, 'voiced' sounds from 'voiceless' sounds across languages.
7.4.2.2 Lisker & Abramson (L&A) (1964): voice onset time (VOT)

In section 3.2.1 VOT was discussed, and the temporal relations between the articulatory explosion of the stop and the onset of voice pulsing. It was pointed out that this may be specified by a single value. VOT has been shown to be highly reliable in differentiating stops with different voice types in many languages.

In word initial CV context, voicing may start prior to the release burst (voicing lead) giving negative VOT values, it may coincide with the release (coincident phonation) giving short positive VOT values, and it may be considerably delayed after the burst (delayed phonation) resulting in long positive values. Actual VOT values may vary from language to language. Some languages have a two-way variation in their stop systems, others are three-way systems, and yet others manifest a four way-contrast.

In Chapter Five, Shekgalagari stops were focussed on, assessing whether and how effectively VOT can differentiate the voice types. Long VOT lead values were obtained for voiced stops, short-lag values for voiceless unaspirated stops and long-lag values for the voiceless unaspirated stops. There were no overlapping values between any two categories of the stops, and the ranges of the means between the three stop types were considerable. VOT was therefore considered to be an effective distinguishing cue for the three stop types in Shekgalagari.

7.4.2.3 Halle and Stevens (1971) (H&S): laryngeal features

The H&S feature system, as observed in section 3.3.3, is limited to vocal fold activity at the moment of stop release: the tension of the vocal folds (i.e. slack/stiff folds) and glottal aperture. The various manipulation of the vocal cords give the following four features: [spread glottis], [constricted glottis], [stiff vocal cords] and [slack vocal cords], which H&S propose should be incorporated into the universal phonetic feature framework (H&S:201). By means of combination, these four features produce nine distinctive phonetic categories of segments. These phonetic categories and their feature specification are summarised in Table 3.1, obtained from Halle and Stevens (1971:201).

By using the features [spread glottis] and [constricted glottis], stops may be divided into three broad groups: plain [-spread, -constricted], aspirated [+spread, -constricted], and glottalized [-spread, +constricted]. These three groups are further subdivided into three groups by using the features [stiff vocal cords] and [slack vocal cords]. These subdivisions are the voiceless stops, voiced stops
and a third group which incorporates implosives, lax stops such as the Danish [b] and the Korean moderately aspirated stop. Voiceless stops are marked by the features [+stiff, -slack], and examples include voiceless unaspirated stops, voiceless aspirated stops and ejectives. Voiced stops are marked by the features [-stiff, +slack] and they include traditionally voiced stops, e.g. the French [b], ‘murmured’ stops, e.g. the Hindi [bʱ], and a third class of stops which, as opposed to the other two stop types, is glottalized. The third class of stop types has the configuration [-slack, -stiff] and includes the true implosives shown in column 7, the ‘lax’ voiced stops [b̥] such as the one found in, for example, Danish and ‘may occur in initial position for many speakers of English’ (H&S 1971:206), and the Korean slightly aspirated stop [p̥k̊].

Within the H&S feature geometry, Shekgalagari stops may be described as follows. The voiced stops, with relaxed vocal folds to facilitate vibration and fairly regular mode of vocal fold vibration during the stop occlusion, may have feature specification [-stiff, +slack], [-spread, -constricted]. The voiceless aspirated stops, observed to have modal voicing before the stop, no voicing during the stop and considerable delay of voice onset after the stop, may have the features [+stiff, -slack], [+spread, -constricted]. The unaspirated stops, which showed regular mode of phonation before and after the stop and no phonation during the stop, may have the features [+stiff, -slack], [-spread, -constricted].

Some of the problems that may be encountered with the H&S feature specification for Shekgalagari stops could be similar to those already pointed out in section 3.3.3 for other languages. As the feature system is based on laryngeal activity, it seems that more specification may be necessary for every laryngeal activity in the production of the stops, leading to the problem of over-generation of features. For example, the tendency for some speakers to produce the ‘voiced’ stops with no voicing during the stop occlusion which would require the use of the feature [+stiff, -slack] instead of [-stiff, +slack]; or with creaky voice instead of modal phonation, where the specification [-spread, +constricted] would be more appropriate than [-spread, -constricted].

---

17 We observed that, as opposed to true implosives, [ɓ], where both the lowering of the larynx and laryngealized voicing occur, these particular glottalized stops are reliably indicated by accompanying laryngealized voicing only, but the downward displacement of the larynx may not always happen.
Keating’s proposals concerning the representation of phonological contrasts and their relationship with phonetic implementation across linguistic systems have been discussed in detail in section 3.3.4. Here, they are reviewed briefly in order to assess whether and how they address the voicing structure of Shekgalagari stops in word initial context.

Keating mainly deals with phonological representation of segments and their phonetic implementation in individual languages of the world. In the 1984 article, she proposed three kinds of representation; we will review two. The first one proposed that there should be as many phonological features e.g. [voice] e.g. and feature values [+voice], [-voice] as are required to distinguish natural classes in a language. The second proposed that there should be as many phonetic categories e.g. {voiced}, {vls. unaspir.}, {vls. asp.} as are needed to distinguish between categories in any given language. {voiced}, {voiceless unaspirated} and {voiceless aspirated} were proposed as a fixed universal and specified set coding possible contrasts and have acoustic and articulatory correlates. For example, {voiced} involved vocal-fold vibration and low periodicity during consonant closure, which is absent in voiceless stops. Also, these phonetic categories directly map onto VOT lead, short-lag and long-lag respectively for stops in word initial context. Languages implemented their phonological contrast differently by way of these phonetic categories.

English and Polish were considered for illustration. These are two-way category languages, with phonological contrast between [+ voice] i.e. /b, d, g/ and [- voice] i.e. /p, t, k/ stops. The implementation of the phonetic categories by these languages will be as follows. Depending on context, [+ voice] stops in English would be {vls. unaspirated}, while in Polish they would be {+ voice}. Similarly, English [- voice] stops would be {vls. aspirated} while in Polish they would be {vls. unaspirated}. Phonologically, the two languages are the same since they are two-way contrasting languages, but they are phonetically distinct since they implement the phonetic categories differently.\(^ {18} \)

**Phonological representation.** Keating (1984) went on to explain that a more accurate account of segment representation (for two-way contrasting languages) could be done by way of levels of representation. There is the

---

\(^ {18} \) Keating’s (1984) analysis was limited to two-way contrasting languages. Three-way and four-way systems such as Thai and Hindi were excluded from her discussion ‘largely because it is unclear whether such languages should be analysed as having a single non-binary feature [voice], or more than one binary feature’ (pg. 191).
Chapter 7: Some implications of the acoustic investigation for general phonetics and phonology

phonological level of representation and the phonetic category level of representation (level of implementation). At the phonological level of representation, phonological rules are applied to binary feature values (e.g. [± voice]), and their output is the phonetic category values (e.g. {voice}, {vls. aspirated}, {vls unaspirated}) which are different in different languages. These phonological rules cannot anticipate their phonetic output, in which case they can work with whatever output a language allows. Thus two level representation helps to keep phonetic details separate from phonological considerations, and also allows for different implementations of the binary phonological feature across linguistic systems.

Keating cited experimental data on vowel duration before 'voiced' and 'voiceless' stops in several languages (mostly two-category languages) to support her proposals. Vowels tended to be longer before 'voiced' stops than before 'voiceless' stops. Also, vowels tended to be longer before phonetically 'voiced' stops than before 'lax' stops. This, according to Keating (1984:291,292) appeared to show that the relationship between vowel duration and voicing was conditioned by the underlying phonological feature [± voice], rather than being mechanically determined by the phonetic voicing during the occlusion of the stop. This further buttressed the need for separating phonetic and phonological levels of representation.

*Phonetic representation.* Another of the levels of representation proposed by Keating dealt with phonetic categories. It was proposed that there should be as many phonetic categories as are needed to distinguish between categories in any given language. The phonetic categories dealt with here are described in word initial context, in terms of VOT ‘and the voicing dimension’, and limited to three contrasting categories: {voiced}, {vls. unaspirated}, {vls. aspirated}, since it is thought that languages have these categories only. Based on a study Lisker and Abramson (1984), the *three* categories proposed were thought to be ‘the right number’, and ‘are the same three in various languages’. Lisker and Abramson (1984), after studying the timing of voice onset after the release of the stop, concluded that languages exhibited no more than a three-way contrast in their stop systems.19 Linguistic systems which appeared to display more than three-way contrast were thought to have overlapping VOT values between at least two categories which could be distinguished through some other dimension, e.g. vocal fold tension. Thus the three phonetic categories: {voiced}, {vls. aspirated} and {vls. unaspirated} appear to be the basic ones implemented by the languages of

---

19 These authors, however, modified this position following subsequent investigations of languages like Hindi and Korean.
Chapter 7: Some implications of the acoustic investigation for general phonetics and phonology

the world, 'And in fact, they are also sufficient elsewhere,' since no greater number of contrasts is found in any other position' (pg. 297).

**Phonetic categories: the polarization principle.** As we have seen, the implementation of phonetic categories is language-specific. This, instead of being dictated by language-specific phonetic rules, could be derived by the principle of polarization, where two adjacent categories are polarized along the voicing dimension. 'According to this principle, within the limits of the implementation chosen — i.e. the phonetic categories — there is maximal separation of the distributions of values' (pg. 309). For example, Polish stops show 5 ms higher VOT values (and therefore slightly aspirated) than the English ones. According to the polarization principle, the contrast between the stops in Polish can be explained by polarizing the 'slightly aspirated' short-lag stop away from the long-lead voiced stop. But the situation with English is slightly complicated, since the so-called 'voiced' stops may have lead values (which are not as in truly voiced stops), and they may have lag values (which are not as in voiceless stops). Keating (1984:309) believes that this situation could be resolved by regarding English as having a bicategory distribution of VOT values: voicing lead (i.e. long negative VOT values) and short-lag. In this way, the English short-lagged 'voiced' stop can then be polarized away from the long-lagged voiceless stop.

Keating is careful to point out that the polarization theory may not always adequately address the contrast in all languages. It is possible that different VOT values may be obtained for the same phonetic category implementations in a language, making it difficult for the polarization principle to apply.

In the 1990 article, the two levels of representation: phonological and phonetic, are still maintained. Phonological representation describes overall contrast between segments in a language. Phonetic representation express contrast in a given context in a language. But this time phonetic representation has three levels. The first level, categorical phonetic representation, is still the output of the phonology in that it is the implementation of the phonological contrast as discussed above. It is regarded as neutral with regard to articulation and acoustics/perception, which are the concerns of the other two levels of phonetic representation. The relationship between the levels is summarized as shown in figure 7.5.

---

20 That is other than word initial position.
The features this time are for voicing, aspiration and glottalization. These are [voice], [spread glottis] and [constricted glottis], which may be related to particular landmarks during the production of the stop. For example, [voice] expresses events during the closure phase and is therefore associated with the closure node. It distinguishes truly voiced closure periods from voiceless ones. The value [+voice] then indicates vocal fold vibration and low frequency periodicity during stop closure. [spread glottis] and [constricted glottis] do not necessarily relate to a particular point in the articulation of the stop, but rather to the configuration of the vocal folds at the moment of release. [spread glottis] describes whether aspiration is present or not. The feature value [+ spread glottis], indicates the fact that the vocal folds are open at the release of the stop, leading to the presence of aspiration. The reverse is true with the value [- spread glottis], which explains that the vocal folds are closed at the moment of release, leading to lack of aspiration.

Keating (1990:326) went on to describe the traditional VOT categories: VOT lead, short-lag and long-lag using the three features mentioned above as shown on figure 7.6.
Chapter 7: Some implications of the acoustic investigation for general phonetics and phonology

The VOT lead (i.e. voiced) category

Stop

Closure

[-voice]

Release

[-spr gl]

Fig 7.6(a): Representation for VOT lead – for the voiced category of stops (Keating 1990:236)

Short VOT lag (i.e. the voiceless unaspirated category)

Stop

Closure

[-voice]

Release

[-spr gl]

Fig 7.6(b): Representation for short VOT lag – for the voiceless unaspirated category of stops (Keating 1990:236)
Chapter 7: Some implications of the acoustic investigation for general phonetics and phonology

Long VOT lag (i.e. the voiceless aspirated category)

![Diagram of Stop, Closure, Release, [-voice], [+spr gl]]

Shekgalagari word initial stops and Keating. The results obtained for Shekgalagari stops appear to bear out Keating’s theory as explained in terms of the phonological feature [± voice] and the phonetic features {voiced}, {vls. unasp.} and {vls. asp}.

For the [b, d, g] series, as an acoustic correlate, long VOT lead values were obtained. Also, as an articulatory correlate, a voice bar representing vocal fold vibration and low frequency periodicity at the gap corresponding to the stop closure was observed on the spectrograms. Within Keating’s framework, this series could be represented by the phonological feature [+ voice] and realized as {voiced}. For the [p, t, k] series, short-lag VOT values were obtained and gaps corresponding to the stop occlusion on spectrograms were essentially empty. This series could be represented by the phonological feature [- voice] and realized as {vls. unasp}. The [pʰ, tʰ, kʰ] series showed long positive VOT values indicating the presence of aspiration or turbulence after the burst, which, in spectrograms, appeared as aperiodic noise after the burst. This series could be represented by the phonological feature [- voice] and realized as {vls. asp}. In terms of the two levels of representation, Shekgalagari stops may be presented as shown in figure 7.7.
Chapter 7: Some implications of the acoustic investigation for general phonetics and phonology

Representation of Shekgalagari stops based on Keating (1984)'s two levels of representation

The three basic phonetic categories proposed by Keating certainly appear to be 'the right number' for Shekgalagari stops.

Shekgalagari stops and the principle of polarization. According to the polarization principle, the contrast between word initial stops in Shekgalagari can be described by polarizing the long lead stops from the short-lag ones, and the short-lag ones from the long-lag ones. This principle seems to be applicable in the Shekgalagari situation.

One level of phonological representation and three levels of phonetic representation.

Our results for Shekgalagari stops again appear to bear out Keating's model of three levels of representation. Consider figure 7.8.
Figure 7.8: Shekgalagari stops in terms of Keating's model of three levels of representation: phonological, categorical phonetic and articulatory and acoustic parametric.
Chapter 7: Some implications of the acoustic investigation for general phonetics and phonology

7.4.2.5 Kohler (1984) fortis and lenis stops

As already observed in section 3.3.5, Kohler (1984) proposed a single feature [± fortis], as adequate to express phonemic distinctions such as /p, t, k/ vis-à-vis /b, d, g/ in all languages. This feature has a phonetic base, being manifested in articulatory timing and/or phonatory power/tension, where power refers to articulatory tension and the strength of air stream during the production of sounds. Laryngeal power/tension refers to properties such as aspiration, voicing and glottalization, and may be language specific. The language universal factor is that of articulatory timing which, in this theory, deals with the speed of the formation and release of the stricture for obstruents.

The feature [± voice], in the theory, is thought inappropriate to express the /b, d, g etc./ versus /p, t, k etc./ contrast in that there is often lack of vocal fold vibration for both of these categories, a situation which contributes to the continuing confusion between phonetic and phonological voicing. This feature was thought to be more applicable in the phonetics domain, where it would represent low frequency periodicity, and the feature [± fortis] more appropriate for the phonology, where it would express phonemic contrast.

Basically, the feature [± fortis] would work in this way. For two-way contrasting languages, the difference between the [+ fortis] series and the [- fortis] series is described with reference to co-ordination between the oral, velopharyngeal and glottal valves. For the fortis series, a narrower/tighter stricture in the vocal tract has to be developed relatively faster than for the [- fortis] series. Also, the velopharyngeal closure has to be tighter for the fortis stops than for the lenis ones. The combined effect if this is that the [+ fortis] segments are produced with more intensity and are auditorily more salient than the [- fortis] series. The feature [± fortis] was also proposed to be of a gradient nature, so that different states on the scale can be assumed in various contexts and languages, particularly languages with more than a two-category distinction, e.g. Korean and Hindi.

For languages like Korean and Hindi, both the voiced and the voiceless unaspirated stops have a narrow glottal width opening before the release burst whereas their aspirated counterparts have a wide opening. The voiced aspirated stops have a narrow glottal width after the release burst, whereas their voiced cognates do not have any glottal opening after the release. According to Kohler's theory, the voiceless unaspirated stops and their aspirated counterparts belong to the fortis category, whereas the voiced stops and their aspirated cognates are lenis stops.
Shekgalagari stop consonants and Kohler (1984)'s fortis and lenis stops. In Kohler's framework of phonemic representation, Shekgalagari voiceless aspirated and unaspirated stops would be classified as [+ fortis], and the voiced stops as [-fortis] or lax. There is, therefore, an overlap between the voiceless unaspirated and aspirated stops. Although this overlap may be resolved by reference to timing relations between glottal events and supraglottal events and the size of the glottal widths, it still remain unclear how to represent the contrast between the two voice types in terms of Kohler's [± fortis] feature. The aspirated stops, as noted above, are realized by 'wide glottal opening with its maximum at the moment of release, resulting in substantial increase in airflow' (Kohler 1984:161). The unaspirated stops have a narrow glottal width opening before the stop is released. In the theory, similar stops in other languages, e.g. Hindi, were both classified as [+ fortis] (see table 3.2 (a)).

The main problem with Kohler's proposal for the presentation of phonemic distinction is the use of only one feature. The premise for this is that features like [voice], and glottal timing (VOT), which have otherwise been used by phonologists to represent phonemic distinction, belongs to the phonetics and should not be incorporated into phonological feature geometry. Inevitably, languages with more oppositions than the feature can account for end up with unresolved overlap. An example of this could be Hindi. As observed in section 3.3.5, although the voiceless unaspirated and aspirated stops may be differentiated in terms of glottal width, they are both categorized as [+ fortis]. A similarly explanation is given for the voiced and the voiced aspirated stop, which are [-fortis]. Phonetic descriptions are brought up to redeem the insufficiencies of the single feature, but since they are not permitted to express phonemic contrast, the theory remains in intense shortage of adequate distinctive features.

The theory also appears to be more suitable for languages with not more than a two-way distinction, and does not appear to be able to satisfactorily classify Shekgalagari stops.

7.4.2.6 The Element Theory

In the Element Theory approach, segments are differentiated by way of elements; namely: L, H and LH, where L = Low and H = High (Harris 1994: 133 – 135). The phonetic exponents used to express contrast are as follows. For the element L; slack vocal folds, long VOT lead and lowered fundamental frequency (voice bar). These features characterize truly voiced stops. Phonetic exponents for the element H are stiff vocal folds, raised fundamental frequency and long VOT lag. These
features describe voiceless aspirated stops. The LH specification has the exponents: slack vocal cords and raised fundamental frequency and represents the voiced aspirated stops, the so-called ‘murmured’ stops. The features describing voiceless unaspirated stops are short VOT lag. In the Element Theory, this category of stops is not specified. Element specification for the various linguistic systems has been summarized in table 3.3 (repeated here as table 7.1), adapted from (Monaka, Abberton & Harris 1997).

<table>
<thead>
<tr>
<th>Type</th>
<th>Language</th>
<th>short lag</th>
<th>long lead</th>
<th>long lag</th>
<th>breathy</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>German</td>
<td>/p/</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>II A</td>
<td>French</td>
<td>/p/</td>
<td>/b/</td>
<td>/pʰ/</td>
<td></td>
</tr>
<tr>
<td>IB</td>
<td>English</td>
<td>/b/</td>
<td></td>
<td>/pʰ/</td>
<td>/bʱ/</td>
</tr>
<tr>
<td>III</td>
<td>Thai</td>
<td>/p/</td>
<td>/b/</td>
<td>/pʰ/</td>
<td>/bʱ/</td>
</tr>
<tr>
<td>IV</td>
<td>Gujarati</td>
<td>/p/</td>
<td>/b/</td>
<td>/pʰ/</td>
<td>/bʱ/</td>
</tr>
</tbody>
</table>

Table 7.1: Element specification for some linguistic systems. The shaded areas indicate that the system does not have the relevant element specification.

As discussed earlier, the results obtained for Shekgalagari stops in word initial position showed long negative VOT values for the voiced stops. A voice bar in the low frequency regions was also observed on the spectrogram at the gap corresponding to the stop occlusion for these stop types. This may be regarded as the ‘lowered fundamental frequency’ phonetic exponent of the element. Although physiological investigation on vocal fold vibration for voiced stops in Shekgalagari still needs to be done, it is generally understood that voicing is produced with rather relaxed (slack) vocal fold configuration instead of stiff configuration. Where there was voicing decay, active voicing happened for most of the stop duration. This category of stops could be adequately be represented by the element L in the Element Theory.

Aspirated stops showed long positive VOT values. Again voiceless stops are understood to be produced with stiff (as opposed to slack) vocal folds configuration. In this study, no acoustic study was done on the fundamental frequency of the following vowel. Cross-linguistic analysis, however, has shown

---

21 That is, Southern German.
that voiceless aspirated stops have a raised fundamental frequency. This category of stops may be represented by the element H. The voiceless unaspirated stops showed a short positive VOT lag and are therefore unspecified.

The results obtained for Shekgalagari stops are compatible with the Element Theory of segment representation. Shekgalagari therefore falls into the same category of languages as Thai in Table 7.1, i.e. type III languages, with element specification L, H and unspecified element. Ejection, therefore, does not appear to be needed for Shekgalagari since other features can adequately account for the three-way contrast in the language’s stops, both phonetically and phonologically.

But it was observed from the spectrograms on Figure 4.1 for another category of voiceless stops, the unaspirated stops, that the first formant, F1, was observed to be rising from low frequency regions, where it was seen to be lodged for the voiced cognates (contrast with Figure 4.6). In the aspirated stops, F1 was always attenuated (see Figures 4.8 and 4.9). Thus formant frequencies for the voiceless stops appear to be higher than those of the voiced stops.
8 Conclusion

8.1 Introduction

8.1.1 The aim and objectives of the thesis, data and processing of the data

The purpose of the present study has been to investigate the voicing structure and articulatory characteristics of Shekgalagari stop consonants with particular reference to the Shengologa dialect (cf. Chapter Two). The study was primarily phonetic, although phonological aspects were also addressed, and was based on spectrographic, electrolaryngographic (ELG) and electropalatographic (EPG) techniques. ELG analysis provided useful and rather detailed information about the nature of vocal fold vibration which may not otherwise be obtained from speech pressure waveforms and spectrograms alone. EPG monitors lingual-palatal contacts in time by means of an artificial palate and is helpful in studying the nature of tongue to palate articulation involved in producing speech sounds. Simultaneous processing of the laryngograph waveforms: the Lx and the Gx (for gross larynx movement), the speech waveforms, the spectrogram and electropalatograms was performed in this study. For the acoustic study, (cf. Chapters 4 and 5), data comprised homorganic oral stops in the word initial CV context in a real word of the form CV:CV. For the EPG study (cf. Chapter 6), real meaningful words of the form CV\_CVCV:CVC produced by a native speaker of the language were used, where the highlighted C was the stop under investigation.

8.2 Summary of results obtained in the study

8.2.1 Electrolaryngographic and VOT investigation

The electrolaryngographic investigation revealed three main categories of waveforms from the data: voiced, voiceless unaspirated and voiceless aspirated stops. The voiced stops exhibited modal phonation during the occlusion of the stop into the vowel, and long negative VOT values. It was pointed out that these characteristics for these stops conformed to their traditional description as 'voiced' stops. The Gx signal showed distinctive downward movement for the voiced stops which appeared to be more in keeping with the voicing structure of these stops.

For the voiceless unaspirated stops (traditionally described as ejectives), a large proportion of the data was produced with modal phonation on the Lx signal
prior to the stop and at vowel onset (at least for the subjects used in this investigation). There was no cycle-by-cycle fluctuation in periodicity and intensity, no gradual build-up of intensity was observed, and a small, rather insignificant burst was a feature in most of the waveforms for these stops, with the burst not even observable in some of the waveforms. They showed short positive VOT values.

A few tokens were produced with irregular vocal fold vibration at vowel onset for one or two informants for the voiceless unaspirated stops in this study. We thought that this could most probably be due to speaker variation and/or token-to-token variation, which is not an unusual observation in studies like the present one. Nevertheless, it was proposed that, on the whole, the observations made for Shekgalagari voiceless unaspirated stops in this study and for the informants used in the investigation seem to suggest that these were simple plain voiceless unaspirated stops, and not ejectives.

Cross-linguistic investigation has revealed that ejection in stops shows, for instance, irregular vocal fold vibration prior to the target stop and/or at vowel onset, with period-by-period fluctuation in amplitude and intensity; long VOT values, e.g. 80+ ms for some North American languages and some Caucasian languages; low frequency perturbation (slow up and down deflection indicating decreased open quotient) in the Lx waveform at vowel onset; there could even be evidence of whole larynx movement. None of these was observed to any significant extent for the voiceless unaspirated stops in this study. For some of the tokens, the Gx signal appeared to remain relatively straight, possibly suggesting that there may be no indication of whole larynx movement due to the glottalic airstream mechanism. This, however, needs further intensive study.

The voiceless aspirated stops, showed turbulence after the burst and the vowel onset displayed breathiness. This category had long positive VOT values. It was concluded that this seems to conform to, as well as confirm, their traditional description as voiceless aspirated stops. For this and the unaspirated category, the Gx signal was contrastive to the distinctive downward movement it made for the voiced stops, possibly in keeping with their voiceless structure.

8.2.2 Electropalatographic study: articulatory and co-articulatory characteristics

The examination of articulatory characteristics of Shekgalagari stops provided qualitative and quantitative description on the type and extent of articulatory movements for Shekgalagari stops and for the variation observable between the
Chapter 8: Conclusion

voice types. Contact area and the percentage of on-electrodes were observed to be different for the different voice types. The ranking between the stops for these respectively was as follows; voiceless aspirated stops > voiceless unaspirated stops > voiced stops. This difference in contact pattern between the stop types implied slightly higher tongue position for the voiceless aspirated stops, followed by the voiceless unaspirated stops and lastly for the voiced ones.

Also, in terms of place of articulation, the palatal place showed the greatest area of contact, followed by the dental place, and lastly the velar place. It was deduced that for the palatal and velar stops, the tongue seems to be in a convex shape and remained so for a relatively longer period for palatal stops than for velar stops. It appeared that the tongue shape is concave for the dental stops in this study, with the back part of the tongue tending to be raised due to the influence of back vowels.

Coarticulation was also investigated by analysing lingual-palatal contact patterns for V-to-C co-articulation in symmetrical and asymmetrical vowel contexts. The stops showed some susceptibility to co-articulation, by voice type: the aspirated stops were less susceptible, followed by the voiceless unaspirated and lastly the voiced stops; and by place of articulation: the palatal stops appeared to show more resistance to co-articulation, followed by dental stops and lastly by velar stops. The back part of the tongue also seemed to accommodate more contextual influence than the front part. The vowel /i/ also seemed to be exerting more influence than the other vowels, particularly in the /i _ i/ context for all the three voice types within an articulatory place.

The findings of the experiment in this study were discussed in the light of constraints on the tongue during the production of sounds, theories proposed to account for co-articulation in speech and in the light of the findings reported in the literature for similar studies in other languages.

The feature-spreading model. In the present study, segments showed differences in the degree of co-articulation rather than the presence or absence of a feature, both within place of articulation and by voice type within articulatory type. Stops produced at the same place of articulation have exactly the same specification of tongue body features for place, and differ with regard to laryngeal features only. Yet stops produced at the same place of articulation manifested different degrees of co-articulation. Since these stops have exactly the same feature specification for place, nothing is left for the feature-spreading model to claim has spread onto them to account for the differing degrees of co-articulation that the stops exhibited.
It was also observed that different places of articulation manifest co-articulation to varying extents in this order: palatal < dental < velar. It is difficult to see how the feature-spreading model may explain this variation in terms of the spread of a feature, since this will have to occur across places of articulation. Variation in co-articulatory behaviour of segments appeared to be dependent on tongue contact area, and explainable in scalar terms rather than in terms of the presence or absence of binary phonological features.

With regard to voice type within an articulatory place, there was a weak trend for the voiceless aspirated stops to consistently be less accommodating to the influence of flanking vowels, followed by the unaspirated stops and lastly the voiced ones. Again co-articulation between the voice types seems to happen in degree rather than in terms of the presence or absence of a feature. There is, therefore, an incompatibility between the co-articulatory behaviour of the stops in this study and the feature-spreading account of co-articulation.

The window width theory. This model appeared to be compatible with the results obtained in this study, both for articulatory types and voice types, in that segments could be assigned different window widths depending on their susceptibility to co-articulation. Palatal stops could be assigned the narrowest window width, followed by dental stops with a medium sized window width and lastly the velar stops with the widest window width. Similarly, with regard to voice type within place of articulation, the voiceless aspirated stops could be assigned a narrowest window width, then the unaspirated stops and lastly the voiced stops with the widest window width of the three.

Co-articulation resistance theory. The results obtained in this study seem to support the co-articulatory resistance theory. With regard to articulatory place, the palatal stops, which showed the largest contact area compared with stops produced at other places of articulation, also showed the most resistance to contextual influence. The velar stops showed the smallest contact area and appeared to manifest less resistance to the influence of context. The dental stops came somewhere in the middle.

It was also pointed out that the overall tongue configuration could also be a possible contributing factor in studying the co-articulatory behaviour of segments. In this study, it was observed that the percentage on on-electrodes for the palatal stops began to rise from row 3 of the EPG palate, with full contact being made on row 7. This gave the general impression of a very constrained, rather convex tongue shape. For the palatal stops in particular, this observation
appeared to suggest that in addition to the contact area of a segment tongue configuration may be another dimension worth of consideration when studying the co-articulatory behaviour of Shekgalagari palatal stops. Velar stops also appeared to have a convex tongue shape too. The dental, however, appeared to have a concave tongue configuration.

The EPG study has revealed that Shekgalagari stops have different articulatory characteristics, and that this appears to be influential in the co-articulatory behaviour of the segments.

8.3 Suggestions for further research

8.3.1 Acoustics

The voicing structure of stops produced in word initial context (and in other environments for that matter), may be expressed along several phonetic dimensions; e.g. VOT, F0 frequency and perturbation, F1 frequency, closure duration, duration of preceding vowel, burst amplitude, activity of vocal folds during the closure phase, configuration of the vocal folds at or immediately after the release burst, the tensity of the articulators, height of the larynx etc. It is therefore necessary to investigate several phonetic properties when examining sounds in a language. This contributes to accurate description and classification of the natural classes in the language as well as helping to facilitate cross-linguistic comparison.

In this study, vocal fold activity during the production of Shekgalagari stops has been examined by means of ELG, and (positive and negative) VOT has also been measured. But the results obtained here could only be presented as tentative, with solid conclusions based on more informants and further investigation being hoped for. Future research could be done on, for instance, the other acoustic properties as well as physiological properties of the stops.

Nevertheless, it is hoped that the results obtained from VOT analysis would provide robust data for perceptual experiments in the future.

8.3.2 Articulation

The EPG investigation conducted in this study did not examine some aspects which are necessary to fairly examine the other models presented in this thesis.
namely the economy model and the co-production model (cf. § 3.6). Only the feature-spreading model, the window model of Keating and the co-articulation resistance model were examined.

For the assessment of the co-production model, for instance, further research may measure temporal aspects of the intervening stop: closure duration, time distance between V₁ and the stop, the stop and V₂ and V₁ and V₂ etc., in order to determine whether co-articulation is more sensitive to the time distance between segments irrespective of context or spatial aspects of the stops. In this study, only the characteristics of V₁-to-C co-articulation were examined. Future studies could investigated C-to-V₂ and V₁-to-V₂ co-articulation, as well as assess whether the direction of co-articulation could in any way be affected by temporal and spatial characteristics of the segments concerned.

No statistical analysis was done for the EPG investigation in this study. This is recommended for future research.
Appendices
Appendix 1A: for Chapter One - Introduction
Appendix 1A: For Chapter One – Introduction

A.1 Speech analysis

A.1.1 Techniques of speech analysis

A.1.1.1 The laryngeal component. The study of the laryngeal component in speech deals with the examination of the activity of the vocal folds in speech production. There have been a number of studies conducted with the aim of observing or monitoring larynx activity in speech and various devices and techniques of studying the behaviour of vocal folds have also been developed over the years. These devices include a laryngoscope, which allow the movements of vocal folds to be observed by using a mirror; a stroboscope, which produces frames of vocal fold vibration when its flash light frequency is adjusted to adapt as close as possible to the frequency of vocal fold vibration; a fibrescope, which is inserted down the pharynx through the nasal cavity and is suspended above the larynx; transillumination/photoglottography, where a light source is inserted into the pharynx via the nose and hangs above the larynx; and so on. A detailed description of the various direct and indirect methods of vocal fold examination can be found in, for instance, Abberton (1976), Fourcin and Abberton (1971) and the references therein, and also Borden, Harris and Raphael (1994). For the present purpose, it is sufficient to mention that these techniques were invasive (Fourcin & Abberton, 1971). There are, however, other techniques of monitoring vocal fold vibration which do not cause discomfort to the speaker. These include the ultrasonic and the electroglottograph methods (cf. Abberton 1976, and references therein). But perhaps the one whose use has become widespread, and which is more relevant for our study, is the electrolaryngograph (often referred to as the laryngograph), developed by Fourcin at University College London (UCL).

The laryngograph allows indirect observation of vocal fold activity during speech. It is a non-invasive device which monitors change in electrical impedance of the vocal fold by means of two gold-plated guard ring electrodes superficially applied to the skin on the sides of the neck on each wing of the thyroid cartilage (Fourcin, 1975; Fourcin & Abberton, 1971). It monitors the varying electrical conductance between the electrodes in terms of the current flowing between them, once contact between the vocal folds has been achieved. This contact is ‘an electrically-conducting mucosal bridge’ which allows the electric signal to be transmitted along vocal fold tissue (Abberton et al., 1989, Fourcin & Abberton
This contact completes the circuit, and allows the electric signal to be transmitted and detected by the electrodes. Laryngographic analysis thus gives information about the closed phase of a vibratory cycle, and in particular the area and form of the electrically-conducting surfaces of the vocal folds during adduction. Thus, the closed phase in a vibratory cycle is the only one which significantly contributes to the electrical output of the laryngograph. It follows therefore that any change in the area or nature of contact between the vocal folds is reflected in the electrical output, and also, vibration of the folds not involving contact (e.g. a form of breathy voice: Abberton et al., 1989) lacks an output.

The output waveform of the laryngograph (Lx), plots the varying electric flow between the electrodes as a function of time. In modal phonation, the different parts of a vibratory cycle can be related to a characteristic Lx cycle: There is the rapid closure of the vocal folds owing to the Bernoulli effect, starting from the bottom edges of the folds and moving upwards (Abberton et al. 1989). This corresponds to the steep rising portion on the Lx waveform and indicates increase in vocal fold contact. The positive peak represents the moment of maximum excitation of the vocal tract. Contact between the folds at this point is at its greatest, with minimal air from the subglottal system passing through the glottis.

Vocal fold opening occurs much more gradually. The edges of the vocal folds open from the bottom upwards 'as a function of subglottal pressure and the natural properties of the vibrating system comprising the elastic folds' (Abberton et al., 1989). This corresponds to the shallow falling edge on the Lx waveform and is associated with a decrease in vocal fold contact (Ibid.). The open phase represents a trough during which the vocal folds are out of contact. It is often assumed that the height of a trace on the Lx waveform is proportional to the area of contact between the vocal folds, i.e. it represents the amplitude (Lindsey, Hayward & Haruna 1992, Connell 1991). This knowledge of vocal fold activity and the corresponding laryngographic information form the basis for both operating the laryngograph device and interpreting the subsequent Lx waveform (Abberton et al. 1989). The four features of the Lx waveform corresponding to a single laryngeal cycle are shown in figure A.1; a hypothetical Lx waveform adapted from Abberton et al. (1989:285). These are the closing phase (I), maximum contact (II), opening phase (III), and open phase (IV).
The closed phase (CP) is believed to correspond to phases I through III together, since there is some certain amount of contact between the vocal folds in these parts of a cycle, and the open phase (OP) is believed to correspond to phase (IV) and involves no contact of the folds (Abberton et al. 1989).

Electric current conductance between the electrodes occurs with an increase in the area of vocal fold contact (cf. phase I). High-speed photography has confirmed that the Lx waveform deflects positively for increased vocal fold contact, with the steep commencement of the deflection in the waveform indicating the precise moment of closure onset, which is a function of the rapid Bernoulli effect (Ibid., Fourcin 1974). This decrease in impedance is plotted as a positive change on the vertical scale of the Lx waveform. Maximum conductance occurs at the peak of a vibratory cycle. This corresponds to the closure at the positive peak of the waveform (Abberton et al. 1989). An increase in electrical impedance corresponds to the more gradual fall in the trace. This is associated with vocal fold parting as the subglottal air pressure increases. The flatter base of the waveform corresponding to vocal fold separation is associated with a trough representing minimum or zero conductance across the abducted folds.

These distinctive parts of the Lx waveform 'provide an exceptionally accurate basis for the measurement of vocal fold vibration, period and frequency (Fourcin & Abberton, 1977). According to Abberton,
no complex processing of the acoustic signal is necessary to arrive at laryngeal features and the instrumentation is impervious to acoustic interference from the speaker's surroundings. The output thus obtained, Lx, has an extremely good signal-noise ratio, gives information in its own right, and can be processed to provide further information. (Abberton, 1976).

**Analysing the laryngeal component: Fundamental frequency (Fx).**

Fundamental frequency relates to excitation repetition rate derived from the vibrating vocal folds. Extracting the excitation component from acoustic information (cf. section A.1.1.2 below) to derive fundamental frequency is not an easy task, and the processing involved slow, given the complexity of the acoustic pressure waveform (Abberton 1976). Cepstral techniques which have attempted this (Noll 1967, cited in Ibid.) produced only average values for a single period of oscillation and, consequently, for the rate of vocal fold vibration (Abberton 1976; Fourcin & Abberton 1971; Fourcin 1974). The laryngograph output enables us to calculate the rate of excitation from the action of the vocal folds in a relatively simple manner and in real time. This frequency can be processed with considerable accuracy using each vibratory cycle since the Lx produces a very clear instant of vocal fold contact.

Fundamental frequency is inversely proportional to the period (Tx), which is the time between sequential periodic vocal fold closures. This is presented as follows:

\[ Fx = \frac{1}{Tx}, \text{ or } \log [Fx] = -\log [Tx] \]

Derivation of Fx on a period-by-period basis gives the true frequency of vocal fold excitation, given the fact that significant variation can occur from one vibratory cycle another in speech, and this period-to-period variation being perceptually significant (Fourcin *et al.* 1995).

Fundamental frequency of vibration is the major physical correlate of pitch. The prosodic systems of tone (e.g. Setswana and Chinese) and intonation (e.g. English) employ variation in pitch pattern in a linguistically significant

---

1 Formula representations for fundamental frequency, period, open and closed phases and open and closed quotient (Qx) used here are all obtained from Abberton *et al.* (1989).
manner. In Setswana, for instance, two different tone contours on the same word gives two different meanings to that word. In English, varying intonation patterns on a sentence can give it different statuses — e.g. that of a question or a statement, and play complex roles in discourse functions.

Fundamental period ($T_x$). This relates to the time distance between two regular epochs of excitation (Fourcin et al. 1995). This is illustrated in figure 2.14, which shows hypothetical $L_x$ waveforms (adapted from Lindsey, Hayward & Haruna (1992:513)) showing fundamental period of differing duration.

---

**Fig. A.2**: Hypothetical $L_x$ waveforms showing fundamental period of differing durations, short period on (a) and longer period on (b). Adapted from Lindsey, Hayward and Haruna (1992:513).
Open and closed phases. The duration of the open and closed phases of the vocal folds can be expressed in various ways. They can be expressed as a ratio:

a. (i) the open to closed phase ratio: \( \frac{OP}{CP} \)
(ii) the close to open phase ratio: \( \frac{CP}{OP} \)

They can also be expressed as a quotient which is the percentage of the fundamental period for which the open or closed phase exists: \( Q_x \)

b. (i) open quotient \( Q_O = \left( \frac{OP}{T_x} \right) \times 100 \) %
(ii) close quotient \( Q_C = \left( \frac{CP}{T_x} \right) \times 100 \) %

\( Q_x \) can be used as an indicator of voice quality and is linked to spectral characteristics in the spectrogram. Rapidity of vocal fold closure and regularity of vocal tract excitation provide further information on voice quality. \( T_x \) values can be processed further to provide further information on voice production and pitch range. This can be in the form of first-, second- and third-order distribution probability histograms (\( D_x \)) and the \( C_x \) plot scattergrams.

Intensity. The intensity of the acoustic excitation can also be calculated by the \( L_x \) software.

The form of \( G_x \). \( G_x \) is the output of a particular kind of laryngograph adjusted to respond to gross movements of the larynx; it principally provides information on laryngeal adjustments before, during, and after phonation. However, other activities not directly linked with phonation, like swallowing, or moving the head or neck can contribute to \( G_x \) output. Initial vocal fold adduction is associated with a small percussive positive peak in the \( L_x \). All these adjustments are reflected in the \( L_x \) as low-frequency baseline movements (Abberton et al., 1989).

Laryngographic analysis thus gives access to laryngeal activity without being invasive, and the information derived from the laryngeal components of speech, rather than from the acoustic signal, gives more accurate qualitative and quantitative analysis of the laryngeal function in speech production (Abberton 1976). The laryngograph, however, does not provide information about variation in the glottal area when the vocal folds are out of contact; neither does it indicate which parts of the folds are in contact during adduction.
The Lx waveform and the acoustic waveform (Sp). The Lx waveform has a consistent relationship with the speech pressure waveform (discussed immediately below in section A.1.1.2) which can also be produced simultaneously with it. They are often time aligned 'so that the vocal fold closure corresponds to the acoustic epoch of excitation in the speech pressure waveform' (Fourcin et al., 1977). The typically rapid closure of the vocal folds constitutes an epoch of vocal tract excitation and is subsequently followed by an onset of maximum acoustic activity which happens during the closure phase (Ibid., Abberton et al. 1989). The availability of the laryngograph output and the simultaneously produced acoustic signal has been fundamental in analysing and processing many physical aspects of the production of voice in a more accurate and reliable way than has been possible with the acoustic signal only (Fourcin et al. 1995, Abberton et al. 1989).

A.1.1.2 The acoustic component: spectrography. Spectrography, developed in the 1940s, has been a useful device for the study of acoustic signals (Kent & Read 1992). One of the advantages which spectrography introduced in speech analysis was the development of a running short-term spectrum producing spectrographic displays called spectrograms. Spectrograms made it possible to make visual analysis of aspects of speech such as amplitude\(^2\) (i.e. intensity of energy in an acoustic signal\(^3\)) and frequency of vocal fold vibration. The amplitude is determined on the grey scale (i.e. the darkness or brightness of the pattern). The more intense portions of the signal are registered as darker patterns on the spectrograms compared with the ones with less energy. Analysis of the frequency components of the speech signal is conveyed in terms of the harmonics it consists of or the resonances it incorporates (Borden, Harris & Raphael 1994:239). Fundamental frequency, which is one of the principal goals of speech analysis, is presented differently in the two types of spectrograms: the narrowband and the wideband, because of the different filters used to generate these spectrograms.

The bandwidths of filters used to produce narrowband spectrograms are of the order of 30 Hz to 50 Hz (Borden, Harris & Raphael 1994). Consequently, they resonate to a small number of frequencies, responding to each harmonic separately in the speech signal. Continued response in the filter is maintained because of light damping due to the nature of the band, which allows only a limited number of frequencies to pass through. This means that by the time each

---

\(^2\) Spectrograms delineate change of energy concentration (amplitude) with respect to time (X-axis) and frequency (Y-axis) (Kent & Read 1992).

\(^3\) This is the product of the sound source and the resonating filter (Kent & Read 1992:18).
glottal pulse reaches the resonator, the resonator itself will still be analysing and resonating to the former one. This continued response is what underlines the equi-distanced unbroken horizontal bands associated with the fundamental and its harmonics in narrowband spectrograms, with each harmonic being an integer multiple of the fundamental frequency. Changes in the fundamental frequency itself are usually not perceptible. This makes it less easy to quantify the fundamental frequency by analysing the fundamental itself. Higher frequencies delineate these changes more saliently, and are normally used to quantify the fundamental frequency by measuring one of the higher harmonics at a particular point in time (the 10th or a higher harmonic; cf. Borden, Harris & Raphael (1994:241)) and dividing its frequency by the number of the harmonic. Narrowband spectrograms, therefore, have good frequency resolution and are suitable for the determination of fundamental frequency. However, current devices can scale variation in fundamental frequency as a function of time from automated measurements in a more reliable and accurate manner, eliminating the manual calculations of fundamental frequency by way of higher frequency harmonics.

Filters used to generate wideband spectrograms generally have a bandwidth of between 300 Hz and 500 Hz (Borden, Harris & Raphael 1994:424). Unlike narrowband filters, these do not resolve the energy within their range into individual, separate harmonics, but tend to collapse and resonate in an identical manner to up to three harmonic frequencies that fall within their limits. This translates into large bands of energy corresponding to formants, with the centre of each band representing the formant frequency, and the gamut of frequencies covered by the band corresponding to the bandwidth of the formant (Ibid. 241). Thus individual harmonics are not depicted separately in wideband spectrograms, as is the case with narrowband spectrograms.4

Broadband filters are also good at depicting dynamic variation in the energy of the acoustic signal as the vocal tract is excited by the glottal output. This information is more reliably and clearly depicted in wideband spectrograms. This includes, for instance the effective registration of aperiodic excitation originating from a random source as a region of noise energy, ‘mostly at the frequencies of the second and third formants of contiguous pattern segments’ (Lisker & Abramson 1964), with the attenuation of the first formant often visible

---

4 According to Borden, Harris and Raphael (1994: 242), ‘this lack of specificity (i.e. of individual harmonics)... is what the investigator wants, because the formants that characterize a particular sound for a particular speaker will be essentially the same, no matter what the frequencies of the fundamental of the harmonics — or, indeed, even if there were no fundamental/harmonic frequencies at all, as when sounds are whispered.’
(Lieberman et al. 1958), brief energy corresponding to the plosive release of a consonant, gaps associated with stop occlusions, conspicuous and long bands of energy associated with vowels, and so on.

Wideband filters also have good temporal resolution. They respond intermittently and rapidly to onset and offset of glottal output, resonating to each pulse separately. This is represented as thin vertical striations on the spectrograms, each denoting a single glottal pulse, and each being darkest (within the bandwidth of a formant) at the peaks of resonance and lighter or discontinuous in frequency regions which are less intense. Because of the rapid response of broadband filters to glottal excitation, broadband displays have often been used for making temporal measurements, such as the duration of an acoustic segment and voice onset time. This task has now been taken over by modern devices which have automated measurements.

Thus in speech analysis, narrowband spectrograms have been crucial for the determination of harmonic frequencies, and hence the fundamental frequency, and wideband spectrograms, because of their excellent depiction of the various dynamic changes in the resonances of the vocal tract, have been used to make temporal measurements. It is worth noting that spectrograms are a product of a sound source, either at the larynx or the mouth, and the acoustic filter, which is almost always the oral cavity. The narrowband and wideband spectrograms are illustrated in figure A.3, adapted from Borden, Harris and Raphael (1994:240).
A.1.1.3 Electropalatography (EPG)

EPG is a non-invasive\(^5\) technique used for examining contact patterns between the tongue and the hard palate as well as recording the timing of these contacts in speech. There are two types of EPG systems: the Reading system and the Japanese Rion system (which will not be discussed further because it is not used in this study).

EPG technique in general uses a made-to-fit, thin artificial palate made from acrylic, with built-in (gold or silver) electrodes exposed to the surface of the tongue as contacts. The Reading system in particular uses 62 silver electrodes (see Hardcastle et al. 1989), prearranged in eight horizontal rows on the basis of the

\[^5\] The EPG palate may introduce discomfort to the speaker beyond the junction between the hard and soft palate. For this reason it may be considered to be partially non-invasive.
anatomical divisions of the hard palate. The front row carries six electrodes, and the rest eight. Consider the illustration given in figure A.4 (from Hardcastle et al. 1989: 3).

![Figure A.4: The Reading electrode embedded artificial palate and plaster model showing the positions of the electrodes (Hardcastle 1989:3).](image)

These electrodes conduct electricity when contact between the tongue and the palate occurs. They thus detect lingual-palatal contact. In speech analysis, this makes it necessary for the sounds being studied to have considerable (or measurable cf. Ibid.) amount of tongue-to-palate contact. For example, [t, d, k, g, s, z, z, 3, tʃ, dʒ, n, ɾ, j, ɾ, ɾ, ɪ, ɪ, ɛ] where the sides of the tongue make contact with the palate for non-open vowels. The signals resulting from lingual-palatal contacts are then electrically processed, displayed on a monitor (or printed out) for inspection or they could be saved on a PC for future analysis. The displays (or printouts) appear in the form of stylised frames of the artificial palate and show details of how lingual-palatal contacts vary in time (ms), with a specific time interval (10 ms) between two adjacent frames. EPG frames provide detailed information about spatio-temporal properties of sounds which could otherwise be difficult to extract from acoustic information alone (cf. Hardcastle et al. 1989: 13). This includes, for instance, qualitative analysis of the extent of frontness of an articulation — determined on the basis of the number of the most front row of the frame on which contact is registered, and of the degree of closeness of an articulation.

---

6 It must be mentioned that these divisions of the palate, far from being standard, actually differ from author to author.
articulation — the sum of contacts in any given row(s) (*cf.* Ibid.). Simultaneous processing is especially crucial when synchronising the activity of the tongue in the mouth with that of other articulatory or phonatory activities elsewhere, for instance, in the larynx.

Like all analytical instruments, it must be mentioned that there are shortcoming involved in using the EPG technique. These include, for instance, the inability of the technique to provide information about the part of the tongue involved in a particular contact pattern, the impossibility of assuming continued closure where two juxtaposed electrodes register contact, lack of and only partial accessibility to post-velar and velar articulations respectively and the possibility of the technique interfering with the natural production of speech.

EPG only records a signal indicating lingual-palatal contact. It does not give information with regard to the part of the tongue responsible for producing a particular contact pattern. This information has to be extrapolated from the place and timing of contact accompanied by the knowledge of the structure of the tongue. For some parts of the palate, the reliability of deducing information in this way may be reduced considerably. For example, when contact is made with a point forward of the alveolar ridge, it is not easy to determine whether it was made with the front or tip of the tongue.

Another shortcoming relates to the fact that electrodes on the artificial palate are isolated and distanced from each other — making it hard to interpret whether contact in two adjacent electrodes indicate unbroken tongue-palate contact. An example may be, for instance, the lateral approximant /l/ for which two neighbouring electrodes at the side of the palate may register contact, though the articulatory closure may, in fact, be incomplete.

The artificial palate can only go back as far as the junction between the hard and soft palates. Beyond that it introduces discomfort. This means that the technique is not accessible to articulation posterior to the velar place. On the one hand, sounds produced in the region posterior to the velar place, e.g. uvular sounds, simply cannot be analysed using EPG. On the other hand, velar sounds can be investigated. But this investigation ‘may be incomplete, and analysis of results of these articulations must take this factor into account’ (*Connell* 1991: 74).

The other point relates to the possibility of the technique interfering with the natural production of speech, since it involves the attachment of a foreign object to an articulator. In fact, subjects tend to salivate excessively in reaction to this, which in turn affects the acoustic properties of sounds being studied. In order to offset this problem, informants are often required to wear the palate for some
time (between 2 and 4 hours) before the recording to familiarise them with a foreign object in the mouth. In order to detect whether speech has in any way been affected, control measures are often taken, and these involve spectral analysis of data with and without the artificial palate (cf. Connell 1991, Hardcastle et al. 1989).

One way in which analysts attempt to overcome the above mentioned shortcomings is to do simultaneous analysis of data in tandem with EPG. For instance, processing the laryngeal component (electrolaryngography - cf. Section A.1.1.1 above), analysing the acoustic signal (cf. section A.1.1.2 above) and conducting aerodynamic investigation. In this way, ‘a more comprehensible picture of articulatory activity than is possible with EPG alone’ (Hardcastle et al. 1989) can be built up. Such attempts have been made by, for instance, Connell (1991). See also Hardcastle et al. (1989) and the references listed there.

The EPG technique is thus useful in the study of the spatio-temporal dynamics of tongue-palatal contacts in the mouth, and can provide detailed information on this aspect that would not otherwise be obtained through acoustic investigation only.

To sum up, the laryngograph enables analysis of the activity of the larynx and is not invasive. Laryngographic investigation of the laryngeal component of speech is more reliable than information derived from spectrographic analysis, which has been used in the past to provide qualitative and quantitative analysis of speech sounds. EPG gives details of contact patterns between the tongue and the palate which could not be obtained any other way.
Appendix 2B: for Chapter Two – Shekgalagari; genetic classification and linguistic issues
Appendix 2B: For Chapter Two – Shekgalagari, genetic classification and linguistic issues

B.1 Southern Bantu Languages: Geographical Considerations

Following Doke (1954), southern Bantu languages may be divided into the Southwestern zone, the South-central zone and the South-eastern zone. The Southwestern zone belongs to the western Bantu region incorporating Namibia and Angola and are classified as ‘Southern’ mainly from the point of view of geography (i.e. southern vs. eastern). The South-central zone comprises the Shona and Ndebele groups of Zimbabwe, and the languages of Northern Mozambique and Zambia. The South-eastern zone includes the Republic of South-Africa, Southern Mozambique and the three former British Protectorates of Botswana, Lesotho and Swaziland.

B.2 The South-eastern zone; Language groupings and clusters

An analysis of the divisions of language groupings and clusters within the various zones would go beyond the scope of this study. But the South-eastern zone, which is more relevant for our purpose, is divided into the following groups and clusters (Doke, 1954).

---

It is worth pointing out that this classification of Shona and Ndebele groups of languages is based on very early research on these languages. Nowadays the Ndebele language is thought to be more likely affiliated to the Zulu language spoken mostly in South Africa, since the two languages are mutually intelligible. It is therefore probable that Zulu belongs to the South-eastern group of languages, and not the South-central.

Owing to constant migrations and the introduction of boundaries by colonial masters, it is the case that languages belonging to a zone or a group could be found living within a different zone/group. For instance, some members of the Shiyeyi and Sesubia languages, though belonging to the Western and Central zones, are living in Botswana, which belongs to the South-eastern zone. In some cases the languages can become influenced away from members of their grouping so much that their classification is not quite clear. Examples of such languages include the Karanga of Mhari, the Tavara and Budya of Korekore and the ‘Shekgalagari of Setswana’ (Doke 1954).
As can be seen on the diagram, the Sotho group is further divided into the Southern Sotho, Northern Sotho and Tswana clusters. The Tswana cluster is further divided into various dialects which are still spoken in Botswana as follows.

It is evident from the above divisions that Shekgalagari is not incorporated in Doke (1954)'s classification of the clusters and dialects of the Sotho group. Rather, concerning this language which is spoken in Botswana, C. M. Doke (1954) writes: ‘Kgalagadi is sufficiently distinct from Tswana to warrant a separate classification’ (Doke 1954: 24).
B.3 Some phonetic and phonological aspects of South-eastern Bantu languages

B.3.1 The vowels. Southern Bantu languages' vowel system falls into three groups: those displaying five vowel inventory system and five allophones, e.g. Nguni; those with five vowel phonemes and seven allophones, e.g. Shona; and those with seven phonemes and nine allophones, e.g. Sotho (cf. Doke 1954). These are respectively illustrated below.

(3) a. Five phonemes and five allophones e.g. *Nguni*

\[
i \quad u
\]
\[
e \quad o
\]
\[
a
\]

b. Five phonemes and seven allophones e.g. *Shona*

\[
i \quad u
\]
\[
\{ e \quad o \}
\]
\[
a
\]

c. Seven phonemes and nine allophones e.g. Sotho

\[
\hat{r} \quad \hat{u}
\]
\[
i \quad u
\]
\[
\{ e \quad o \}
\]
\[
a
\]

It can be observed from the above illustrations that Southern Bantu languages have one low vowel /a/, and an equal number of front and back vowel counterparts. There are no centre vowels. For those systems displaying the types (b) and (c) inventories, the mid front and back vowels /e/ and /o/ each constitute a single phoneme, each with two allophones: the raised and lowered versions, the choice of which, by rule, is determined by the quality of the vowel in the
subsequent syllable. Basically, a close vowel triggers the raised variant, otherwise the lowered variant is used. Typically, Southern Bantu has no diphthongs. Successive vocalic segments in words like *mae* ‘eggs’ are, in fact, two successive syllables with empty onset positions.

**B.3.2 The consonants: stops.** Doke (1954:30) gives the following taxonomy for stops in Southern Bantu.

(4) Stops

<table>
<thead>
<tr>
<th></th>
<th>Bilabial</th>
<th>Dental</th>
<th>Alveolar</th>
<th>Retroflex</th>
<th>Palatal</th>
<th>Velar</th>
</tr>
</thead>
<tbody>
<tr>
<td>Radical</td>
<td>p</td>
<td>t</td>
<td></td>
<td></td>
<td>k</td>
<td></td>
</tr>
<tr>
<td>Ejectives</td>
<td>p’</td>
<td>t’</td>
<td>t’</td>
<td>c’</td>
<td>k’</td>
<td></td>
</tr>
<tr>
<td>Asp.</td>
<td>p’h’</td>
<td>t’h’</td>
<td>t’h’</td>
<td>c’h’</td>
<td>k’h’</td>
<td></td>
</tr>
<tr>
<td>Voiced</td>
<td>b</td>
<td>d</td>
<td>d</td>
<td>j</td>
<td>g</td>
<td></td>
</tr>
</tbody>
</table>

Figure B.1 shows the traditional complexion of Southern Bantu stop consonants, indicating four main types. These are as follows: the voiceless forms, namely; the radicals and the ejected types (i.e. the so-called ‘ejectives’) where ejected stops are considered ‘a typical South-eastern feature’ (Doke, 1954:31), and are frequently represented in the phonemic inventories of South-eastern languages of the Sotho cluster (e.g. Pai, Pulana and Kutswe; Ibid.) to which Shekgalagari belongs; the aspirated stops; and the voiced stops.

These so-called ‘ejectives’ are supposedly produced on a laryngeal airstream mechanism which entails simultaneous constriction in the glottis and

---

2. Radical stops are sometimes referred to as plain stops. They are found in, for instance, Shona, a South-central Bantu language, but not in South-eastern Bantu, although the position with regards to Shekgalagari is not clear (cf. Doke 1954:31).

3. However, countering this view, Snyman (1991) argues that the so-called ‘ejectives’ are more commonly produced by means of the pulmonic airstream mechanism, thereby losing their ‘ejective’ status. Ladefoged and Maddieson (1996) also comment that ‘the presence or absence of a glottalic mechanism is often a matter of degree’. These arguments raise interesting questions about the precise nature of the production of the so-called ‘ejectives’ in southern and/or South-eastern Bantu languages. An acoustic analysis of the so-called ‘ejectives’ in Shekgalagari, a South-eastern Bantu language, forms one of the bases of this study.
oral cavity. The whole larynx is then abruptly hoisted up, compressing the air trapped in the chambers above the larynx, and rendering a characteristic explosive effect upon release (Doke 1954:31). The plain counterparts (i.e. the radicals) lack this particular characteristic in their production and are therefore not called 'ejectives'.

Aspirated sounds are easily identified by an audible rush of air after the release, which could be derived from a random source in the larynx or from a site in the vocal tract. Voiced sounds are identified by vocal fold vibration in the larynx.

B.4 The classification of Shekgalagari

Shekgalagari belongs to the South-eastern group of Bantu languages and is often classed with the Sotho-Tswana cluster. Most works in the past have linked Sekgalagadi to Sesotho and Setswana and described it as a dialect of both of these languages. For example, it has been described as one of the mixed dialects of Sesotho, along with ‘extreme Sotho types’ of Pai and Lobedu, displaying elements of what might be considered to be proto-Sotho (Doke 1954). It has also been classified as a dialect of Setswana (Cole 1955; Doke 1954). Cole (1955), for instance, describes it as ‘definitely belonging to Tswana’. There is no evidence given to support any of these views; which would have perhaps shed light on the grounds for this classification. In spite of this classification, however, there seems to be an awareness (noted by both Doke 1954 and Cole 1955) that Shekgalagari displays phonetic and grammatical features which are sufficiently distinct from both Sesotho and Setswana literary forms to warrant a separate classification. In keeping with this view, Cole (1955) says that Shekgalagari has ‘developed or become influenced away from the norm in such a way that it tends to stand apart from Tswana’. Shekgalagari is the first language of the Bakgalagari people, who use Setswana as a second language. In Botswana, this language is spoken in the region between Lake Ngami near the Okavango delta and Lokhwabe in the Kgalagadi district. The dialects of this language include Shengologa, Shelala, Sheboloongwe, Sheshaga, Shekoma, Shekue and Shekhana.

Shekgalagari is spoken by the Bakgalagari people who live mainly in the Kgalagadi — this is the area around Hukuntsi on the map (cf. Figure 2.1, pg. 28) of the thesis—and Ghanzi districts, the west part of Southern and Kweneng districts, and around the Ngamiland (Maun) area. The present population of Bakgalagari in Botswana is not known. Andersson and Jason (1997) estimate that their number could possibly be between 10 000 and 15 000, and add; ‘but this
estimate is highly uncertain' (Andersson & Jason 1997). Almost all Bakgalagari know and can speak Setswana, but Shekgalagari is marginally, if at all, intelligible to speakers of Setswana, or of any other language spoken in the country for that matter.

B.5 Previous studies on Shekgalagari
Most of the previous studies on Shekgalagari have been theoretical analysis into some of the dialects of the language. These will be mentioned briefly by researcher.

B.5.1 Van der Merwe and Schapera (1943)
The first work to be published on Shekgalagari was a study conducted by van der Merwe and Schapera (1943). This was a side-by-side study of the Sekwena dialect of Setswana and the Sheboloongwe dialect of Shekgalagari, making comparative references to other Sotho-Tswana languages. An outline of their phonemic inventory records Seboloongwe as a three-way contrasting system: namely, the voiced series, the voiceless aspirated series and the voiceless unaspirated series. They record only one ejective, which they transcribe as [k'].

B.5.2 Doke (1954)
Doke (1954), based on an earlier work by Merwe and Schapera (1943), analysed the classification of Shekgalagari as one of the Bantu languages and outlined its consonantal system, making comparisons with the other Sotho-Tswana consonantal inventories. His consonant system shows that the consonants are produced at four places of articulation: namely, bilabial, alveolar, palatal and velar; it also indicates that it is basically a three-way distinctive system with voiced sounds, voiceless aspirated and voiceless unaspirated sounds. He also records one ejective [k'], following Merwe and Schapera (1943).

B.5.3 Du Plessis and Kruger (1968)
Du Plessis and Kruger (1968) investigated the consonant system of the Boloongwe dialect. Their phonemic inventory shows that the stop consonants in this dialect are produced at five places of articulation: namely, bilabial, alveolar, palatal velar, and uvular. Sheboloongwe is shown to be a three-way contrasting system with voiced stops, voiceless aspirated and voiceless unaspirated sounds.

Dickens (below) argues rightly that the realization of this sound has been transcribed incorrectly as [k'] for [q'].

---

4 Dickens (below) argues rightly that the realization of this sound has been transcribed incorrectly as [k'] for [q'].
B.5.4 Dickens (1987)

A attempted experimental investigation was carried out by Dickens (1987) on the Sheshaga, Shekoma and Sheboloongwe dialects with few recordings made on tape. Most of the data were gathered by auditory impressionistic methods and direct visual observation of the places of articulation, tokens being transcribed as they were produced. Palatography was also used to aid in the phonetic analysis. His phonemic taxonomy shows that stops in these dialects may be found at five articulatory places: that is, bilabial, dental, palatal, velar and uvular. It also indicates that they display a three-way distinction: namely, voiced, voiceless aspirated and voiceless unaspirated stops which were 'always produced pulmonically except on a few occasions when "ejectives" were used metalinguistically' - for instance, emphasizing with an ejection when a voiceless unaspirated sound was incorrectly imitated as aspirated. He goes on to say that a hypothesis with regard to 'ejectives' is that they may be stored mentally as 'ejectives' but only realized as such when particular attention is paid to the sound during articulation.
Appendix 3C: For Chapter Three ~
Stops: VOT and $F_o$ perturbation, phonological representation and issues of co-articulation
C.1 The glottalic airstream mechanism

The larynx is the initiator of the egressive and ingressive glottalic\(^1\) airstream mechanisms. In this type of initiation, the glottis is closed so that only the air above the glottis is used to generate sounds. A constriction is simultaneously formed in the chambers above the glottis.\(^2\) The velum is raised to block airflow through the nasal cavity. And the relevant sounds are produced on the trapped air between the glottal and the supraglottal constrictions. Initiation involves abrupt vertical displacement of the whole larynx, initiating airflow; upwards in the egressive mechanism and downwards in the ingressive mechanism, and since pulmonic airflow is not involved, the whole impression is that of the larynx acting as a piston in the pharynx (see figure C.1 (a) and (b) for illustrations).

---

\(^1\) Also known as laryngeal or pharyngeal (Pike, 1943: 90).

\(^2\) The constriction in the supraglottal cavities can either be complete - as in the articulation of a stop, or partial - as in the production of a fricative. When the complete closure is followed by a fricative portion during the release phase, ejective affricates are produced.
In the egressive mechanism, the glottis is always closed, so that the sounds produced are always voiceless (although VOT variations are in principle possible). In the ingressive mechanism, however, vibration of the vocal cords occurs most of the time. Voiceless ingressive sounds produced on this mechanism are very rare.

Ejectives are articulated on the egressive laryngeal pressure initiation which entails simultaneous constriction in the glottis and oral cavity. The whole larynx is then rapidly raised up, compressing the air trapped in the supralaryngeal cavities, and forcing it out of the mouth. In the production of plosives, this results in a characteristic explosive effect upon release. Ejectives are a feature of many African languages. In K’ekchi, a Quichean language spoken in Guatemala (Pinkerton 1986), contrast between voiceless egressive pulmonic plosives and their glottalic counterparts is illustrated in the following words (Pinkerton 1986: 130 cited in Laver 1994):

(1) [t’oqok] ‘to throw’  [toqok] ‘to break’
    [fɔ:ʃ’ɔk] ‘you threw it’  [fɔ:ʃtok] ‘you broke it’

---

According to Ladefoged and Maddieson (1996: 78), the degree of pressure generated behind the occlusion is two times that of the normal pulmonic pressure.
Sounds generated by the rapid downward movement of the larynx are commonly called implosives. This gesture increases the volume of the supraglottal cavities, hence the air trapped between the glottal constriction and the closure in the mouth is thus rarefied in pressure (Laver 1990: 173), and when the oral stricture is released, there is an inrush of atmospheric air to fill the relative vacuum (Ibid.). There are two initiators involved in the production of these sounds: namely, the respiratory system and the larynx. Thus airflow is a combination of the ingressive glottalic mechanism — this effectively produces voiceless implosives when it is the only type of airflow mechanism involved — and the egressive pulmonic mechanism — thus producing voiced implosives. These are more commonly used in languages than voiceless implosives. Consider the following examples (The Tojolabal and Cakchiquel examples were adopted from Laver 1990, and the Shona examples from Doke 1954):

(2)

Voiceless implosives: Tojolabal  
\[\text{[bof]}\] ‘to be able’
\[\text{[p\textsuperscript{b}op\textsuperscript{b}]}\] ‘straw mat’

Cakchiquel  
\['\text{q\textdagger}\text{olone}\text{t}'] \text{‘Saviour’}
\['\text{q\textdagger}\text{olone}\text{t}'] \text{‘a stripper of bark’}

The most important characteristic of these sounds is the movement (and direction) of the larynx, since it is often the case that both lung air and glottalic airflow mechanism are employed in their production. Ashby (1990), describes the various articulatory possibilities for the production of implosives, which, although vocally possible, do not appear to be employed linguistically.

Lung air escaping through the approximated vocal cords in the production of voiced implosives can be sufficient to ‘offset the suction action of the downward larynx movement so that there is little or no inward airflow through the mouth, even to such an extent that the net airflow is actually egressive’ (Clark & Yallop 1990:57, also Ladefoged 1971, and Ladefoged & Maddieson 1996).

The occurrence of these sounds in the languages of the world are comparatively rare (i.e. when compared with the voiced implosives). They have been reported in Tojolabal in Mexico and Cakchiquel in Guatemala, where they are phonologically distinctive (Laver 1994: 173).

During the downward movement of the larynx, which initiates the ingressive laryngeal airstream, the vocal folds are often separated, allowing pulmonic air to escape and causing phonation in the process.

As in the case of ejectives, the English system does not employ implosives contrastively, but it can sometimes make allophonic use of them, especially the energetic production of stops ‘as in “absolutely billions and billions”’ (Ladefoged 1993: 133).

[\text{q}] is a voiceless uvular implosive produced with a constriction made at the glottis and between the uvula and the back of the tongue.
Voiced implosives: Shona

[dededza] ‘bite’
[dededza] ‘toddle’

Hausa
[bardoo] ‘long tailed dove’
[bardoo] ‘Fulani man’
Appendix 4D: For Chapter Four ~
Acoustics: Description and classification of speech and electrolaryngograph waveforms
Appendix 4D: For chapter Four ~ Acoustics: Description and classification of speech and electrolaryngograph waveforms

Shekgalagari data

Bilabial

1. Voiceless unaspirated stops
   ぴ:na ‘song’
   ぴ:mo ‘sight’
   ぴ:ri ‘goat’
   ぴ:xa ‘hang’
   ぴ:la ‘refuse’

2. Voiceless aspirated stops
   ぴ:*i:ri ‘hyena’
   ぴ:*l:o ‘state of being healed’
   ぴ:*e:l:o ‘wind’
   ぴ:*a:k:á ‘receive’
   ぴ:*u:c:á ‘pack’

3. Voiced
   び:na ‘dance’
   べ:pá ‘bellow’
   ば:lá ‘read’
   ぼ:na ‘see’
   ぶ:ta ‘break’

Dental

1. Voiceless unaspirated stops
   ち:pa ‘to overload’
   ち:mo ‘dam’
   ち:qa ‘find, lost property’
   つ:ra ‘dear, expensive’
2. Voiceless aspirated stops
th'ě:ra  ‘stop suddenly’
th'ě:pe  ‘water tap’
th'á:ma  ‘bind’
th'ű:je  ‘wild rat’

3. Voiced
de:pá  ‘take in large quantity’
de:lá  ‘bring for’
dá:la  ‘be full’
dó:mo  ‘stupid, dull’
duri  ‘ignorant’

Palatal
1. Voiceless unaspirated stops
cú:la  ‘to duck’
cá:ca  ‘to not welcome’
cómá  ‘fix firmly, especially on the ground’
cú:ba  ‘to hit’

2. Voiceless aspirated stops
e'ň:na  ‘dirt, dirty’
e'ň:ma  ‘to get stuck’
e'ň:xa  ‘type of bird’
e'ň:ibá  ‘steal away’
e'ň:upa  ‘stick’

3. Voiced
jú:sa  ‘make to eat’
jé:la  ‘eat for, eat someone’s food’
já:qa  ‘sojourn’
jó:ba  ‘Job’
jó:ko
Velar

1. Voiceless unaspirated stops
ki:ka ‘mortar’
ke:mo ‘wedding’
ká:ma ‘fill’
Kö:ma ‘traditional song’
kú:ri ‘tick’

2. Voiceless aspirated stops
khí:ba ‘apron’
kʰê:me ‘wild melon’
kʰá:la ‘keep quiet’
kʰú:ba ‘type of bird’
kʰú:mo ‘wealth’

3. Voiced
gi:ri ‘stop suddenly’
qa:ná ‘be arrogantly uncooperative’
go:ba ‘grievous toil’
go:go ‘take in large numbers’

Uvular

1. Voiceless unaspirated stops
qí:di ‘onomatopoeic for noisy swallowing’
qé:da ‘I have arrived’
qá:la ‘branch’
qá:ma ‘act of chewing the bone’
qú:du ‘onomatopoeic for noisy swallowing’
Xhosa data

**Bilabial**

1. **Voiceless unaspirated stops**
   - *pópó:la* ‘to examine something thoroughly, e.g. before buying’
   - *penápe:na* ‘throb with pain’
   - *pá:pa* ‘porridge/ mealie-meal’

2. **Voiced**
   - *bú:xá* ‘saying something wrongly, e.g. mispronouncing a name’
   - *bámbá:tá* ‘turning/pressing something over, e.g. when ironing and pressing clothes’
   - *bó:ka* ‘praise’
   - *be:ě:rt:a* ‘hit/hitting’ (see also *be:t:a*)
   - *bamípá* ‘hold’

3. **Implosive stops**
   - *bě:ě:t:a* ‘hit/hitting’
   - *bále:ka* ‘run/running’
   - *bó:na* ‘see’

**Alveolar**

1. **Voiceless unaspirated stops**
   - *tá:ka* ‘duck/ hurry up’
   - *tótó:ba* ‘walking slowly, especially by old people’
   - *taitá* ‘father’

**Velar**

1. **Voiceless unaspirated stops**
   - *keké:la* ‘not moving properly’
   - *kó:t:á* ‘but’
   - *kokó:sa* ‘love/value something very much/ despise’
   - *kú:ku* ‘fat cake’

---

1. This word does not appear to be used in the main-stream language. It seems to be a jargon word.

2. Where the action is done by the third person singular/plural.
2. Voiced stops

ngawú:la ‘to cut/chop’
gâ:jiga ‘to grab’
ge:za ‘to be silly/naughty’
Appendix 5E.
For Chapter 5. Acoustics: Voice Onset Time
Table 5E(1): Prevoicing, VOT, means and standard deviations (SD) for the three stop types in Shekgalagari for speaker EN. Mean 1 represents the means for the stops per place of articulation and across vowel contexts (as well as SD 1). Mean 2 represents the means for the vowels across the place of articulation for the stops, (as well as SD 2). Voicing decay and VOT for the voiced stops are indicated where applicable but their means are not calculated since their occurrence was very limited.
Table 5E(2): Prevoicing, VOT, means and standard deviations (SD) for the three stop types in Shekgalagari for speaker CM. Mean 1 represents the means for the stop per place of articulation and across vowel contexts (as well as SD 1). Mean 2 represents the means for the vowels across the place of articulation for the stops, (as well as SD 2). Voicing decay and VOT for the voiced stops were not always represented, therefore the means for these may not be reliable. The means for decay and VOT per place of articulation are shown on the table below (this table is regarded as part of the big table).
<table>
<thead>
<tr>
<th>Vowel context</th>
<th>[i]</th>
<th>[e/e]</th>
<th>[a]</th>
<th>[o/u]</th>
<th>[u]</th>
<th>Mean 1</th>
<th>SD 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>[b]</td>
<td>-118.1</td>
<td>-57.4</td>
<td>-39.4</td>
<td>-82.0</td>
<td>-85.0</td>
<td>-76.6</td>
<td>29.9</td>
</tr>
<tr>
<td>[d]</td>
<td>-100.3</td>
<td>-75.9</td>
<td>-41.6</td>
<td>-68.1</td>
<td>-95.2</td>
<td>-71.5</td>
<td>24.2</td>
</tr>
<tr>
<td>[g]</td>
<td>No prev. (VOT 24.2)</td>
<td>No prev. (VOT 14.9)</td>
<td>No prev. (VOT 35.5)</td>
<td>No prev. (VOT 40.4)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean 2</td>
<td>-99.6</td>
<td>-72.6</td>
<td>-50.8</td>
<td>-80.3</td>
<td>-90.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SD 2</td>
<td>18.8</td>
<td>13.8</td>
<td>17.9</td>
<td>11.1</td>
<td>7.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[p]</td>
<td>8.3</td>
<td>7.1</td>
<td>4.3</td>
<td>15.9</td>
<td>24.9</td>
<td>11.5</td>
<td>7.2</td>
</tr>
<tr>
<td>[t]</td>
<td>8.4</td>
<td>-</td>
<td>5.4</td>
<td>9.7</td>
<td>16.2</td>
<td>9.9</td>
<td>4.5</td>
</tr>
<tr>
<td>[k]</td>
<td>44.6</td>
<td>-</td>
<td>17.9</td>
<td>24.8</td>
<td>26.3</td>
<td>28.4</td>
<td>11.4</td>
</tr>
<tr>
<td>[g]</td>
<td>16.1</td>
<td>19.2</td>
<td>12.6</td>
<td>20.6</td>
<td>17.1</td>
<td>21.2</td>
<td>9.1</td>
</tr>
<tr>
<td>Mean 2</td>
<td>22.8</td>
<td>12.4</td>
<td>9.4</td>
<td>19.2</td>
<td>20.3</td>
<td>15.8</td>
<td>7.1</td>
</tr>
<tr>
<td>SD 2</td>
<td>6.2</td>
<td>6.2</td>
<td>5.7</td>
<td>4.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[pʰ]</td>
<td>66.0</td>
<td>66.3</td>
<td>52.7</td>
<td>77.2</td>
<td>90.3</td>
<td>70.5</td>
<td>14.1</td>
</tr>
<tr>
<td>[tʰ]</td>
<td>57.3</td>
<td>75.5</td>
<td>63.5</td>
<td>64.6</td>
<td>55.3</td>
<td>63.2</td>
<td>7.9</td>
</tr>
<tr>
<td>[cʰ]</td>
<td>123.3</td>
<td>100.4</td>
<td>80.8</td>
<td>89.2</td>
<td>116.8</td>
<td>102.1</td>
<td>17.9</td>
</tr>
<tr>
<td>[kʰ]</td>
<td>98.2</td>
<td>82.3</td>
<td>65.7</td>
<td>64.2</td>
<td>89.5</td>
<td>80.8</td>
<td>14.8</td>
</tr>
<tr>
<td>Mean 2</td>
<td>86.2</td>
<td>81.1</td>
<td>65.7</td>
<td>73.8</td>
<td>88.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SD 2</td>
<td>30.4</td>
<td>14.4</td>
<td>11.6</td>
<td>11.9</td>
<td>25.2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table SE(3): Prevoicing, VOT, means and standard deviations (SD) for the three stop types in Shekgalagari for speaker EN. Mean 1 represents the means for the stops per place of articulation and across vowel contexts (as well as SD 1). Mean 2 represents the means for the vowels across the place of articulation for the stops, (as well as SD 2). Voicing decay and VOT for the voiced stops are indicated where applicable but their means are not calculated since their occurrence was very limited.
Table 5E(4): Prevoicing, VOT, means and standard deviations (SD) for the three stop types in Shekgalagari for speaker TM. Mean 1 represents the means for the stops per place of articulation and across vowel contexts (as well as SD 1). Mean 2 represents the means for the vowels across the place of articulation for the stops, (as well as SD 2). Voicing decay and VOT for the voiced stops were not always represented, therefore the means for these may not be reliable. The means for decay and VOT per place of articulation are shown on the table below (this table is regarded as part of the big table).
### Between-Subjects Factors

<table>
<thead>
<tr>
<th>Factor</th>
<th>Levels</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>VOICING</td>
<td>asp</td>
<td>64</td>
</tr>
<tr>
<td></td>
<td>unasp</td>
<td>63</td>
</tr>
<tr>
<td></td>
<td>voiced</td>
<td>56</td>
</tr>
<tr>
<td>PLACE</td>
<td>bilab</td>
<td>47</td>
</tr>
<tr>
<td></td>
<td>dental</td>
<td>46</td>
</tr>
<tr>
<td></td>
<td>palatal</td>
<td>46</td>
</tr>
<tr>
<td></td>
<td>velar</td>
<td>44</td>
</tr>
<tr>
<td>VOWEL</td>
<td>a</td>
<td>46</td>
</tr>
<tr>
<td></td>
<td>i</td>
<td>47</td>
</tr>
<tr>
<td></td>
<td>o</td>
<td>45</td>
</tr>
<tr>
<td></td>
<td>u</td>
<td>45</td>
</tr>
<tr>
<td>SEX</td>
<td>F</td>
<td>93</td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>90</td>
</tr>
</tbody>
</table>

### Levene's Test of Equality of Error Variances

<table>
<thead>
<tr>
<th>Source</th>
<th>df1</th>
<th>df2</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>195</td>
<td>.95</td>
</tr>
<tr>
<td>VOICING</td>
<td>2</td>
<td>198</td>
<td>.67</td>
</tr>
<tr>
<td>PLACE</td>
<td>3</td>
<td>198</td>
<td>.55</td>
</tr>
<tr>
<td>VOWEL</td>
<td>3</td>
<td>198</td>
<td>.55</td>
</tr>
<tr>
<td>SEX</td>
<td>1</td>
<td>198</td>
<td>.52</td>
</tr>
<tr>
<td>VOICING * PLACE</td>
<td>6</td>
<td>198</td>
<td>.00</td>
</tr>
<tr>
<td>VOICING * VOWEL</td>
<td>6</td>
<td>198</td>
<td>.00</td>
</tr>
<tr>
<td>PLACE * VOWEL</td>
<td>9</td>
<td>198</td>
<td>.06</td>
</tr>
<tr>
<td>VOICING * PLACE * VOWEL</td>
<td>18</td>
<td>198</td>
<td>.94</td>
</tr>
</tbody>
</table>

### Tests of Between-Subjects Effects

<table>
<thead>
<tr>
<th>Source</th>
<th>Type III Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>Hypothesis</td>
<td>1268.869</td>
<td>1</td>
<td>1268.869</td>
<td>6.378</td>
</tr>
<tr>
<td>Error</td>
<td>158.949</td>
<td>1</td>
<td>158.949</td>
<td>63.78</td>
<td>6.378</td>
</tr>
<tr>
<td>VOICING</td>
<td>Hypothesis</td>
<td>849173.923</td>
<td>2</td>
<td>424586.961</td>
<td>1623.182</td>
</tr>
<tr>
<td>Error</td>
<td>523.154</td>
<td>2</td>
<td>261.577</td>
<td>1623.182</td>
<td>.001</td>
</tr>
<tr>
<td>PLACE</td>
<td>Hypothesis</td>
<td>9697.823</td>
<td>3</td>
<td>3232.608</td>
<td>81.965</td>
</tr>
<tr>
<td>Error</td>
<td>116.316</td>
<td>3</td>
<td>39.439</td>
<td>81.965</td>
<td>.002</td>
</tr>
<tr>
<td>VOWEL</td>
<td>Hypothesis</td>
<td>1954.231</td>
<td>3</td>
<td>651.410</td>
<td>3.526</td>
</tr>
<tr>
<td>Error</td>
<td>554.251</td>
<td>3</td>
<td>184.752</td>
<td>3.526</td>
<td>.154</td>
</tr>
<tr>
<td>SEX</td>
<td>Hypothesis</td>
<td>198.949</td>
<td>1</td>
<td>198.949</td>
<td>5.59</td>
</tr>
<tr>
<td>Error</td>
<td>858.297</td>
<td>2</td>
<td>429.149</td>
<td>5.59</td>
<td>.521</td>
</tr>
<tr>
<td>VOICING * PLACE</td>
<td>Hypothesis</td>
<td>14535.815</td>
<td>6</td>
<td>2423.302</td>
<td>27.918</td>
</tr>
<tr>
<td>Error</td>
<td>520.603</td>
<td>6</td>
<td>86.800</td>
<td>27.918</td>
<td>.000</td>
</tr>
<tr>
<td>VOICING * VOWEL</td>
<td>Hypothesis</td>
<td>19893.571</td>
<td>6</td>
<td>3315.595</td>
<td>58.009</td>
</tr>
<tr>
<td>Error</td>
<td>342.933</td>
<td>6</td>
<td>57.156</td>
<td>58.009</td>
<td>.000</td>
</tr>
<tr>
<td>PLACE * VOWEL</td>
<td>Hypothesis</td>
<td>4413.842</td>
<td>9</td>
<td>490.427</td>
<td>2.891</td>
</tr>
<tr>
<td>Error</td>
<td>1526.864</td>
<td>9</td>
<td>169.652</td>
<td>2.891</td>
<td>.065</td>
</tr>
<tr>
<td>VOICING * PLACE * VOWEL</td>
<td>Hypothesis</td>
<td>1541.997</td>
<td>18</td>
<td>85.666</td>
<td>4.62</td>
</tr>
<tr>
<td>Error</td>
<td>3336.061</td>
<td>18</td>
<td>185.448</td>
<td>4.62</td>
<td>.945</td>
</tr>
<tr>
<td>VOICING * SEX</td>
<td>Hypothesis</td>
<td>523.154</td>
<td>2</td>
<td>261.577</td>
<td>5.59</td>
</tr>
<tr>
<td>Error</td>
<td>577.85</td>
<td>2</td>
<td>288.925</td>
<td>5.59</td>
<td>.521</td>
</tr>
</tbody>
</table>

Table 5E (5a): Detailed ANOVA test for between-subjects effects.
Appendix 5E: For chapter 5. Acoustics: Voice Onset Time

Table 5E (5b): Detailed ANOVA test for between-subjects effect

<table>
<thead>
<tr>
<th>Source</th>
<th>Hypothesis Type III Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>VOICING * PLACE * SEX</td>
<td>520.803</td>
<td>6</td>
<td>86.800</td>
<td>.467</td>
<td>.824</td>
</tr>
</tbody>
</table>
| VOWEL * SEX Hypothesis  | 3396.600                           | 18.266 | 186.069
| Error                   | 1864.750                           | 3  | 61.502
| VOWING * VOWEL * SEX    | 12.976                             | .311 | 41.702
| Error                   | 3418.633                           | 6  | 57.156      | .307  | 925  |
| PLACE * VOWEL *SEX      | 342.938                            | 6  | 57.156      | .307  | 925  |
| Error                   | 3418.633                           | 3  | 186.069
| VOWING * PLACE * VOWEL  | 1521.664                           | 9  | 169.652     | .906  | .538 |
| Error                   | 3467.445                           | 18.567 | 186.757
| VOWING * VOWEL * SEX    | 3338.061                           | 18 | 185.448     | .545  | .928 |
| Error                   | 29919.510                           | 87 | 340.445

Table 5E (5b): Detailed ANOVA test for between-subjects effect
Appendix 6F.
For chapter 6. Articulatory analysis: Electropalatography
Appendix 6F. For chapter 6: Articulatory analysis ~ electropalatography

**EPG Data**

**Stops in symmetrical vowel contexts**

*a*–*a*

**Dental**

matápo:le  ‘potatoes’
madále:> nonsense word which could mean ‘to the old man’
matáre:> locative word for branches

**Palatal**

páeáxa:na  ‘be bound together’
majaqo:> ‘gone to sojourn/where people to sojourn’
mae<ac>bë:> locative word for problems

**Velar**

rákáné:na  ‘meet/meet for each other’
magánó:> nonsense word
mae<ac>ané:> locative word for small trees

*u*–*u*

**Dental**

buurú:> ‘at the sour milk — meaning gone for the / to collect sour milk’
qúduqú:du onomatopoeic word for eating fast and swallowing noisily.
sur<ul>ó:xa  to appear suddenly, especially from hiding

**Palatal**

rueešé:la  ‘be educated for — (someone or something)’
muiúte:> ‘to the Jew’
mue<usi:> locative word for helper
Appendix 6F. For chapter 6. Articulatory analysis: Electropalatography

Velar
rukule:la ‘de-seed/remove the seeds on the melon for (someone /something)
muguru: an nonsense word
muk:uk:u: an locative word for small hut

i~ i

Dental
bitibiti nonsense word which could mean a trick
qidiqidi onomatopoeic word for eating fast and swallowing noisily.
sitaime:la to disappear suddenly, especially into hiding

Palatal
riciba:la ‘calm, be calmed’
bicoso: an locative word for stuff traditionally believed to be used for bewitching’
biecibea: (they are) struggling for

Velar
bikiri: an ‘at/gone for/to collect a traditionally made tin mug/ locative word’
bigirina nonsense word
biecibea: an locative word for T-shirts

Stops in asymmetrical vowel contexts

a~i

Dental
mairime:la ‘strayed cattle’
mairike:la nonsense word
laribe:la nonsense word which could mean the sun becoming
overcast with clouds

Palatal
laisare:aa ‘to follow on the trek (of something/someone) for
(something/someone)
baibaz ‘the people who eat it (food)’
kaeibe:la the sun becoming overcast with clouds
Appendix 6F. For chapter 6. Articulatory analysis: Electropalatography

Velar
bao\text{\textasciitilde}sa\text{\textasciitilde}na ‘cause each other to never repeat the same mistake again/to hurt someone in a non-forgettable way’
magiri\text{\textasciitilde}n ‘at the spot where the markings of someone or something’ sudden stop are visible on the ground/locative’
bae\text{\textasciitilde}iba\text{\textasciitilde}na name of the people living in a particular ward

i\text{\texttilde}a

Dental
ri\text{\textasciitilde}ale\text{\textasciitilde}n ‘starving, being in hunger, going through a period of hunger because of lack of food’
ria\text{\textasciitilde}aq\text{\textasciitilde}qo ‘they, e.g. goats, are coming on the other side’
bi\text{\textasciitilde}re\text{\textasciitilde}n locative for trees

Palatal
bie\text{\textasciitilde}la\text{\textasciitilde}la ‘to flatten oneself on the ground behind e.g. the bush in an attempt to hide’
ri\text{\textasciitilde}are\text{\textasciitilde}n ‘in years (past, present or future)’
ic\text{\textasciitilde}cicc\text{\textasciitilde}n (person, etc. who) solicits love, admiration

Velar
rik\text{\textasciitilde}x\text{\textasciitilde}x\text{\textasciitilde}na ‘surround’
rig\text{\textasciitilde}go\text{\textasciitilde}x\text{\textasciitilde}a ‘be burnt by the fire on the surface’
\text{\textasciitilde}i\text{\textasciitilde}de\text{\textasciitilde}la

a\text{\textasciitilde}u

Dental
ma\text{\textasciitilde}ru\text{\textasciitilde}n ‘at/gone to where there is a lot of a variety of sour milk/ lots and lots of sour milk/locative word for sour milk’
mad\text{\textasciitilde}pu\text{\textasciitilde}lo ‘marks or wounds indicating where someone was hit’
a\text{\textasciitilde}h\text{\textasciitilde}ile\text{\textasciitilde}la nonsense word
Appendix 6F. For chapter 6. Articulatory analysis: Electropalatography

Palatal
açūpū:xa ‘someone/something falling from a considerable height and hitting the ground with a loud painful bang’
açūpū:xa ‘someone/something falling from a considerable height and hitting the ground with a loud painful bang’
bae_kusi:ŋ locative word for helper

Velar
ma_kukā:na ‘water container, often used for carrying water on the backs of donkeys’
ma_gup:u:ŋ nonsense word
ma_kub:u:ŋ locative word for hills

u-a

Dental
buङa:na ‘to break something’
mudape:le ‘first born/ one who comes arrives first’
surha:sut:a to dust/clean

Palatal
cuekacu:ca ‘to burn with a flicker, especially when the fire is having difficulty burning’
mujabo:j:wa ‘the heir’
phe_kophu:ca to pack

Velar
sukāx:ana ‘be crowded’
sugāx:ana nonsense word
sukāx:ana nonsense word

i-u

Dental
rifuritu ‘lots of houses/huts’
bridu:xa ‘suddenly come out a hiding place’
si:lo:xa nonsense word
Appendix 6F. For chapter 6. Articulatory analysis: Electropalatography

Palatal
bieúloŋ ‘to/at gone for the chairs’
bijúxwa:na nonsense word
biekule:xa something that can be hit

Velar
rikúkuŋ locative word for fat cakes
miguπuŋ nonsense word
rikúkwa:na:na beetles

u−i

Dental
mufimiri word improvised from Setswana meaning a lost person
mufikón nonsense word
muribele:lo nonsense word

Palatal
cuęise:za ‘light the fire for’
lujikelo nonsense word
mucęibe:ri herdsman

Velar
súkiriŋ locative word for sugar
mugiloŋ nonsense word
súkiriŋ nonsense word which could be locative word for sugar
Bibliography
Bibliography


Bibliography


Bibliography


Bibliography


Bibliography


